International Journal of Methods in Psychiatric Research
Int. J. Methods Psychiatr. Res. 17(1): 12–29 (2008)
Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/mpr.236

Ascertaining late-life depressive symptoms in Europe: an evaluation of the survey version of the EURO-D scale in 10 nations. The SHARE project

ERICO CASTRO-COSTA,[1,2,3] MICHAEL DEWEY,[1] ROBERT STEWART,[1] SUBE BANERJEE,[1] FELICIA HUPPERT,[4] CARLOS MENDONCA-LIMA,[5] CHRISTOPHE BULA,[6] FRIEDEL REISCHES,[7] JOHANNES WANCATA,[8] KAREN RITCHIE,[9] MAGDA TSOLAKI,[10] RAIMUNDO MATEOS,[11] MARTIN PRINCE[1]

1 Institute of Psychiatry, Department of Health Services and Population Research, King's College London, London, UK
2 Centro de Pesquisa René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Brazil
3 Faculdade da Saúde e Ecologia Humana (FASEH), Vespasiano, Brazil
4 Department of Psychiatry, University of Cambridge, Cambridge, UK
5 Department of Psychiatry, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
6 Service of Geriatric Medicine & Geriatric Rehabilitation, University of Lausanne Medical Center (CHUV), Lausanne, Switzerland
7 Universitatsklinik und Hochschulambulanz fur Psychiatrie und Psychotherapie, Campus Benjamin Franklin, Berlin, Germany
8 Department of Psychiatry, Division of Social Psychiatry, Medical University of Vienna, Vienna, Austria
9 Institut National de la Santé et la Recherche Médicale, E361 Pathologies of the Nervous System: Epidemiological and Clinical Research, Montpelier, France
10 Department of Neurology, Aristotle University of Thessaloniki, Thessaloniki, Greece
11 Departamento de Psiquiatria, Facultad de Medicina, Santiago de Compostela, Spain

Abstract

The reported prevalence of late-life depressive symptoms varies widely between studies, a finding that might be attributed to cultural as well as methodological factors. The EURO-D scale was developed to allow valid comparison of prevalence and risk associations between European countries. This study used Confirmatory Factor Analysis (CFA) and Rasch models to assess whether the goal of measurement invariance had been achieved; using EURO-D scale data collected in 10 European countries as part of the Survey of Health, Ageing and Retirement in Europe (SHARE) (n = *22,777). The results suggested a two-factor solution (Affective Suffering and Motivation) after Principal Component Analysis (PCA) in 9 of the 10 countries. With CFA, in all countries, the two-factor solution had better overall goodness-of-fit than the one-factor solution. However, only the Affective Suffering subscale was equivalent across countries, while the Motivation subscale was not. The Rasch model indicated that the EURO-D was a hierarchical scale. While the calibration pattern was similar across countries, between countries agreement in item calibrations was stronger for the items loading on the affective suffering than the motivation factor. In conclusion, there is evidence to support the EURO-D as either a uni-dimensional or bi-dimensional scale measure of depressive symptoms in late-life across European countries. The Affective*

Suffering sub-component had more robust cross-cultural validity than the Motivation sub-component.

Key words: depression, late-life, EURO-D, factor analysis item response theory

Introduction

Among late-life psychiatric disorders, depression is probably the most frequent cause of emotional suffering and has a significant impact on quality of life (Blazer, 2003). A wide range in depression prevalence has been found between studies (0.4–35%). Whilst there may be important contextual differences or differences in risk factor prevalence accounting for this, methodological issues require further evaluation: in particular, potential between-centre variations in the performance of diagnostic instruments (Beekman et al., 1999).

The EURO-D (Prince et al., 1999b) was originally developed in an effort to harmonize data on late-life depression from population-based studies in 11 European countries as part of the EURODEP collaboration. The EURO-D items were all taken from the Geriatric Mental State (GMS: Copeland et al., 1986). However, they were also used in several of the other assessments to be harmonized – the SHORT-CARE, the Centre for Epidemiological Studies Depression scale (CES-D) the Zung Self-rating Depression Scale and Comprehensive Psychopathological Rating Scale (CPRS) (Prince et al., 1999b). Thus, the EURO-D has intrinsically strong face validity. The initial validation showed adequate internal consistency, with the optimal cut-off point of 3/4 for prediction of both GMS depression and SHORT-CARE pervasive depression (Prince et al., 1999b). The EURO-D was found to be reliable and validated for detection of DSM-IIIR depression in older people in Spain (Larraga et al., 2006). Principal Component Analysis (PCA) generated two factors (Affective Suffering and Motivation) that were common to nearly every participating country in the EURODEP studies (Prince et al., 1999b) and for Indian, Latin-American and Caribbean centres in the 10/66 Dementia Research Group pilot studies (Prince at al., 2004).

Confirmatory Factor Analysis (CFA) and Rasch models can be used to assess whether a scale measures the same trait dimension, in the same way, when applied in qualitatively distinct groups (Reise et al., 1993); in this instance populations from different countries and cultures. The aim of this analysis was to investigate whether data on late-life depression symptoms collected using EURO-D in 10 European countries participating in the Survey of Health, Ageing and Retirement in Europe (SHARE) support the cross-cultural validity of the measure.

Method

Design

We use data from Release 1 of the SHARE 2004 baseline study (Borsch-Supan et al., 2005). This comprised cross-sectional surveys of representative samples of community residents aged 50 years and over from 10 European countries from Scandinavia (Denmark and Sweden) through Central Europe (Austria, France, Germany, Switzerland and the Netherlands) to the Mediterranean (Spain, Italy and Greece). The actual fieldwork in SHARE was carried out by a different agency for each country, but the programming of the individual instruments was done centrally by CentER-data, a survey research institute affiliated with Tilburg University in the Netherlands. The data were collected using a computer assisted personal interviewing (CAPI) program, supplemented by a self-completion paper and pencil questionnaire. The set-up of this CAPI program allowed each country involved to use exactly the same underlying structure of meta-data and routing. The only difference across countries was the language. Next to the CAPI instrument, a Case Management System (CMS) was developed to manage the co-ordination of the fieldwork. In the participating SHARE countries the institutional conditions with respect to sampling are so different that a uniform sampling design for the entire project was infeasible. As a result the sampling designs have varied from simple random selection of households to more complex multi-stage designs according to the choice of the Principal Investigator (PI) from each country to seek maximal representativeness. A train-the-trainer (TIT) programme was developed by the Survey Research Centre (SRC) of the University of Michigan at Ann Arbor for the SHARE project, providing centralized training of local survey agency

DOI: 10.1002/mpr

trainers in order to facilitate standard training of interviewers and standardization of the data collection processes in the respective countries.

SHARE was primarily carried out to investigate economic issues around and after retirement age and was designed to provide comparable data to the US Health and Retirement Study (HRS, 2006) and the English Longitudinal Study of Ageing (ELSA, 2006). It has the advantage of encompassing cross-national variation in public policy, culture and socio-political history. The data is publicly accessible (www.share-project.org).

Mental health and psychological status assessment

The SHARE database includes variables and indicators created by the AMANDA RTD-project under the European Union's 5th framework programme. This module included the objective of developing and validating indicators of mental health and cognitive functioning suitable for use in European surveys. Mood was measured using EURO-D (Prince et al., 1999a; Prince et al., 1999b) with 12 items covering depression, pessimism, suicidality, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment and tearfulness in the last month (see Appendix). It is scored, by summing item scores for individual symptoms that are coded as 0 and 1 when they are 'not present' and 'present', respectively. Total score ranging from 0 to 12 a higher score indicating more depressive symptoms.

Statistical analyses

First we describe for each country the sample characteristics (age, gender, marital status, education and retirement status), together with mean scores for EURO-D, mean factor scores for Affective Suffering (depression, tearfulness, suicidality, sleep disturbance, guilt, irritability and fatigue) and Motivation (interest, enjoyment, concentration and pessimism) factors originated from the PCA carried out in this study, and Cronbach's alpha for all 12 EURO-D items.

PCA was based on a covariance matrix of tetrachoric correlations to reduce bias in the estimation of factor loadings (Olsson, 1979). The cut off used to assume that an item loaded on a given factor was 0.55. A varimax rotation was carried out with an eigenvalue of one as initial extraction criterion.

In CFA a model is tested that specifies in advance the relations between observed variables and latent factors. Such a model contains parameters that are (a) fixed to a certain value, (b) constrained to be equal to other parameters, or/and (c) free to take on any unknown value. We wished to test for measurement invariance (Sorbom, 1974); that is whether the best factor solution was related to the latent trait or traits in the same way across the 10 countries represented in our sample. We compared two models in which all item loadings were (i) constrained and (ii) not constrained to be identical between countries. In each case factor variances and covariances were sample specific. We aimed to test for Partial Measurement Invariance if Full Measurement Invariance was rejected. Partial Measurement Invariance occurs if a majority of the non-fixed values are still invariant across groups and if these invariant loadings define the latent metric. For each model the absolute fit of the model was evaluated by means of a χ^2 statistic. The χ^2 test, however, is very sensitive to technical conditions, particularly large sample size or a violation of the multivariate normality assumption (Bentler and Bonett, 1980). As recommended (Bollen and Long, 1993; Kline, 1998; MacCallum, 1990) we used several other absolute and relative indices of fit. Akaike's Information Criterion (AIC: Akaike, 1987) adjusts the model chi-square to penalize for model complexity. The lower the AIC value, the better the fit of the model (Burnham and Anderson, 1998). The Tucker–Lewis Index (TLI: Tucker and Lewis, 1973), indicates the proportion of covariation among indicators explained by the model relative to a null model of independence, and is independent of sample size. Values near zero indicate poor fit, whereas values near 1.0 indicate good fit; those greater than 0.90 are considered satisfactory (Dunn et al., 1993; Marsh et al., 1996). In contrast, the Root Mean Square Error of Approximation (RMSEA) assesses badness-of-fit per degree of freedom in the model and equals zero if the model fits perfectly; RMSEA values of less than 0.05 indicate close fit and 0.05 to 0.08 reasonable fit of a model (Browne, 1990).

Rasch analysis belongs to a family of statistical models developed from Item Response Theory (IRT) (Nunnally and Bernstein, 1994; Reise et al., 1993). Whereas CFA models account for the covariance between test items, IRT models account for patterns of item responses. The Rasch model (Bond and Fox, 2001) suggests that the responses to a set of items can be justified by a person's position on the underlying trait that is being measured and by the characteristics (parameters) of the items. The key parameter is the item

DOI: 10.1002/mpr

severity (also referred to as item calibration). It is measured on the same scale as the severity of the underlying trait. Participants with severity scores below the calibration for the given item are more likely to deny than endorse the item, while those with severity scores above the calibration for the item are more likely to endorse it. It follows that the items with higher calibrations are less frequently endorsed. Also, a participant who endorses an item of middling severity is likely to have endorsed all items with lower calibrations. By the same token, one who denies an item at midrange is likely to deny all items with higher calibrations. These assumptions are only probabilistically true; not all participants will follow these expected patterns exactly. INFIT (inlier–sensitive or information–weighted fit) assesses the extent to which observed response patterns were consistent, or inconsistent with the item calibrations. The probability of a positive response in each cell of the person-by-item matrix is calculated. The INFIT statistic compares the actual response to the probabilistically expected response in that cell. It is an 'information-weighted' fit statistic for each item that is sensitive to responses by persons with severity scores in the range near the calibration of the particular item. Mean square INFIT (MNSQ) statistics of <1.3 indicate that the subscale items contribute to a single hierarchical underlying construct (unidimensionality). While there can be no single satisfactory test of model adequacy, techniques can be used to assess different aspects of goodness-of-fit of Rasch models. The Andersen Z fit tests the assumption that the estimations of the item calibrations are the same, whatever the level of the latent trait (person homogeneity) (Andersen, 1973). The sample is divided into groups as a function of the score and the calibrations are estimated in each of these groups. The statistics follow, under the null assumption, a chi-square distribution.

The invariance of the item calibrations by country was determined through plotting a matrix of Spearman correlations, which was used to estimate the consistency between countries in the rank order of the calibrations assigned to the items. Measurement invariance with respect to item calibration would be suggested by consistently high correlations across all pairs of countries. Overall agreement across the 10 countries was assessed with an intraclass correlation coefficient.

Description of data, PCA, Rasch models and Spearman's correlation coefficients were conducted with STATA 9.1, and CFA with AMOS 5 (Arbuckle, 2003). The parameters of the Rasch model and the evaluation of its fit were estimated in STATA 9.2 with the incorporation of the Raschtest module developed by Hardouin (2001).

Results

Characteristics of participants across the 10 countries

The SHARE probability samples represent the non-institutionalized population aged 50 and older in 10 European countries. The breakdown of the 2004 sample (Release 1) by country, sex and age has been reported elsewhere (Borsch-Supan et al., 2005). The household response rate across all countries was 57.4%, with the lowest rate in Switzerland (37.6%) and the highest in France (69.4%). Individual response (proportion of eligibles interviewed in consenting households) ranged from 73.8% (Italy) to 93.0% (Denmark) with 86.0% overall (Borsch-Supan et al., 2005). In Table 1, we summarize the characteristics of participants who answered the EURO-D. Their mean (standard deviation, SD) age was 64.2 (10.5) years, 10,742 (54.5%) were female, 15,906 (70.6%) were married and 10,749 (47.9%) were retired. The mean EURO-D score ranged from 1.80 (Denmark) to 3.10 (Spain) with an overall mean of 2.25 (SD = 2.24). Affective Suffering mean factor scores ranged from 0.20 (Sweden) to 0.29 (Germany), and the Motivation mean factor scores from −0.07 (Sweden) to 0.18 (Switzerland). In all countries, the EURO-D scale was moderately internally consistent with Cronbach's alpha ranging from 0.62 to 0.78.

Principal components analyses (PCA)

The PCA before rotation, with an eigenvalue of 1 or more as extraction criterion, gave rise to two factors in all countries other than Switzerland where a three-factor solution was found. The first factor explained between 41.4% and 69.5% of the variance (eigenvalues 4.55 to 6.14), and the second 10.3% to 18.4% (eigenvalues between 1.20 and 1.83) (Table 2). In Switzerland, the third factor had an eigenvalue of 1.25. The depression item loaded most strongly (≥0.80) for factor 1 in all countries, followed by tearfulness. Others items such as suicidality, guilt, irritability, sleep disturbance and fatigue also typically loaded over 0.60 on factor 1. Pessimism was the highest loading item on factor 2 (≥0.70). Enjoyment, concentration and interest were also highly correlated with factor 2, however loadings varied from 0.55 to 0.70 between countries. Enjoyment and pessimism loaded on a separate third factor in

DOI: 10.1002/mpr

Table 1. The SHARE Release 1 – basic demographic characteristics and the EURO-D score for depressive symptoms (n = 22,777)

	Switzerland (N = 1010)	Greece (N = 2142)	Denmark (N = 1732)	France (N = 1842)	Italy (N = 2559)	Spain (N = 2419)	The Netherlands (N = 3000)	Sweden (N = 3067)	Germany (N = 3020)	Austria (N = 1986)
Age (years/%)										
50–59	37.03	37.54	39.95	40.34	33.04	31.06	41.69	35.43	33.97	30.65
60–74	41.74	42.62	39.52	38.70	51.68	44.51	42.28	44.85	49.54	50.57
≥75	21.23	19.84	20.53	20.96	15.28	24.44	16.03	19.72	16.49	18.78
Mean (range)	65 (50–96)	64 (50–97)	64 (50–104)	64 (50–99)	64 (50–100)	66 (50–103)	63 (50–99)	65 (50–102)	64 (50–97)	65 (50–100)
Female (%)	53.66	57.94	54.68	56.89	55.76	58.50	54.11	53.57	54.14	58.71
Married (%)	67.60	69.23	62.68	66.89	77.00	72.26	77.28	68.78	75.21	59.01
Education (years)										
Mean	12	8	13	8	7	5	11	10	13	11
Retired (%)	45.6	45.5	50.9	50.7	54.5	35.0	32.5	52.6	51.1	64.5
Euro-D										
Mean score	1.9	2.2	1.8	2.8	2.8	3.1	1.9	2.0	2.0	1.9
Affective Suffering										
Mean factor score	0.27	0.26	0.27	0.30	0.27	0.24	0.24	0.20	0.29	0.28
Motivation										
Mean factor score	0.18	0.09	0.07	0.07	0.09	0.11	0.11	−0.07	0.03	0.01
Cronbach's α	0.62	0.72	0.67	0.70	0.73	0.78	0.68	0.64	0.70	0.72

Table 2. The SHARE Release 1 – PCA based on a tetrachoric correlation matrix for EURO-D (SHARE, 2004)

	Two-factor solution		Number of factors with eigenvalue > 1	Content of factors not captured in two-factor solution
	Factor 1	Factor 2		
Pool			2	
Variance (%)	48.1	11.5		
Item loading	**depression**	**enjoyment**		
	tearfulness	**pessimism**		
	suicidality	interest		
	guilt	concentration		
	irritability			
	sleep			
Switzerland			3	Enjoyment, pessimism
Variance (%)	41.4	12.8		
Item loading	**depression**	**irritability**		
	tearfulness	**appetite**		
	guilt	**interest**		
	suicidality	*fatigue*		
Greece			2	
Variance (%)	48.3	12.6		
Item loading	**depression**	**interest**		
	tearfulness	**pessimism**		
	suicidality	concentration		
	guilt			
	irritability			
	fatigue			
	sleep			
	appetite			
Denmark			2	
Variance (%)	44.8	15.2		
Item loading	**depression**	**enjoyment**		
	tearfulness	**pessimism**		
	sleep	*concentration*		
	suicidality			
	irritability			
	appetite			
	interest			
France			2	
Variance (%)	42.8	12.2		
Item loading	**depression**	**pessimism**		
	sleep	**enjoyment**		
	tearfulness	**interest**		
	guilt	concentration		
	suicidality			
	fatigue			
	irritability			
Italy			2	
Variance (%)	46.6	12.3		
Item loading	**depression**	**enjoyment**		
	tearfulness	**pessimism**		
	guilt	appetite		
	irritability	concentration		
	suicidality	interest		
	fatigue			
	sleep			

DOI: 10.1002/mpr

Table 2. *Continued*

	Two-factor solution		Number of factors with eigenvalue > 1	Content of factors not captured in two-factor solution
	Factor 1	Factor 2		
Spain			2	
Variance (%)	51.2	11.3		
Item loading	**depression**	**pessimism**		
	guilt	**enjoyment**		
	irritability	interest		
	tearfulness	concentration		
	suicidality	appetite		
	sleep			
The Netherlands			2	
Variance (%)	51.2	11.3		
Item loading	**depression**	**pessimism**		
	guilt	**enjoyment**		
	irritability	interest		
	tearfulness	concentration		
	suicidality	appetite		
	sleep			
Sweden			2	
Variance (%)	69.5	18.4		
Item loading	**depression**	**pessimism**		
	tearfulness	suicidality		
	irritability	enjoyment		
	sleep			
Germany			2	
Variance (%)	50.9	10.3		
Item loading	**depression**	**enjoyment**		
	tearfulness	**pessimism**		
	suicidality	interest		
	irritability	*concentration*		
	appetite			
	fatigue			
	sleep			
Austria			2	
Variance (%)	53.9	11.4		
Item loading	**depression**	**enjoyment**		
	suicidality	**pessimism**		
	interest	*concentration*		
	fatigue			
	sleep			
	tearfulness			
	appetite			
	irritability			

Bold typeface: items loading at ≥0.70; Roman typeface: items loading at ≥0.60; italic typeface: items loading at 0.55–0.59.

DOI: 10.1002/mpr

Switzerland. When data were pooled from all centres, a two-factor solution also emerged, with factors 1 and 2 accounting for 48.1% and 11.5% of the variance respectively.

Factors 1 and 2 had been identified in the earlier EURO-D development (Prince et al., 1999b), with a very similar item loading pattern, and were then referred to as Affective Suffering and Motivation, respectively. Only appetite tended to crossload with average loadings of 0.53 on Affective Suffering and 0.59 on Motivation.

Confirmatory Factor Analysis (CFA)

First we fitted a one-factor solution with no constraints (all item loadings freely estimated). This provided a reasonable fit for the 12-item EURO-D scale (χ^2 = 4352.4; $p < 0.001$; df = 540; AIC = 3272.4; TLI = 0.84; RMSEA = 0.018).

Table 3 shows results for confirmatory factors for the two-factor solution (Affective Suffering and Motivation) derived from the PCA in the current study. Three different models were obtained, with different constraints applied in each case (Model 1: no constraints, as per the one factor solution; Model 2: both factors are constrained to load equally across all countries; Model 3: only items from Affective Suffering factor were constrained). As expected, each of the three two-factor solution models fitted better (lower AIC value, TLI ≥ 0.90 and RMSEA ≤ 0.05) than the one-factor solution (Model 1: χ^2 = 2637.9, df = 430, AIC = 1777.9, TLI = 0.89, RMSEA = 0.015; Model 2: χ^2 = 3375.3, df = 511, AIC = 2353.3, TLI = 0.88, RMSEA = 0.016; Model 3: χ^2 = 3050.4, df = 484, AIC = 2082.4, TLI = 0.89, RMSEA = 0.015).

Comparison between Model 2, with all items loadings constrained, and Model 1 showed that the imposed constraints increased the chi-square values and degrees of freedom, with a change of 737.4 in chi-square value and 81 more degrees of freedom. More importantly, the AIC value for Model 2 (2353.3), was higher than the AIC model for Model 1 (1777.9), and the goodness-of-fit indices (TLI = 0.88, RMSEA = 0.016) were slightly worse than those for Model 1. Therefore, the full invariance model did not hold across countries, and was rejected.

In the partial invariance model (Model 3), specified after closer inspection of Model 1, the factor loading parameters of motivation items were freely estimated, while those for the affective suffering items were constrained. While the chi-square change (412.5 on 54 degrees of freedom) still suggests that Model 1 is superior, the three absolute goodness-of-fit indices provide strong evidence that the partial invariance model (Model 3) is adequate and does not clearly differ from Model 1, which was without any restriction.

Item performance of EURO-D using Rasch model

Average item calibrations and INFIT values for each EURO-D item for the 10 countries included in the SHARE project are presented in Table 4. Lower item calibrations suggest a high probability of item endorsement by those with low scores on the trait. Depression (followed by sleep disturbance) had the lowest item calibrations overall, while guilt and suicidality had the highest. Overall, the goodness-of-fit of the EURO-D items was satisfactory for all countries. The Andersen Z Likelihood test for person heterogeneity was statistically significant for all countries suggesting that items do not have the same meaning for people at different scores. The suicidality item had the worst fit, with INFIT indices ranging from 0.71 in Greece and Austria to 0.80 in Italy, and interest showed a slight misfit in Austria.

The rank ordering of EURO-D item calibration values was similar for all countries. Spearman correlation coefficients between the set of EURO-D item calibrations for pairs of SHARE project countries ranged from 0.70 to 0.97 (Table 5). The intraclass correlation coefficient for agreement in item calibrations across all 10 countries was 0.89 (0.78–0.95). There is therefore strong evidence for measurement invariance with respect to item calibration. `

However, Figures 1 and 2 indicate clearly that there is more heterogeneity in item calibration between countries for the motivation than for the affective suffering items. This was confirmed by further calculations of intraclass correlations for the two subsets of items. For the subset of items loading on affective suffering the item calibration by country (ICC) was 0.94 (0.85–0.99) whereas for the four motivation items it was 0.65 (0.30–0.96).

Discussion

This is the first time that the EURO-D has been administered as a self-contained scale rather than nested within its parent GMS instrument. The survey version of the scale was found to be feasible and easily administrable by professional social survey organizations across Europe. We were able to evaluate its

Table 3. CFA – two-factor solutions with or without any constraint

	Two-factor solution										
	Affective Suffering							Motivation			
	Dep	Tear	Suic	Slep	Guil	Irrit	Fat	Int	Enj	Pess	Con
Model 1											
Switzerland	1.00	0.66	0.27	0.61	0.21	0.38	0.69	1.00	2.19	1.75	3.26
Greece	1.00	0.81	0.30	0.60	0.25	0.52	0.86	1.00	1.34	1.01	1.05
Denmark	1.00	0.56	0.27	0.64	0.21	0.56	0.59	1.00	0.67	0.32	1.06
France	1.00	0.72	0.46	0.71	0.38	0.51	0.77	1.00	1.05	1.37	1.39
Italy	1.00	0.72	0.25	0.68	0.23	0.61	0.84	1.00	1.09	0.75	1.45
Spain	1.00	0.82	0.48	0.71	0.23	0.67	0.62	1.00	0.63	0.81	0.95
Netherlands	1.00	0.67	0.30	0.64	0.34	0.58	0.74	1.00	0.33	0.44	1.14
Sweden	1.00	0.67	0.19	0.57	0.15	0.50	0.61	1.00	0.59	0.67	0.95
Germany	1.00	0.71	0.34	0.72	0.19	0.47	0.71	1.00	0.68	0.85	1.39
Austria	1.00	0.70	0.26	0.72	0.17	0.42	0.81	1.00	0.84	1.05	1.01
Factor loading											
Means (SD)	—	0.70 (0.07)	0.31 (0.09)	0.66 (0.05)	0.24 (0.07)	0.52 (0.09)	0.72 (0.10)	—	0.94 (0.52)	0.90 (0.41)	1.36 (0.68)
χ^2	2637.9	$p < 0.001$									
df	430										
χ^2 change	—										
df change	—										
AIC	1777.9										
TLI	0.89										
RMSEA	0.015										

DOI: 10.1002/mpr

Model 2													
Switzerland								1.00	0.76				
Greece								1.00	1.35				
Denmark								1.00	0.67				
France								1.00	1.06				
Italy	1.00	0.71	0.30	0.66	0.23	0.54	0.74	1.00	1.09	0.74	0.54	0.23	0.67
Spain								1.00	0.64				
Netherlands								1.00	0.33				
Sweden								1.00	0.60				
Germany								1.00	0.67				
Austria								1.00	0.82				
χ^2	3375.3												
df	511												
χ^2 change	737.4												
df change	81												
AIC	2353.3												
TLI	0.88												
RMSEA	0.016												
Model 3													
Switzerland								1.00	1.72				3.22
Greece								1.00	1.02				1.06
Denmark								1.00	0.32				1.07
France								1.00	1.38				1.40
Italy	1.00	0.71	0.30	0.67	0.23	0.54	0.74	1.00	0.76	0.74	0.54	0.23	1.45
Spain								1.00	0.81				0.96
Netherlands								1.00	0.44				1.14
Sweden								1.00	0.69				0.96
Germany								1.00	0.84				1.38
Austria								1.00	1.03				0.99

Table 3. *Continued*

	Two-factor solution										
	Affective Suffering							Motivation			
	Dep	Tear	Suic	Slep	Guil	Irrit	Fat	Int	Enj	Pess	Con
Factor loading											
Means (SD)									0.94 (0.52)	0.90 (041)	1.36 (0.68)
χ^2	3050.4										
df	484										
χ^2 change	412.5										
df change	54										
AIC	2082.4										
TLI	0.89										
RMSEA	0.015										

df: degree of freedom; AIC: Akaike's Information Criterion; TLI: Tucker–Lewis index; RMSEA: Root Mean Square Error Approximation; Dep: Depression, Tear: Tearfulness, Suic: Suicidality, Guil: Guilty, Slep: Sleep; Irrit: Irritability; Apet: appetite; Fat: Fatigue; Int: Interest; Enj: enjoyment; Pess: Pessimism, Con: Concentration.

DOI: 10.1002/mpr

Table 4. Comparison between different countries of EURO-D item Rasch rating scale calibrations, goodness-of-fit (INFIT) values (SHARE, 2004)

	Switzerland	Greece	Denmark	France	Italy	Spain	The Netherlands	Sweden	Germany	Austria
Depression										
Calibration	−0.99	−0.13	−1.03	−1.10	−0.93	−0.45	−0.28	−0.65	−0.73	−0.75
INFIT	0.87	0.81	0.82	0.82	0.82	0.79	0.81	0.84	0.85	0.82
Rank of difficulty	(1)	(2)	(1)	(1)	(1)	(2)	(1)	(2)	(1)	(1)
Sleep										
Calibration	−0.29	−0.54	−0.81	−0.62	−0.32	−0.10	−0.02	−0.46	−0.41	−0.69
Infit	0.94	0.92	0.94	0.96	0.94	0.96	0.96	0.95	0.97	0.97
Rank of difficulty	(2)	(1)	(3)	(2)	(4)	(3)	(2)	(3)	(2)	(2)
Fatigue										
Calibration	−0.23	0.22	−0.95	−0.38	−0.47	−0.72	0.07	−0.71	0.10	−0.45
INFIT	0.87	0.76	0.98	0.85	0.84	1.04	0.88	0.93	0.89	0.84
Rank of difficulty	(3)	(4)	(2)	(4)	(3)	(1)	(4)	(1)	(4)	(3)
Irritability										
Calibration	0.10	0.39	−0.32	−0.39	−0.53	0.41	0.70	0.09	0.84	0.55
INFIT	0.98	1.02	0.94	1.06	1.04	0.97	0.89	0.92	0.94	0.97
Rank of difficulty	(5)	(5)	(4)	(3)	(2)	(6)	(6)	(5)	(6)	(8)
Tearfulness										
Calibration	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
INFIT	0.90	0.95	0.91	0.89	0.91	0.91	0.97	0.92	0.96	0.87
Rank of difficulty	(4)	(3)	(5)	(5)	(7)	(4)	(3)	(4)	(3)	(5)
Concentration										
Calibration	0.78	0.48	0.57	0.27	−0.27	0.07	0.42	0.64	0.79	0.40
INFIT	0.93	1.09	0.90	0.98	0.91	0.96	0.95	0.95	0.92	0.97
Rank of difficulty	(6)	(6)	(6)	(6)	(5)	(5)	(5)	(6)	(5)	(7)
Enjoyment										
Calibration	1.11	1.08	0.81	1.28	−0.01	1.39	1.15	1.18	1.55	0.31
INFIT	0.97	0.93	1.01	0.92	1.04	0.98	1.10	0.95	0.99	1.05
Rank of difficulty	(7)	(7)	(7)	(10)	(6)	(9)	(7)	(7)	(7)	(6)

DOI: 10.1002/mpr

Table 4. *Continued*

	Switzerland	Greece	Denmark	France	Italy	Spain	The Netherlands	Sweden	Germany	Austria
Appetite										
Calibration	1.89	1.99	1.20	1.60	1.68	1.40	2.21	1.69	2.17	1.25
INFIT	0.79	0.79	0.80	0.88	0.87	0.86	0.79	0.78	0.79	0.82
Rank of difficulty	(10)	(11)	(8)	(11)	(11)	(10)	(11)	(11)	(11)	(9)
Pessimism										
Calibration	1.45	1.17	1.70	0.31	0.33	0.70	1.61	1.35	1.71	−0.20
INFIT	0.93	1.01	0.95	1.03	1.09	1.02	0.99	0.91	0.92	1.08
Rank of difficulty	(8)	(8)	(12)	(7)	(8)	(8)	(8)	(8)	(8)	(4)
Interest										
Calibration	2.45	1.52	1.30	1.60	1.17	1.27	1.84	1.40	2.06	1.59
INFIT	0.83	0.92	0.78	0.83	0.76	0.74	0.71	0.80	0.75	0.68
Rank of difficulty	(12)	(9)	(10)	(11)	(9)	(9)	(10)	(10)	(9)	(10)
Guilt										
Calibration	1.95	1.93	1.22	1.15	1.64	2.42	1.64	1.38	2.43	2.28
INFIT	0.80	0.93	0.89	0.90	0.97	1.02	0.87	0.99	0.87	0.91
Rank of difficulty	(11)	(10)	(9)	(8)	(10)	(12)	(9)	(9)	(12)	(12)
Suicidality										
Calibration	1.83	2.42	1.43	1.23	2.02	1.87	2.42	2.19	2.15	2.19
INFIT	0.76	0.71	0.74	0.79	0.80	0.78	0.73	0.76	0.73	0.71
Rank of difficulty	(9)	(12)	(11)	(9)	(12)	(11)	(12)	(12)	10	11
Andersen LR test[1]										
Z	158.47	431.14	226.71	267.76	404.61	396.94	515	311.17	292.32	351.37
df	110	110	110	110	110	110	110	110	110	110
p	0.0017	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

[1]There is a statistical difference between participants with severe depressive symptoms and non-severe depressive symptoms across all countries.

DOI: 10.1002/mpr

Table 5. Spearman's correlation coefficients of EURO-D item calibrations across countries

	Switzerland	Greece	Denmark	France	Italy	Spain	The Netherlands	Sweden	Germany
Switzerland									
Greece	0.87								
Denmark	0.92	0.77							
France	0.95	0.81	0.80						
Italy	0.83	0.84	0.79	0.85					
Spain	0.82	0.84	0.85	0.79	0.78				
The Netherlands	0.93	0.91	0.86	0.87	0.80	0.86			
Sweden	0.92	0.83	0.97	0.88	0.79	0.91	0.90		
Germany	0.95	0.83	0.87	0.89	0.73	0.83	0.95	0.89	
Austria	0.78	0.70	0.72	0.78	0.74	0.81	0.80	0.77	0.82

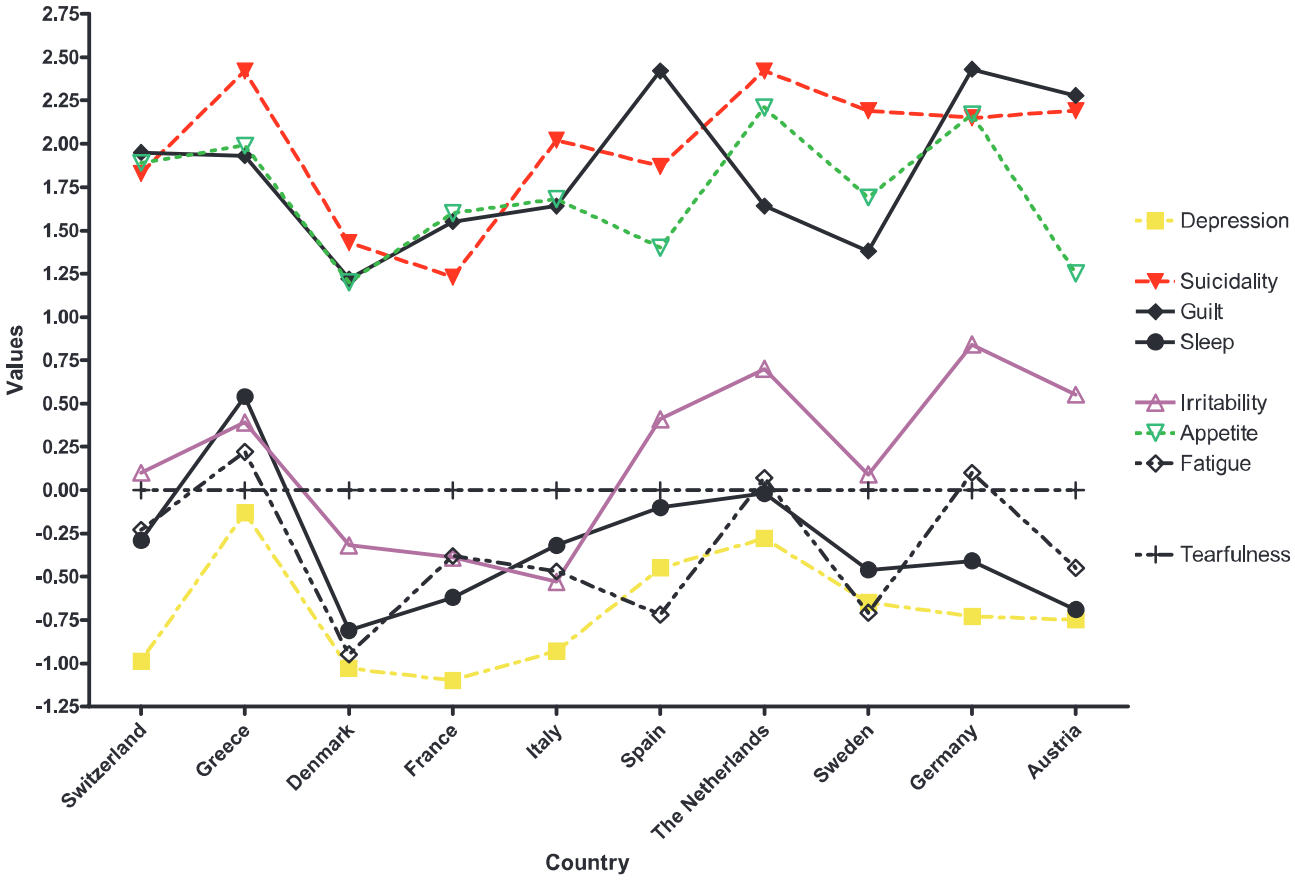

Figure 1. Average Affective Suffering item calibration by country.

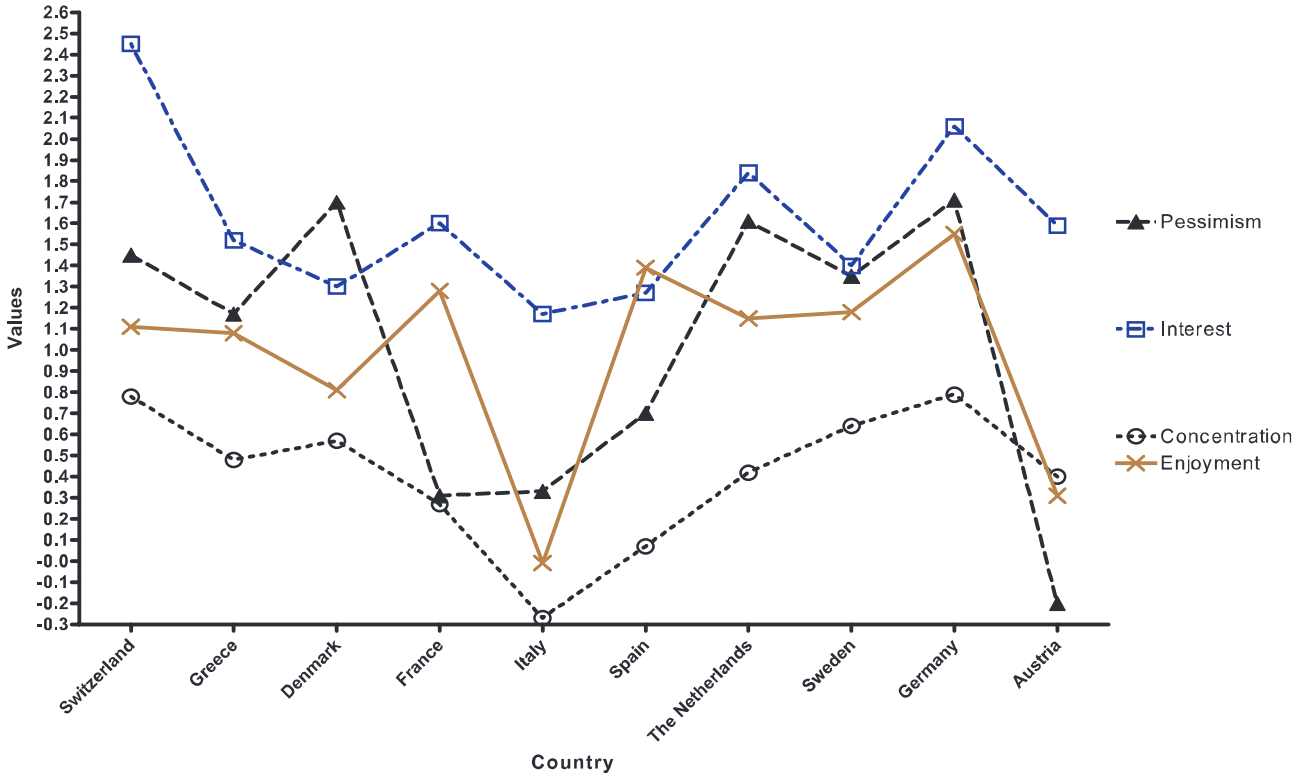

Figure 2. Average Motivation item calibration by country.

psychometric properties in detail in a very large and multicultural sample of older people from 10 European countries. In all countries, the EURO-D held up well as a uni-dimensional scale as evidenced by the moderately high Cronbach's alpha, the reasonable fit of the one factor solution and the goodness-of-fit of the Rasch model. For all these reasons, it can be commended as a brief indicator of depressed mood for large scale generic surveys of health and social circumstances in aged populations across Europe. However, there is also evidence that it may be measuring two underlying factors, Affective Suffering (well characterized and invariant across cultures) and Motivation (less well characterized and variable across cultures).

The EURO-D factor structure

The application of PCA using a covariance matrix based on a tetrachoric correlation should have reduced bias in estimation of factor loadings, and increased precision in the estimation of numbers of underlying factors. In contrast with the earlier cross-national EURO-D PCA (prince et al., 1999b) where most centres had more than two-factors component solutions, in the current study all but one (Switzerland) generated a two-factor component solution. Also, in this analysis the variance explained by the principal components was greater (Prince et al., 1999b). Both depression and tearfulness were more consistently loaded on Affective Suffering than the other symptoms (suicidality, guilt, irritability, fatigue and sleep) that tended to load less consistently across the 10 European countries. In the factor denominated Motivation, both enjoyment and pessimism were the higher loading items and they were present in almost all countries. Other items such as concentration and interest were also loaded strongly on this factor, although they were distributed more heterogeneously among the different cultures.

CFA is superior to exploratory factor analysis when some knowledge has been accumulated about the characteristics of a measure (Dunn et al., 1993; Joreskog and Sorbom, 1988). In this study, the two-factor solution suggested by previous research using PCA (and by the PCA findings for the current study discussed in the previous paragraph) showed better goodness-of-fit indices than the one-factor solution for EURO-D, as judged by the much lower AIC. The absolute

DOI: 10.1002/mpr

goodness/badness-of-fit indices, TLI and RMSEA suggested a generally adequate fit for the two-factor solution. Although TLI was just under the more stringent recommended threshold of 0.90, it was higher than the more liberal cut-off point of 0.80 proposed by others (Ullman, 2001). The CFA model (Model 3) in which Affective Suffering loadings were constrained to be equal between centres while Motivation factor loadings were estimated freely provided nearly as good a fit as the model with no constraints, and a better fit than the model in which loadings for both factors were constrained. Therefore measurement invariance was present for the Affective Suffering, but not for the Motivation factor.

The EURO-D hierarchical structure

A second objective was to investigate the ranking of individual items using Rasch modelling and to clarify the extent to which this ranking varied across countries. In all countries the 12 EURO-D items conformed well to a Rasch model, with the INFIT indices in particular suggesting that it could be regarded as an adequate uni-dimensional scale with hierarchical scaling properties. Items loading on Affective Suffering fell into two groups, those with high item calibrations (guilt, suicidality and appetite) and those with low calibrations (irritability, tearfulness, fatigue, sleep and depression), with little variability between countries. Items loading on Motivation calibrations varied markedly between countries, interest, pessimism and enjoyment generally occupied the mid-range between the two sets of Affective Suffering items with concentration falling into the lower calibration group. Therefore, to the extent that EURO-D can still be regarded as a uni-dimensional scale measuring a single underlying trait, the motivation items may provide useful additional discriminating power in the mid-range of trait scores.

Clinical and research implications

The Affective Suffering factor shows excellent cross-cultural measurement properties with strong evidence for measurement invariance with respect both to item loadings onto a common underlying latent trait, and to item calibrations. It has strong face validity as a depression measure. For future use, Affective Suffering factor scores can be calculated from the culturally invariant factor loadings in Model 3 in Table 3. A simple Affective Suffering subscale could be constructed from the seven main items that load on the factor (depression, guilt, suicidality, irritability, tearfulness, fatigue and sleep), however discrimination may be poor over the mid-range unless further appropriate items with middling calibrations can be identified. While the motivation factor emerges as a clear second factor in all countries, it exhibits considerable heterogeneity in factor loading patterns and item calibrations between countries. It cannot therefore be considered to be sufficiently well characterized or stable for use in cross-cultural research. Efforts should first be concentrated on clarifying the construct and its clinical and predictive validity. Concerns were raised regarding its clinical salience in our earlier paper (Prince et al., 1999b). Motivation factor scores, but not Affective Suffering scores, increased with increasing age. Did this correlation, considered in the context of the orthogonal relationship with Affective Suffering suggest that an expressed lack of interest and enjoyment, together with pessimism might be affectively neutral statements representing an adaptive cognitive appraisal of activity limitation in older age (Prince et al., 1999a)? Alternatively, might this factor represent some of the clinical features of vascular depression (Alexopoulos et al., 1997) with impaired executive functioning associated with vascular damage to subcortical structures? Future research might usefully test whether disability is associated with motivation but not with affective suffering, and whether motivation is specifically associated with cognitive impairment, particularly tests of executive function such as verbal fluency. It would also be important to investigate further the stability or otherwise of the measurement properties of the affective suffering and, particularly, the motivation factors across the life stages in later life. In the SHARE study, participants' ages ranged from 55 to extreme old age; we could test for measurement invariance with respect to CFA and Rasch models, comparing explicitly different age groups. Age might also be a useful selection factor in future qualitative research, recruiting younger-old and oldest-old people with high motivation factor scores in an attempt to understand the meaning of lack of motivation at different ages. Qualitative research involving older people with high Motivation factor scores accompanied by high and low Affective Suffering scores may also help to clarify the underlying construct or constructs, and to generate additional items that could be used to improve the psychometric characteristics of

DOI: 10.1002/mpr

the subscale, should this be desirable. In the meantime, we propose to use both the full 12-item EURO-D scale, and the Affective Suffering factor score derived from it in our further analyses of SHARE data and would commend this approach to others using this publicly accessible data resource.

Despite the measurement invariance demonstrated for the affective suffering factor we cannot conclude that the between country differences in levels of this trait described in Table 1 relate to real differences in psychological morbidity. Compositional differences (for example in age, gender or educational level) may have contributed. More significantly, culturally determined differences in norms or expectations or expressions of mood and mental health might be implicated. Such influences have been clearly established in the case of self-reported global health, where the use of anchoring vignettes has been advocated to identify and adjust for the consequent response bias (Salomon et al., 2004). In principle, a similar approach might improve the cross-cultural comparability of self-reported mental health assessments.

Acknowledgements

The SHARE data collection has been primarily funded by the European Commission through the 5th framework programme (project QLK6-CT-2001-00360 in the thematic programme Quality of Life). Additional funding came from the US National Institute on Ageing (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, Y1-AG-4553-01 and OGHA 04-064). Data collection in Austria (through the Austrian Science Fund, FWF), Belgium (through the Belgian Science Policy Office) and Switzerland (through BBW/OFES/UFES) was nationally funded. E. Castro-Costa is funded by the Social Psychiatry Research Trust for running this analysis.

References

Akaike H. Factor analysis and AIC. Psychometrika 1987; 52: 317–32.

Alexopoulos GS, Meyers BS, Young RC. Vascular depression hypothesis. Arch Gen Psychiatry 1997; 54: 915–22.

Andersen EB. A goodness of fit for the Rasch model. Psychometrika 1973; 38(1): 123–40.

Arbuckle JL. AMOS 5.0 update to the Amo's user's guide SPSS, Chicago, IL; SPSS, 2003.

Beekman ATF, Copeland JRM, Prince MJ. Review of the community prevalence of depression in later life. Br J Psychiatry 1999; 174: 307–11.

Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariances structures. Psychol Bull 1980; 88: 588–606.

Blazer DG. Depression in late life: review and commentary. J Gerontol A Biol Sci Med Sci 2003; 58A(3): 249–65.

Bollen KA, Long JS. Testing Structural Equation Models. Newbury Park, CA: Sage, 1993.

Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. Mahwah, NJ: Lawrence Erlbaum Associates, 2001.

Borsch-Supan A, Brugiavini A, Jurges H, Mackenbach J, Siegrist J, Weber G. Health, Ageing and Retirement in Europe – First Results from the Survey of Health, Ageing and Retirement in Europe. Manheim: MEA, 2005.

Browne MW. MUTMUM PC: user's guide Columbus, OH: Ohio State University, 1990.

Burnham KP, Anderson DR. Model Selection and Inference: A Practical Information – Theoretic Approach New York: Springer-Verlag, 1998.

Copeland JR, Dewey ME, Griffiths-Jones HM. Computerised pscychiatric diagnostic system and case nomenclature for elderly subjects: GMS and AGECAT. Psychol Med 1986; 16: 89–99.

Dunn G, Everitt B, Pickles A. Modelling Covariances and Latent Variables using EQS, 1st edition. London: Chapman & Hall, 1993.

English Longitudinal Study of Ageing (ELSA). Available: http://www.natcen.ac.uk/elsa/, 2006. Assessed 6 October 2006.

Hardouin JB. Rasch analysis: estimation and tests with the Raschtest module. The Stata J 2001; 1(1): 1–20.

Joreskog KG, Sorbom D. PRELIS. A Program for Multivariate Data Screening and Data Summariztion. User's Guide, 2nd edition. Chicago, IL: Scientific Software International, 1988.

Kline RB. Principles and Practice of Structural Equation Modelling. New York: Guilford Press, 1998.

Larraga L, Saz P, Dewey ME, Marcos G, Lobo A, ZARADEMP Workgroup. Validation of the Spanish version of the EURO-D scale: an instrument for detecting depression in older people. Int J Geriatr Psychiatry 2006; (12): 1199-205.

MacCallum RC. The need for alternative measures of fit in covariance structure modelling. Multivariate Behav Res 1990; 25: 157–62.

Marsh HW, Balla JR, Hau KT. An evaluation of incremental fit indices: a clarification of mathematical and empiracal properties. In Marcoulides GA, Schumacker RE (eds) Advanced Structural Equation Modelling: Issues and Techniques. Mahwah, NJ: Lawrence Erlbaum Associates, 1996, pp. 315–55.

Nunnally JC, Bernstein IH. Psychometric Theory, 3rd edition. New York: McGraw-Hill, 1994.

Olsson U. Maximum likelihood estimation of the polychoric correlation coefficient. Psychometrika 1979; 44: 443–60.

Prince M, Acosta D, Chiu H, Copeland J, Dewey M, Scazufca M, Varghese M, 10/66 Dementia Research Group. Effects of education and culture on the validity of the Geriatric

Mental State and its AGECAT algorithm. Br J Psychiatry 2004; Nov(185): 429–36.
Prince M, Beekman ATF, Deeg DJH, Fuhrer R, Kivela SL, Lawlor B, Lobo A, Magnusson H, Meller I, Van Oyen H, Reischies F, Roelands M, Skoog I, Turrina C, Copeland JR. Depression symptoms in late life assessed using the EURO-scale. Br J Psychiatry 1999a; 174: 339–45.
Prince M, Reischies F, Beekman ATF, Fuhrer R, Jonker C, Kivela SL, Lawlor B, Lobo A, Magnusson H, Fichter MM, Van Oyen H, Roelands M, Skoog I, Turrina C, Copeland JR. Development of the EURO-D scale – a European Union initiative to compare symptoms of depression in 14 European centres. Br J Psychiatry 1999b; 174: 330–8.
Reise SP, Widaman KF, Pugh RH. Confirmatory Factor Analysis and Item Response Theory: two approaches for exploring measurement invariance. Psychol Bull 1993; 114(3): 552–66.
Salomon JA, Tandon A, Murray CJL, World Health Survey Pilot Study Collaborating Group. Comparability of self rated health: cross sectional multi-country survey using anchoring vignettes. BMJ 2004; 328(333).
Sorbom D. A general method for studying differences in factor means and factor structure between groups. Br J Matha Stat Psychol 1974; 27: 229–39.
Tucker L, Lewis C. A reliability coefficient for maximum likelihood factor analysis. Psychometrika 1973; 38: 1–10.
Ullman JB. Structural equation modelling. In Tabachnick BG, Fidell LS (eds) Using Multivariate Statistics. Needham Heights, MA: Allyn & Bacon, 2001.
US Health and Retirement Study (HRS). Available: http://www.hrsonline.hrs.umich.edu, 2006. Assessed 6 October 2006.

Appendix

The EURO-D scale

Question 1: SAD OR DEPRESSED LAST MONTH
'In the last month, have you been sad or depressed?'
0 No
1 Yes

Question 2: HOPES FOR THE FUTURE
'What are your hopes for the future?'
0 Any hopes mentioned
1 No hopes mentioned

Question 3: FELT WOULD RATHER BE DEAD
'In the last month, have you felt that you would rather be dead?'
0 No such feelings
1 Any mention of suicidal feelings or wishing to be dead

Question 4: FEELS GUILTY
'Do you tend to blame yourself or feel guilty about anything?'
0 No such feelings
1 Obvious excessive guilt or self-blame, mentions guilt or self-blame, but it is unclear if these constitute obvious, or excessive guilt or self-blame

Question 5: TROUBLE SLEEPING
'Have you had trouble sleeping recently?'
0 No trouble sleeping
1 Trouble with sleep or recent change in pattern

Question 6: LESS OR SAME INTEREST IN THINGS
'In the last month, what is your interest in things?'
0 No mention of loss of interest, non-specific or uncodeable response
1 Less interest than usual mentioned

Question 7: IRRITABILITY
'Have you been irritable recently?'
0 No
1 Yes

Question 8: APPETITE
'What has your appetite been like?'
0 No diminution in desire for food, non-specific or uncodeable response
1 Diminution in desire for food

Question 9: FATIGUE
'In the last month, have you had too little energy to do the things you wanted to do?'
0 No
1 Yes

Question 10: CONCENTRATION
'How is your concentration?' (Difficulty in concentrating on entertainment or reading)
1 Difficulty in concentrating on entertainment
2 No such difficulty mentioned

Question 11: ENJOYMENT
'What have you enjoyed doing recently?'
0 Mentions any enjoyment from activity
1 Fails to mention any enjoyable activity

Question 12: TEARFULNESS
'In the last month, have you cried at all?'
0 No
1 Yes

Correspondence: Erico Castro-Costa, Section of Epidemiology, PO Box 060, Institute of Psychiatry, De Crespigny Park, London, SE5 8AF, UK.
Telephone (+44) (0)20 7848 0341
Fax (+44) (0)20 7277 0283
Email: dacosta.bhe@terra.com.br