Measuring disability across cultures – the psychometric properties of the WHODAS II in older people from seven low- and middle-income countries. The 10/66 Dementia Research Group population-based survey

RENATA M. SOUSA,[1] MICHAEL E. DEWEY,[1] DAISY ACOSTA,[2] JOTHEESWARAN A.T.,[3] ERICO CASTRO-COSTA,[1] CLEUSA P. FERRI,[1] MARIELLA GUERRA,[4] YUEQIN HUANG,[5] K.S. JACOB,[6] JUANA GUILLERMINA RODRIGUEZ PICHARDO,[7] NAYELI GARCIA RAMÍREZ,[8] JUAN LLIBRE RODRIGUEZ,[9] MARINA CALVO RODRIGUEZ,[9] AQUILES SALAS,[10] ANA LUISA SOSA,[8] JOSEPH WILLIAMS[11] & MARTIN J. PRINCE[1]

1 Institute of Psychiatry, King's College London, Health Services & Population Research Department, London, UK
2 Universidad Nacional Pedro Henriquez Ureña (UNPHU), Internal Medicine Department, Geriatric Section, Santo Domingo, Dominican Republic
3 Institute of Community Health, Voluntary Health Services, Chennai, India
4 Psychogeriatric Unit, National Institute of Mental Health 'Honorio Delgado Hideyo Noguchi', Lima, Peru
5 Peking University, Institute of Mental Health, Beijing, China
6 Christian Medical College, Vellore, India
7 Colegio Dominicano de Estadisticos y Demografos (CODE), Santo Domingo, Dominican Republic
8 National Institute of Neurology and Neurosurgery of Mexico, National University Autonomous of Mexico, Mexico City, Mexico
9 Facultad de Medicina Finlay-Albarran, Medical University of Havana, Havana, Cuba
10 Medicine Department, Caracas University Hospital, Faculty of Medicine, Universidad Central de Venezuela, Caracas, Venezuela
11 Department of Community Health, Voluntary Health Services, Chennai, India

Correspondence
Renata M. Sousa, King's
College London, Institute of
Psychiatry, Health Services &
Population Research
Department, De Crespigny Park,
PO Box 60, SE5 8AF, London,
UK.
Telephone (+44) 20 7848 0177
Fax (+44) 20 7848 5056
Email: r.sousa@iop.kcl.ac.uk

Abstract

We evaluated the psychometric properties of the 12-item interviewer-administered screener version of the World Health Organization Disability Assessment Schedule – version II (WHODAS II) among older people living in seven low- and middle-income countries. Principal component analysis (PCA), confirmatory factor analysis (CFA) and Mokken analyses were carried out to test for unidimensionality, hierarchical structure, and measurement invariance across 10/66 Dementia Research Group sites.

PCA generated a one-factor solution in most sites. In CFA, the two-factor solution generated in Dominican Republic fitted better for all sites other than rural China. The two factors were not easily interpretable, and may have been an artefact of differing item difficulties. Strong internal consistency and high factor loadings for the one-factor solution supported unidimensionality. Furthermore, the WHODAS II was found to be a 'strong' Mokken scale. Measurement invariance was supported by the similarity of factor loadings across sites, and by the high between-site correlations in item difficulties.

The Mokken results strongly support that the WHODAS II 12-item screener is a unidimensional and hierarchical scale confirming to item response theory (IRT) principles, at least at the monotone homogeneity model level. More work is needed to assess the generalizability of our findings to different populations. *Copyright © 2010 John Wiley & Sons, Ltd.*

Introduction

The number of people aged 60 years and over is estimated to grow worldwide to two billion by 2050 (United Nations, 2007). By 2025, approximately three-quarters will be living in low- and middle-income countries (LAMIC) (United Nations, 2007). The accompanying health transition will see a rapid increase in the burden arising from age-related chronic non-communicable diseases, mainly because of their contribution to years lived with disability (Murray and Lopez, 1996). There is a need for convenient, feasible and valid measures of disability, for population health surveys, health service evaluation and routine clinical practice. The applicability of such measures to older people, and to those living in LAMIC will also be of increasing relevance.

The World Health Organization's (WHO's) International Classification of Functioning, Disability and Health (ICF) defines disability as 'the negative aspects of the interaction between an individual (with a health condition) and that individual's contextual factors (personal and environmental factors)' (WHO, 2001a). Interactions include: impairments (affecting the body); activity limitations (affecting actions or behaviour); participation restrictions (affecting experience of life). The WHO

Disability Assessment Schedule, version II (WHODAS II) was developed to be consistent with the ICF classification, and to identify the consequences of any type of disorder that has an impact on everyday functioning, treating all disorders at parity when determining level of functioning (Chopra *et al.*, 2004). It was developed, field-tested and validated in 16 languages in 14 different countries (WHO, 2001b).

The psychometric properties of the 36-item WHODAS II have been explored in a variety of clinical populations including those with stroke (Posl *et al.*, 2007), inflammatory arthritis (Baron *et al.*, 2008), back pain (Chwastiak and Von Korff, 2003), ankylosing spondylitis (van Tubergen *et al.*, 2003), systemic sclerosis (Hudson *et al.*, 2008), acquired hearing loss (Chisolm *et al.*, 2005), psychosis (Chopra *et al.*, 2004; McKibbin *et al.*, 2004), depression (Chwastiak and Von Korff, 2003), and among mental health service users (Chavez *et al.*, 2005). It performed well in all of these contexts, with high internal consistency, moderate to good test–retest reliability, and good concurrent validity against indicators of disease severity, disease specific and other generic disability assessments. It seems to be at least as responsive as other generic measures to clinical change in anxiety (Perini *et al.*, 2006), depression (Chwastiak and Von Korff, 2003;

Mogga *et al.*, 2006) and back pain (Chwastiak and Von Korff, 2003).

The 36-item WHODAS II has previously been used in two population-based studies, the World Mental Health (WMH) Surveys of adults across 16 countries (Von Korff *et al.*, 2008) and the Kwangju community survey of physical and psychiatric morbidity among older adults in South Korea (Kim *et al.*, 2005). Findings from these surveys again supported internal consistency and concurrent validity; in the WMH surveys WHODAS II scores were consistently correlated with the Sheehan Disability Scale (Von Korff *et al.*, 2008). The Sheehan Disability Scale (SDS) is a self-report measure of levels of mental-health related functional impairment in primary care settings. Patients are asked to respond in a 10-point visual analogue scale how much their symptoms interfered with three domains of life (work, social life, and family life). SDS has shown high internal consistency reliability and good construct validity (Leon *et al.*, 1997). In Korea, physical health, depression and cognitive function explained 40% of the variance in WHODAS II scores and effects of socio-demographic variables were no longer apparent after controlling for these health outcomes.

The WHO website reports a clear unidimensional structure for the 36-item WHODAS II with very high loadings of all six domain scores on a Global Disability latent variable (WHO, 2001b). For the WMH survey, confirmatory factor analysis (CFA) provided only relatively weak support for unidimensionality in the four countries in which this was carried out (Von Korff *et al.*, 2008). However, the domain subscale scores all loaded >0.40 on a Global Disability latent variable and, to this extent the utility of a Global Disability score was supported.

The shorter 12-item 'screener' version of the WHODAS II has been little used to date (Norton *et al.*, 2004; Rehm *et al.*, 1999). This is surprising, since it takes only five minutes to administer, and covers all six domains of the full WHODAS II (Rehm *et al.*, 1999). In the WHO pilot studies, the correlation between the score from the screener and the score of the WHODAS II was 0.95, meaning that the screener explained more than 90% of the total variation of the full 36-item WHODAS II (Rehm *et al.*, 1999). CFA indicated a unidimensional scale with good classical scaling properties. However, its lack of compatibility with item response theory (IRT) was considered to limit its cross-cultural applicability (Rehm *et al.*, 1999). Accordingly, five of the 12 items were subsequently replaced with others from the 36-item version to improve its IRT characteristics, resulting in the version currently approved by the WHO. Changes were made in the following domains: (1) understanding and communication – items addressing difficulties in understanding and remembering were replaced with items on difficulties in concentrating and learning, (2) self-care – the item on difficulties with feeding was substituted by an item on difficulties with dressing, and (3) participation in society – items addressing difficulties in carrying out plans and in living in dignity were replaced with items addressing difficulties in joining community activities and with being emotionally affected.

Our aim in the current analysis was to explore the psychometric properties of the 12-item screener version of the WHODAS II in an epidemiological survey of older adults across a wide variety of LAMIC settings. Specifically, we wished to examine whether the 12-item WHODAS II would meet criteria for measurement invariance across cultures, a requirement for international comparative research, as well as to assess its factor structure and unidimensionality. The 10/66 Dementia Research Group (10/66 DRG) studies in 11 sites in seven LAMIC provide an opportunity to address these aims, at the same time redressing the imbalance in previous research towards younger participants in high income countries.

Methods

Design

A secondary analysis of data from the 10/66 DRG surveys of representative samples of older people in seven developing countries (urban sites in Cuba, Dominican Republic and Venezuela, and rural and urban sites in Mexico, Peru, China and India) were carried out. Full details of the study protocol can be found elsewhere (Llibre Rodriguez *et al.*, 2008; Prince *et al.* 2008a). Briefly, a cross-sectional one phase survey was carried out in geographically defined catchment areas. All residents aged 65 years and over were included in the survey and an informant was also interviewed. The sample size for each country was between 2000 and 3000 participants. All studies were approved by local ethical committees and by the King's College London ethical committee.

Disability assessment

A copy of the 12-item interview-administered WHODAS II screener assessment is provided in the Appendix [this and the full 36-item version are also available on the WHODAS II website (WHO, 2001b). Both versions cover six domains encompassing: understanding and communicating with the world; moving and getting around; self care; getting along with people; life activities; and participation in society. Scores for each question range from

zero (no difficulty) to four (extreme difficulty/can not do). The standardized global score ranges from zero (non-disabled) to 100 (maximum disability).

Statistical analysis

For each site, we estimated the mean WHODAS II global disability score, the proportion with non-zero scores, and the 90th centile. We also calculated the mean age, and the proportion of participants reporting three or more chronic limiting physical impairments.

Principal component analysis (PCA) of WHODAS II items was based on covariance matrix of polychoric correlations used for analysis of ordinal variables (Joreskog and Sorbom, 1993; Joreskog, 1994). The cut off used to assume that an item loaded on a given factor was 0.60. A varimax rotation was carried out with an eigenvalue of one as initial extraction criterion (Castro-Costa *et al.*, 2008). Given the *a priori* hypothesis of unidimensionality (Rehm *et al.*, 1999), we then tested and compared between sites the goodness-of-fit of a one-factor solution, using CFA.

CFA and Mokken scaling analysis can be used to assess both the dimensionality of a scale, and its ability to measure the same latent trait in the same way across different groups, in our case older people from different countries and cultures. This property has been referred to as psychometric equivalence, distinct from functional and conceptual equivalence, which are a *sine qua non* of cross-cultural comparability. CFA models contain parameters that are (a) fixed to a certain value, (b) constrained to be equal to other parameters, and (c) free to take on any unknown value (Castro-Costa *et al.*, 2008). In testing for psychometric invariance, two models are fitted and then compared for goodness-of-fit; one in which the factor loadings are unconstrained that is estimated separately for all sites, and the second in which they are constrained to be equal across sites, the null hypothesis being that items load to a similar extent on the same latent trait or traits across sites. Markedly superior fit of the first model would challenge the hypothesis of measurement invariance. Chi square statistics can be used to evaluate the absolute fit for each model tested in CFA, but in large data sets trivial differences in goodness-of-fit may be highly statistically significant. Therefore it is recommended to use other absolute and relative indices: the Akaike's Information Criterion (AIC: Akaike, 1987) – 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) – values near 1.0 indicate good fit and those greater than 0.90 are considered

satisfactory (Dunn *et al.*, 1993; Marsh *et al.*, 1996); and the root mean square error of approximation (RMSEA) (Browne, 1990) – values of less than 0.05 indicate close fit and 0.05 to 0.08 reasonable fit for the model.

Mokken scaling involves the application of a non-parametric item response model (Mokken, 1971) to measure the hierarchical properties of items in a scale, assessing if the items can be ordered by degree of difficulty, so that any individual who endorses a particular item will also endorse all the items ranked lower in difficulty. Three basic assumptions are required for a monotone homogeneity model (MHM): (1) unidimensionality (one latent variable summarizes the variation in the item scores in the questionnaire), (2) local independence (after conditioning on the position on the latent trait, the item scores are statistically independent), and (3) monotonicity (for all items the probability of a positive response increases monotonically with increasing values of the latent trait). These assumptions being met, an individual's position on the latent trait can conveniently be estimated as the rank of the highest item in the hierarchy that they endorse, or their total number of positive responses (Dijkstra *et al.*, 1999). Double monotonicity models (DMMs) require in addition that for any value of the latent trait, the probability of a positive response decreases with the difficulty of the item. This means that the order of item difficulties remains invariant over all values of the latent trait and thus, that the item response function curves do not intersect (Sijtsma *et al.*, 2008; Van der Ark *et al.*, 2007). To assess single monotonicity, we estimated Loevinger coefficients for each item (Hi) and for the whole scale (H), where values between 0.3 and 0.4 suggest weak scalability, values between 0.4 and 0.5 moderate, and values above 0.5 strong scalability. We also tested formally for violations of monotonicity (using the Stata loevH monotonicity command) and non-intersection (using the Stata loevH nipmatrix command) between pairs of items (minimum violation 0.03, alpha = 0.05), using overall criteria values as an indication of the likelihood of assumption violation; ≤40 'satisfactory', 40 to 79 'questionable violation', 80 and over 'strongly suggesting an assumption violation' (Molenaar and Sijtsma, 2000).

To establish the agreement between the difficulties (the proportions of people falling above each of the four cut-points) for each of the 12 items, we formed the correlation matrix separately for each cut-point and then pooled them using Fisher's hyperbolic arctangent transformation – performing the correlation on all four cut-points simultaneously would fail to take account of the fact that difficulties within items must be non-decreasing.

Overall agreement between the 11 sites was assessed with an intraclass correlation coefficient.

Descriptive statistics and PCA were conducted with STATA 10.0 (Stata Corporation, 2007). Mokken model was calculated in STATA 10.0 after downloading the LoevH add-on program from http://www.anaqol.org. Intraclass correlation coefficient was calculated in SPSS 15.0 (SPSS Inc., 2005), CFA was carried out with AMOS 5 (SPSS Inc., 2003), and Fisher's correlation was calculated in R (R Development Core Team, 2007).

Results

Sample characteristics

The 10/66 Release 1.7 dataset was used in this analysis. Response rates for the survey were excellent varying from 72% in urban India to 95% in Dominican Republic (Llibre Rodriguez *et al.*, 2008).

The 12 items contributing to the WHODAS global disability score were highly internally consistent with Cronbach's alpha varying between 0.90 and 0.97 by site (Table 1). There was considerable variation in the distribution of WHODAS II scores between sites (Figure 1 and Table 1). However, most of this was accounted for by two outlier sites, urban China with a very low proportion of non-zero scores (22.2%) and rural India, with a very high proportion (97.7%). There were also marked compositional differences between the sites in their age distributions (younger mean age in rural China, and in the Indian sites), and in the proportion of respondents reporting three or more limiting physical impairments (high proportions in Dominican Republic and Venezuela, and low proportions in Cuba, rural Peru, rural China and urban India).

Principal component analyses (PCA)

The PCA gave rise to a one-factor solution in most countries, with the exception of Cuba, the Dominican Republic, rural China and rural India, where a two-factor solution was generated. The second factor varied between centres, constituting the getting around, participation in society, self care and life activities domains in Cuba; the self care and getting along with people domains in the Dominican Republic; the getting around and life activities domains and the learning item (understanding and concentration) in rural India and the standing (getting
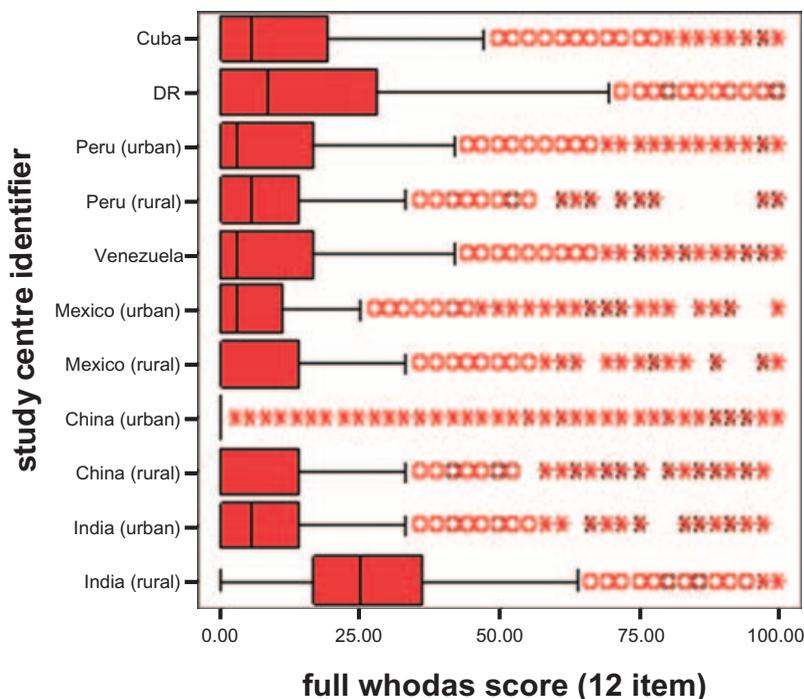

Figure 1 Distribution of WHODAS II scores by study site (box plot): o – outlier (more than one and a half box lengths above the 75th centile); x – extreme value (more than three box lengths above the 75th centile).

Table 1 The 10/66 DRG release 1.7 – socio-demographic characteristics and the WHODAS II mean total scores ($n = 14\,991$)

	Cuba (N = 2937)	Dominican Republic (N = 2001)	Urban Peru (N = 1380)	Rural Peru (N = 552)	Venezuela (N = 1947)	Urban Mexico (N = 1002)	Rural Mexico (N = 1000)	Urban China (N = 1160)	Rural China (N = 1002)	Urban India (N = 1001)	Rural India (N = 999)
WHODAS score distribution											
WHODAS mean scores (SD)	13.4 (20.0)	16.5 (20.3)	13.0 (20.6)	10.4 (14.5)	10.7 (16.4)	10.0 (17.3)	11.1 (19.1)	8.1 (20.1)	8.0 (14.5)	10.5 (15.4)	28.3 (18.3)
Percentage with non-zero scores	62.2%	68.6%	59.6%	66.4%	58.6%	51.7%	48.7%	24.3%	43.6%	52.3%	97.7%
Mean scores (SD) omitting zeros	21.5 (21.7)	24.0 (20.5)	21.9 (22.7)	15.7 (15.4)	18.3 (17.8)	19.3 (19.9)	22.7 (22.0)	33.2 (28.9)	18.7 (17.2)	16.9 (16.6)	28.9 (17.9)
90th Centile	38.9	47.2	41.7	27.8	30.6	30.6	36.1	33.1	22.2	30.0	50.0
Cronbach's alpha	0.93	0.90	0.95	0.91	0.92	0.92	0.93	0.97	0.91	0.93	0.95
Sample characteristics											
Mean age (SD)	75.1 (7.0)	75.2 (7.5)	75.0 (7.4)	74.2 (7.3)	72.5 (6.9)	74.5 (6.6)	74.1 (6.7)	73.9 (6.1)	72.4 (6.0)	71.4 (6.1)	72.6 (5.8)
Three or more limiting impairments, number (%)	292 (9.9)	465 (23.1)	224 (16.2)	40 (7.2)	486 (25.4)	158 (15.8)	185 (18.5)	208 (17.9)	39 (3.9)	41 (4.1)	168 (16.9)
Response proportion (%)	94	95	80	88	80	84	86	74	96	72	98

Table 2 Factor structure and loadings[a] from principal component analysis based on polychoric correlation matrix for WHODAS II

Site	Cuba	Dominican Republic	Urban Peru	Rural Peru	Venezuela	Urban Mexico	Rural Mexico	Urban China	Rural China	Urban India	Rural India	Pooled
Items						Factors						
U&C Concentration	1	1[c]	1	1	1	1	1	1	1	1	1	1
U&C Learning	1	1	1	1	1	1	1	1	2	1	2[b]	1
Getting around Standing	2	1	1	1[b]	1	1	1	1	2[b]	1	2	1
Getting around Walking	2	1	1	1	1	1	1	1	1	1	2	1
Self-care washing	1/2[b]	2[b]	1	1	1	1	1	1	1	1	1	1
Self-care dressing	1/2[b]	2[b]	1	1	1	1	1	1	1	1	1	1
Getting along with people Dealing	1	2	1	1	1	1	1	1	1	1	1	1
Getting along with people Friendship	1	2	1	1	1	1	1	1	1	1	1	1
Life activities Household	2	1	1	1	1	1	1	1	2	1	2	1
Life activities Work	2[b]	1	1	1	1	1	1	1	1	1	2	1
Participation in society Community	2[b]	1	1	1	1	1	1	1	1	1	1[b]	1
Participation in society Emotionally	2	1	1	1	1	1	1	1	1	1	1/2[b]	1
Factor 1 Eigenvalue	9.08	7.69	9.43	8.29	9.02	8.87	9.65	10.94	9.76	9.08	8.83	9.01
Factor 1 Variance (%)	43.7	44.6	78.6	69.1	75.2	73.9	80.4	91.1	60.2	75.6	46.0	75.1
Factor 2 Eigenvalue	1.02	1.02	–	–	–	–	–	–	1.11	–	1.08	–
Factor 2 Variance (%)	40.4	28.0	–	–	–	–	–	–	30.4	–	36.6	–

[a] One factor solutions were generated in all sites other than Cuba, Dominican Republic, rural China and rural India, which generated two factor solutions. Value 1 indicates items loading on factor one, value 2 indicates items loading on a second factor (where one was generated). value 1/2 indicates items cross-loading on both factors. Item loadings are signified as follows – all items loading at ≥0.70, other than those marked.
[b] Loaded at 0.60–0.69.
[c] Loaded at 0.50–0.59.

around), household tasks (life activities) and learning (understanding and concentration) items in rural China. In sites where a one-factor solution was found, this single factor explained between 69.1% and 91.1% of the variance with eigenvalues ranging from 8.29 to 10.94. In sites that generated a two-factor solution (Cuba, Dominican Republic, rural China and rural India) the first factor explained between 43.7% and 60.2% of the variance (eigenvalues between 7.69 and 9.76) and the second factor 28% to 40.4% (eigenvalues between 1.02 and 1.11) (Table 2). When data were pooled from all sites a one-factor

solution emerged, accounting for 75.1% of the variance (eigenvalue = 9.01).

Confirmatory factor analysis (CFA)

We next tested the goodness-of-fit, in each site, of the one-factor solution, and the different two-factor solutions arising from the PCA, using CFA. Upon inspection of the four two-factor solutions, it was clear that that arising from the Dominican Republic fitted best across all sites, other than in rural China where the locally generated

Table 3 Goodness-of-fit within centres between 10/66 sites

Site	One factor solution				Two factor solution					
	χ^2	TLI	RMSEA	AIC	χ^2	TLI	RMSEA	AIC	χ^2 change	*p* Values
Cuba	4215.9	0.71	0.16	4263.9	1990.4	0.86	0.11	2040.4	2225.5	<0.001
Dominican Republic	1738.4	0.77	0.12	1786.4	898.1	0.88	0.09	948.1	840.3	<0.001
Urban Peru	1275.7	0.84	0.13	1323.7	773.1	0.90	0.10	823.1	502.6	<0.001
Rural Peru	1003.6	0.66	0.18	1051.6	778.7	0.73	0.16	828.7	224.9	<0.001
Venezuela	2151.1	0.76	0.14	2199.1	1066.2	0.88	0.10	1116.2	1084.9	<0.001
Urban Mexico	947.7	0.78	0.13	995.7	606.1	0.86	0.10	656.1	341.6	<0.001
Rural Mexico	2300.7	0.68	0.20	2348.7	1033.9	0.86	0.14	1083.3	1266.8	<0.001
Urban China	2444.0	0.79	0.19	2492.0	2190.0	0.81	0.18	2240.0	254.0	<0.001
Rural China	3142.8	0.54	0.24	3190.8	2912.7	0.57	0.23	2962.7	230.1	<0.001
Urban India	1413.1	0.73	0.16	1461.1	1199.4	0.76	0.15	1249.4	213.7	<0.001
Rural India	2292.8	0.49	0.20	2340.8	946.2	0.79	0.13	996.2	1346.6	<0.001

two-factor solution provided a better fit. The best fitting (Dominican Republic) two-factor solution was then compared with the one-factor solution (Table 3). The one-factor solution provided only a moderately poor fit with RMSEA varying between 0.13 and 0.24 and TLI varying between 0.49 and 0.84. The two-factor solution provided a better fit in all sites with lower RMSEA (0.09 to 0.23), higher TLI (0.57 to 0.90) and lower AICs. Although the fit of the two-factor solution was generally better, TLI only exceeded 0.90 in one site, and was between 0.85 and 0.90 in a further five sites, all in Latin America.

Next we tested for measurement invariance for the one-factor (Table 4) and best-fitting two-factor solution (Table 5) across all sites with loadings not constrained (model 1) and constrained (model 2). In each case, there was little difference in the goodness-of-fit of the constrained and unconstrained models, although according to the AIC the unconstrained models fitted somewhat better [one-factor solution – AIC (unconstrained) = 23455.6, AIC (constrained) = 25855.4; two-factor solution – AIC (unconstrained) = 14946.3, AIC (constrained) = 17645.4]. Again, the overall fit of constrained and unconstrained models was much better for the two-factor solution than for the one-factor solution.

Mokken analysis

Item and scale Loevinger *H* coefficients were estimated separately for each site, using a polytomous Mokken analysis. There was robust evidence that the WHODAS II and its 12 constituents items conformed to a 'strong' Mokken scale (Table 6). The item scalability coefficients from the

Mokken analysis exceeded 0.40 for all items in all sites, and exceeded 0.50 in most cases. The coefficient *H* values for the scale as a whole varied between 0.52 and 0.81 by site. There were only two statistically significant violations of monotonicity, learning a new task in rural China and walking a kilometre in rural India. There were however a number of violations with respect to non-intersection (double monotonicity), all of which were statistically significant; and several of which were linked with overall criteria values >80, strongly suggesting an assumption violation. Certain items and sites seemed particularly implicated. Intersection violations were most apparent in Cuba, Dominican Republic, rural China, and rural India. With respect to items, intersection violations were most apparent for walking, standing, and learning a new task. The item difficulties (for the no difficulty/ some difficulty threshold) across the 11 sites are summarized in Table 6. Lower item difficulties (easy items) suggest a high probability of item endorsement by those with low scores on the trait. Standing (followed by walking) had the lowest item difficulty overall, while maintaining friendship and dressing had the highest. The rank order of item difficulty was similar in all sites. The correlation coefficients for the set of WHODAS II item difficulties between pairs of 10/66 sites ranged from 0.70 to 0.98, other than those that included rural China, which ranged between 0.43 and 0.61 (Table 7). Closer inspection of the pattern of item difficulties by site (Table 6) suggested some possibility of differential item functioning for the household responsibilities and learning a new task, both of which had relatively high levels of endorsement, particularly at the higher thresholds of difficulty.

Table 4 Confirmatory factor analysis – one-factor solution without constraints, and goodness-of-fit parameters with and without constraints

One-factor solution

Factor 1

Site	Standing	Household	Community	Emotional	Walking	Washing	Dressing	Work	Learning	Concentration	Dealing	Friendship
Cuba	0.58	0.75	0.75	0.50	0.61	0.74	0.71	0.77	0.54	0.56	0.55	0.64
Dominican Republic	0.63	0.74	0.67	0.63	0.64	0.54	0.52	0.75	0.61	0.62	0.39	0.50
Urban Peru	0.51	0.74	0.72	0.62	0.67	0.79	0.77	0.82	0.71	0.66	0.67	0.71
Rural Peru	0.41	0.69	0.69	0.59	0.66	0.68	0.67	0.72	0.67	0.70	0.62	0.67
Venezuela	0.60	0.80	0.77	0.68	0.62	0.63	0.60	0.75	0.73	0.65	0.50	0.52
Urban Mexico	0.66	0.75	0.69	0.66	0.65	0.62	0.60	0.75	0.59	0.59	0.50	0.48
Rural Mexico	0.66	0.75	0.80	0.72	0.64	0.78	0.76	0.70	0.71	0.72	0.72	0.72
Urban China	0.79	0.91	0.90	0.77	0.86	0.87	0.79	0.89	0.80	0.69	0.67	0.68
Rural China	0.48	0.43	0.77	0.74	0.70	0.80	0.73	0.80	0.45	0.72	0.65	0.65
Urban India	0.55	0.77	0.65	0.64	0.55	0.67	0.58	0.73	0.78	0.68	0.66	0.63
Rural India	0.19	0.40	0.41	0.31	0.21	0.55	0.44	0.39	0.41	0.51	0.95	0.93
Factor loading Means (SD)	0.55 (0.16)	0.70 (0.15)	0.71 (0.12)	0.62 (0.13)	0.62 (0.15)	0.70 (0.11)	0.65 (0.11)	0.73 (0.12)	0.64 (0.13)	0.64 (0.07)	0.62 (0.14)	0.65 (0.12)

	Without constraints	With constraints		
χ^2	22927.6	25547.4	χ^2 change	2619.8
df	594	704	df change	110
AIC	23455.6	25855.4	p-value	$P < 0.001$
TLI	0.72	0.74		
RMSEA	0.050	0.049		

Note: SD, standard deviation.

Table 5 Confirmatory factor analysis – two-factor solution without constraints, and goodness-of-fit parameters with and without constraints

	Two-factor solution											
	Factor 1								Factor 2			
	Learning	Concentration	Community	Emotional	Work	Household	Walking	Standing	Washing	Dressing	Dealing	Friendship
Cuba	0.53	0.52	0.75	0.55	0.80	0.81	0.66	0.63	0.93	0.90	0.52	0.62
Dominican Republic	0.61	0.61	0.68	0.65	0.76	0.76	0.64	0.64	0.79	0.79	0.47	0.53
Urban Peru	0.74	0.66	0.73	0.64	0.82	0.78	0.70	0.55	0.89	0.88	0.65	0.70
Rural Peru	0.74	0.72	0.73	0.62	0.68	0.74	0.65	0.42	0.85	0.88	0.59	0.59
Venezuela	0.73	0.63	0.78	0.69	0.75	0.82	0.64	0.62	0.88	0.88	0.50	0.48
Urban Mexico	0.59	0.58	0.69	0.67	0.75	0.77	0.68	0.69	0.82	0.83	0.50	0.46
Rural Mexico	0.74	0.73	0.82	0.75	0.68	0.82	0.69	0.72	0.94	0.94	0.74	0.71
Urban China	0.81	0.68	0.91	0.77	0.88	0.92	0.86	0.80	0.88	0.85	0.75	0.75
Rural China	0.50	0.70	0.79	0.75	0.80	0.48	0.73	0.53	0.75	0.77	0.79	0.77
Urban India	0.82	0.69	0.66	0.69	0.68	0.81	0.54	0.56	0.58	0.70	0.81	0.75
Rural India	0.73	0.37	0.64	0.60	0.66	0.75	0.56	0.52	0.53	0.43	0.97	0.93
Factor loading Means (SD)	0.62 (0.11)	0.63 (0.10)	0.74 (0.08)	0.67 (0.07)	0.75 (0.07)	0.77 (0.11)	0.67 (0.08)	0.61 (0.10)	0.80 (0.13)	0.80 (0.14)	0.57 (0.16)	0.66 (0.14)

	Without constraints	With constraints		
χ^2	14396.3	17295.4	χ^2 change	2899.1
df	583	683	df change	100
AIC	14946.3	17645.4	p-value	$P < 0.001$
TLI	0.82	0.82		
RMSEA	0.040	0.040		

Table 6 Polytomous Mokken analysis: Loevinger *H* values and item difficulties by 10/66 sites

10/66 Site		Cuba (n = 2944)	Dominican Republic (n = 2011)	Urban Peru (n = 1381)	Rural Peru (n = 552)	Venezuela (n = 1971)	Urban Mexico (n = 1003)	Rural Mexico (n = 1000)	Urban China (n = 1160)	Rural China (n = 1002)	Urban India (n = 1005)	Rural India (n = 999)	Coefficient *Hi* (pooled across sites)
WHODAS II items													
1. Standing	Coefficient *Hi*	0.58	0.52	0.60	0.41	0.59	0.62	0.67	0.76	0.69	0.61	0.71	0.61
	Item difficulty (rank)	0.66 (2)	0.63 (3)	0.63 (1)	0.55 (1)	0.61 (1)	0.68 (1)	0.69 (2)	0.85 (5)	0.67 (1)	0.54 (1)	0.06 (1)	
2. Household responsibilities	Coefficient *Hi*	0.72	0.55	0.69	0.59	0.70	0.65	0.68	0.85	0.69	0.67	0.78	0.67
	Item difficulty (rank)	0.69 (4)	0.68 (4)	0.76 (4)	0.73 (3)	0.73 (4)	0.78 (4)	0.76 (3)	0.86 (6)	0.69 (3)	0.77 (7)	0.25 (4)	
3. Learning a new task	Coefficient *Hi*	0.57	0.50	0.66	0.55	0.65	0.55	0.57	0.79	0.61	0.67	0.76	0.58
	Item difficulty (rank)	0.74 (5)	0.68 (5)	0.76 (5)	0.73 (4)	0.80 (7)	0.79 (5)	0.79 (5)	0.85 (3)	0.67 (2)	0.77 (6)	0.26 (5)	
4. Joining in community activity	Coefficient *Hi*	0.69	0.54	0.68	0.53	0.66	0.63	0.67	0.85	0.72	0.67	0.71	0.65
	Item difficulty (rank)	0.77 (7)	0.74 (7)	0.77 (7)	0.75 (5)	0.77 (5)	0.85 (7)	0.82 (6)	0.85 (4)	0.90 (6)	0.67 (3)	0.27 (7)	
5. Emotionally affected	Coefficient *Hi*	0.45	0.48	0.63	0.54	0.59	0.54	0.59	0.73	0.63	0.58	0.75	0.56
	Item difficulty (rank)	0.67 (3)	0.61 (2)	0.66 (2)	0.65 (2)	0.69 (3)	0.78 (3)	0.78 (4)	0.84 (1)	0.90 (5)	0.76 (5)	0.16 (3)	
6. Concentrating	Coefficient *Hi*	0.59	0.47	0.62	0.58	0.60	0.56	0.62	0.77	0.73	0.62	0.65	0.58
	Item difficulty (rank)	0.87 (8)	0.76 (8)	0.77 (6)	0.78 (7)	0.86 (8)	0.86 (8)	0.85 (8)	0.93 (10)	0.95 (9)	0.81 (8)	0.50 (8)	
7. Walking a kilometre	Coefficient *Hi*	0.66	0.57	0.66	0.55	0.63	0.63	0.67	0.83	0.67	0.61	0.63	0.64
	Item difficulty (rank)	0.60 (1)	0.58 (1)	0.69 (3)	0.76 (6)	0.65 (2)	0.70 (2)	0.68 (1)	0.84 (2)	0.83 (4)	0.56 (2)	0.09 (2)	
8. Washing whole body	Coefficient *Hi*	0.71	0.53	0.72	0.64	0.68	0.61	0.73	0.83	0.74	0.64	0.73	0.65
	Item difficulty (rank)	0.88 (9)	0.86 (9)	0.86 (10)	0.90 (11)	0.92 (10)	0.92 (9)	0.91 (9)	0.86 (7)	0.92 (7)	0.75 (4)	0.80 (11)	
9. Getting dressed	Coefficient *Hi*	0.72	0.55	0.72	0.60	0.67	0.63	0.74	0.82	0.79	0.65	0.77	0.68
	Item difficulty (rank)	0.89 (10)	0.89 (11)	0.88 (12)	0.91 (12)	0.92 (11)	0.94 (11)	0.92 (12)	0.91 (9)	0.96 (10)	0.94 (12)	0.89 (12)	
10. Dealing with people you don't know	Coefficient *Hi*	0.67	0.37	0.62	0.46	0.54	0.55	0.63	0.80	0.77	0.68	0.70	0.58
	Item difficulty (rank)	0.93 (12)	0.87 (10)	0.83 (9)	0.86 (10)	0.92 (9)	0.93 (10)	0.92 (10)	0.94 (12)	0.97 (12)	0.93 (11)	0.73 (9)	
11. Maintaining a friendship	Coefficient *Hi*	0.69	0.46	0.65	0.49	0.58	0.58	0.64	0.79	0.77	0.64	0.71	0.62
	Item difficulty (rank)	0.92 (11)	0.89 (12)	0.87 (11)	0.85 (9)	0.93 (12)	0.95 (12)	0.92 (11)	0.94 (11)	0.97 (11)	0.92 (10)	0.74 (10)	
12. Carrying out work and everyday activities	Coefficient *Hi*	0.70	0.57	0.73	0.62	0.67	0.65	0.64	0.84	0.74	0.72	0.73	0.68
	Item difficulty (rank)	0.75 (6)	0.72 (6)	0.81 (8)	0.81 (8)	0.79 (6)	0.85 (6)	0.82 (7)	0.87 (8)	0.92 (8)	0.88 (9)	0.27 (6)	
Whole scale	Coefficient *H*	0.64	0.52	0.66	0.55	0.63	0.60	0.65	0.81	0.69	0.64	0.72	0.63

Table 7 Between site correlation coefficients for WHODAS II after Fisher's hyperbolic arctangent transformation

	Cuba	Dominican Republic	Urban Peru	Rural Peru	Venezuela	Urban Mexico	Rural Mexico	Urban China	Rural China	Urban India	Rural India
Cuba	–										
Dominican Republic	0.94	–									
Urban Peru	0.82	0.86	–								
Rural Peru	0.87	0.88	0.90	–							
Venezuela	0.95	0.94	0.91	0.92	–						
Urban Mexico	0.94	0.95	0.90	0.91	0.98	–					
Rural Mexico	0.93	0.94	0.85	0.88	0.97	0.99	–				
Urban China	0.85	0.78	0.71	0.72	0.78	0.76	0.73	–			
Rural China	0.50	0.59	0.43	0.48	0.51	0.57	0.61	0.46	–		
Urban India	0.80	0.84	0.78	0.82	0.86	0.88	0.88	0.75	0.62	–	
Rural India	0.92	0.93	0.89	0.88	0.98	0.96	0.96	0.76	0.53	0.82	–

Overall, the intraclass correlation coefficient for consistency of item difficulty (location) across all 11 sites was 0.64 (95% confidence interval, 0.44–0.84).

Discussion

Little research has been carried out into the psychometric properties of the 12-item 'screening' version of WHODAS II. The original version was found to have unfavourable hierarchical scaling properties (Rehm *et al.*, 1999), but the revised version (WHO, 2001b), which we have used, has not previously been tested. The cross-cultural measurement properties of the WHODAS II have not previously been investigated in depth. In the modified 36 item version of the WHODAS II used in the WMH Survey, CFA (of a one-factor solution) could only be attempted in four countries (US, Israel, New Zealand and Ukraine) because of sample size limitations (Von Korff *et al.*, 2008). In a subsequent publication from the European Study of the Epidemiology of Mental Disorders (ESEMED) network of European WMH survey sites, cross-cultural measurement invariance between Mediterranean and non-Mediterranean countries was formally assessed using CFA (Buist-Bouwman *et al.*, 2008). However, because of the filter questions used in the WMH version of the WHODAS II IRT analyses were not applied to these data sets. Since filter questions were not used in the WHODAS II screener, we were not restricted in this way. Another strength of the current analysis was the sampling of older people, who were sparsely represented in the WHODAS development (Rehm *et al.*, 1999) and WMH survey data sets. Older people have a much higher prevalence of cognitive, physical and mental impairments, reflected in the much higher proportion of non-zero scores for the WHODAS II and its items. This strength is also a limitation, in that our findings may not be safely generalized to younger populations.

Is the WHODAS II a unidimensional scale?

In common with many other studies (Chisolm *et al.*, 2005; Chwastiak and Von Korff, 2003; McKibbin *et al.*, 2004) we found that the WHODAS II was a highly internally consistent scale, with Cronbach's alpha ranging from 0.90 to 0.97 by site. Rehm reports an excellent fit for a one-factor solution for his original version of the 12-item WHODAS II, with average factor loadings for the six domains onto a 'global disability' trait of 0.81 or 0.83 (Rehm *et al.*, 1999). A similar factor structure was reported from the ESEMED survey (Buist-Bouwman *et al.*, 2008), using a modified version of the full 36 item WHODAS II, particularly when the frequency items (not included in the 12-item WHODAS II) were omitted. However, Von Korff *et al.* (2008) in their analysis of the WMH survey data found that factor loadings at the dimension level were consistently lower, and concluded that 'the fit statistics did do not support the hypothesis that the five modified WHODAS domains form a unidimensional latent variable of Global Disability'. Our findings are more consistent with those of Von Korff. However, as with the WMH analysis, we found that, for all sites other than rural India, factor loadings for the one-factor solution all exceeded 0.40, hence based on classical test theory, a meaningful unidimensional scale could be constructed by summing items. Our *a priori* hypothesis was for a one-

factor solution, as supported by previous research. However, two-factor solutions emerged from PCA in Cuba, Dominican Republic, rural China and rural India, and one of these, from the Dominican Republic, clearly fitted much better across all sites in the subsequent CFA than did the one-factor solution. This grouped the getting around, life activities, understanding and communication, and participation in society domains in the first dominant factor and the self care and getting along with people domains in the second factor. The underlying constructs were not easily interpretable, other than on the basis of the item difficulty of the associated items. The generation of artefactual 'difficulty' factors from otherwise unidimensional scales is well recognized (Gillespie *et al.*, 1987; Hattie, 1985). Our Mokken scale analysis gave further strong support for the WHODAS II as a unidimensional hierarchical scale (see later).

Does the WHODAS II conform to IRT principles?

The hierarchality of the original version of the WHODAS II screener was tested using Mokken scaling analysis, the conclusion being that it did not conform to IRT principles given only moderate scalability with significant non-intersection violation (Rehm *et al.*, 1999). Our Mokken analysis of the revised WHODAS II screener suggests that it does conform to IRT principles, at least at the level of a MHM. Item scalability coefficients were all positive and easily exceeded the threshold of 0.3, generally accepted as signifying that items meet MHM assumptions. Coefficient *H*, a weighted mean of item coefficients, exceeded 0.50 in all sites, suggesting that the WHODAS II screener is a 'strong' Mokken scale capable of ordering individuals on the latent disability trait. Consistent with these findings, monotonicity diagnostics were satisfactory for all items in nearly all sites. It is not so clear that the revised WHODAS II screener meets the more stringent criteria for a DMM, in which the rank ordering of item difficulties is invariant across all levels of the trait. While there were a number of non-intersection violations these were few with respect to the number of active pairs ($n = 7040$), and the largest amongst them were close to the 0.03 threshold. Also, no non-intersection violations, or only 'questionable' violations were noted in five of the 11 sites. Measurement properties could perhaps be improved with respect to DMM by omitting items one (difficulty standing) and seven (difficulty walking), but to do so would remove the entire 'getting around' domain from the scale. It should be noted that non-intersection is not among the fundamental IRT assumptions (unidimensionality, local independence and monotonicity), and the majority of IRT models do not imply invariant item ordering (Sijtsma and Hemker, 2000). However, overall, there is strong evidence that the current revised version of the 12-item WHODAS II screener is a hierarchical, unidimensional scale conforming to IRT principles, and as such, the sum of its scores can be taken as a measure of the underlying trait.

Measurement invariance

The marked similarity of the goodness-of-fit of the one-factor and two-factor CFA models when loadings were constrained (to be equal across sites) or not constrained (estimated freely in each site) strongly supports measurement invariance with respect to a common underlying factor structure and factor loadings. Buist-Bownman *et al.* (2008) have previously reported measurement invariance for the full version of the WHODAS II, between Mediterranean and non-Mediterranean countries for all but one item (embarrassment) not included in the 12-item WHODAS II screener. Mokken analysis constitutes a less conservative test of measurement invariance than parametric IRT models. Nevertheless, Mokken gives information about the dimensionality of the data and the properties of the items, with an intuitive relationship between individual measurement values and item difficulty and discrimination values. 'Dressing' and 'maintaining friendship' were, consistently, the most difficult items (only likely to be endorsed by those with the highest scores on the trait), while 'getting around' (walking and standing) had the lowest item difficulties (likely to be endorsed by those with relatively low scores on the trait). The high inter-site correlations and overall intraclass correlation for WHODAS II item difficulties across sites supports measurement invariance with respect to common hierarchical scaling properties and common item difficulties between diverse countries and cultures. The exception, rural China, seemed to have been accounted for by the unusually high proportion in that site (19.2%) reporting at least severe difficulty with learning a new task and with household responsibilities (7.7%). This suggests the possibility of some degree of culturally determined differential item function with respect to these two items. In rural China a high proportion (73.2%) (Prince *et al.*, 2008b) of older people lived with their family. It was noted in the course of our field research that all older people, regardless of functional ability, were given support with personal care and core activities of daily living, and were not expected to perform any more complex instrumental activities of daily living (cooking, shopping, managing household budgets). Hence taking

on household responsibilities and learning new tasks would be seen as exceptional activities and, hence, possibly more challenging than in other sites, even for those with little disability.

Distributional properties in different populations

Others have remarked upon the skewed and zero-inflated character of the WHODAS II distribution, and on the large distributional differences between survey populations from different countries (Buist-Bouwman *et al.*, 2008; Rehm *et al.*, 1999; Von Korff *et al.*, 2008). Zero-inflation seems not to be accounted for by the insensitive filter questions in the WMH survey version of the WHODAS II, since the phenomenon was also evident in our samples, where no filter questions were used. Furthermore, our data suggests that variation in the extent of zero-inflation may account for much of the between site variance in WHODAS II scores. Standardizing WHODAS II distributions by dichotomizing at the 90th centile for each population has been proposed (Von Korff *et al.*, 2008). However, this has the disadvantage both of loss of measurement precision, and loss of the ability to model and explore country differences. We would propose zero-inflated negative binomial regression as the appropriate model for dealing with both overdispersion and zero-inflation, including variation in zero-inflation between samples in cross-cultural research. This model allows for 'excess zeros' in count models under the assumption that the population is characterized by two groups, one where members always have zero counts, and one where members have zero or positive counts. The likelihood of being a certain zero is estimated using a logit specification (an effect of country here could be interpreted as a culturally determined propensity for 'naysaying'), while the counts in the second group are estimated using a negative binomial specification.

Conclusion

In this analysis, we assessed the psychometric properties of 12-item WHODAS II in large and representative community samples of older people in 11 sites across seven LAMIC. Explanatory factor analysis gave rise to a one-factor solution in seven of the 11 sites studied. While CFA demonstrated that a two-factor solution fitted better, the underlying constructs were not easily interpretable, other than on the basis of the item difficulty of the associated items. Mokken scale analysis gave strong support for the WHODAS II as a unidimensional hierarchical scale conforming to IRT principles at least at the monotone homogeneity model level. Measurement invariance was

demonstrated both for item calibrations, and for underlying factor structures and factor loadings. These are all highly desirable properties for a brief disability assessment to be used in cross-cultural comparative research. The brief 12-item version of the WHODAS II has been little used since its development, perhaps because of a more negative report on an earlier version of the scale (Rehm *et al.*, 1999). More work is needed to assess the generalizability of our findings to younger aged samples, and to confirm the reliability and validity of the new version of the WHODAS II screener. The demonstration of robust cross-cultural measurement properties for the WHODAS II screener opens the way for further analyses to explore explanations for observed differences in the distribution of scores between sites. The effect of varying degrees of zero-inflation will need to be accounted for, and the influence of both compositional factors (age, gender, education, socio-economic circumstances, physical, mental and cognitive health) and contextual differences explored.

Acknowledgements

Declaration of interest statement

All authors declare no conflicts of interest.

References

Akaike H. (1987) Factor Analysis and AIC. *Psychometrika*, **52**, 317–332.

Baron M., Schieir O., Hudson M., Steele R., Kolahi S., Berkson L., Couture F., Fitzcharles M.A., Gagne M., Garfield B., Gutkowski A., Kang H., Kapusta M., Ligier S., Mathieu J.P., Menard H., Starr M., Stein M., Zummer M. (2008) The clinimetric properties of the World Health Organization Disability Assessment Schedule II in early inflammatory arthritis. *Arthritis & Rheumatism*, **59**(3), 382–390, DOI: 10.1002/art.23314.

Browne M. (1990) *MUTMUM PC: User's Guide*, Ohio State University.

Buist-Bouwman M.A., Ormel J., De Graff R., Vilagut G., Alonso J., Van Sonderen E., Vollebergh W.A. (2008) Psychometric properties of the World Health Organization Disability Assessment Schedule used in the European Study of the Epidemiology of Mental Disorders.

International Journal of Methods in Psychiatric Research, **17**(4), 185–197, DOI: 10.1002/mpr.261.

Burnham K.P., Anderson D.R. (1998) *Model Selection and Inference: A Practical Information-theoretic Approach*, Springer-Verlag.

Castro-Costa E., Dewey M., Stewart R., Banerjee S., Huppert F., Mendonca-Lima C., Bula C., Reisches F., Wancata J., Ritchie K., Tsolaki M., Mateos R., Prince M. (2008) Ascertaining late-life depressive symptoms in Europe: An evaluation of the survey version of the EURO-D scale in 10 nations. The SHARE project. *International Journal of Methods in Psychiatric Research*, **17**(1), 12–29, DOI: 10.1002/mpr.236.

Chavez L.M., Canino G., Negron G., Shrout P.E., Matias-Carrelo L.E., Guilar-Gaxiola S., Hoppe S. (2005) Psychometric properties of the Spanish version of two mental health outcome measures: World Health Organization Disability Assessment Schedule II and Lehman's Quality Of Life Interview. *Mental Health Services Research*, **7**(3), 145–59.

Chisolm T.H., Abrams H.B., McArdle R., Wilson R.H., Doyle P.J. (2005) The WHO-DAS II: Psychometric properties in the measurement of functional health status in adults with acquired hearing loss. *Trends in Amplification*, **9**(3), 111–126.

Chopra P.K., Couper J.W., Herrman H. (2004) The assessment of patients with long-term psychotic disorders: Application of the WHO Disability Assessment Schedule II. *Australian and New Zealand Journal of Psychiatry*, **38**(9), 753–759.

Chwastiak L.A., Von Korff M. (2003) Disability in depression and back pain: Evaluation of the World Health Organization Disability Assessment Schedule (WHO DAS II) in a primary care setting. *Journal of Clinical Epidemiology*, **56**(6), 507–514, DOI: 10.1016/S0895-4356(03)00051-9.

Dijkstra A., Buist G., Moorer P., Dassen T. (1999) Construct validity of the Nursing Care Dependency Scale. *Journal of Clinical Nursing*, **8**(4), 380–388.

Dunn G., Everitt B., Pickles A. (1993) *Modelling Covariances and Latent Variables using EQS*, Chapman & Hall.

Gillespie M., Tenvergert E., Kingma J. (1987) Using Mokken scale analysis to develop unidimensional scales. Do the six abortion items in the NORC GSS form one or two scales? *Quality Quant*, **21**, 393–408.

Hattie J. (1985) Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, **9**, 139–164.

Hudson M., Steele R., Taillefer S., Baron M., Canadian S.R. (2008) Quality of life in systemic sclerosis: psychometric properties of the World Health Organization Disability Assessment Schedule II. *Arthritis & Rheumatism*, **59**(2), 270–278, DOI: 10.1002/art.23343.

Joreskog K., Sorbom D. (1993) *New Features in PRELIS 2.* Chicago University Press.

Joreskog K. (1994) On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika*, **59**(3), 381–389.

Kim J.M., Stewart R., Glozier N., Prince M., Kim S.W., Yang S.J., Shin I.S., Yoon J.S. (2005) Physical health, depression and cognitive function as correlates of disability in an older Korean population. *International Journal of Geriatric Psychiatry*, **20**(2), 160–167, DOI: 10.1002/gps.1266.

Leon A.C., Olfson M., Portera L., Farber L., Sheehan D.V. (1997) Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. *International Journal of Psychiatry in Medicine*, **27**(2), 93–105.

Llibre Rodriguez J.J., Ferri C.P., Acosta D., Guerra M., Huang Y., Jacob K.S., Krishnamoorthy E.S., Salas A., Sosa A.L., Acosta I., Dewey M.E., Gaona C., Jotheeswaran A.T., Li S., Rodriguez D., Rodriguez G., Kumar P.S., Valhuerdi A., Prince M., Dementia Research Group. (2008) Prevalence of dementia in Latin America, India, and China: a population-based cross-sectional survey. *Lancet*, **372**(9637), 464–474, DOI: 10.1016/S0140-6736(08)61002-8.

Marsh H., Balla J., Hau K. (1996) An evaluation of incremental fit indices: A clarification of mathematical and empirical properties. In: *Advanced Structural Equation Modelling: Issues and Techniques* (eds Marcoulides G., Schumacker R.), pp. 315–355, Lawrence Erlbaum Associates.

McKibbin C., Patterson T.L., Jeste D.V. (2004) Assessing disability in older patients with schizophrenia: results from the WHODAS-II. *Journal of Nervous and Mental Disease*, **192**(6): 405–413, DOI: 10.1097/01.nmd.0000130133.32276.83.

Mogga S., Prince M., Alem A., Kebede D., Stewart R., Glozier N., Hotopf M. (2006) Outcome of major depression in Ethiopia: population-based study. *British Journal of Psychiatry*, **189**, 241–246, DOI: 10.1192/bjp.bp.105.013417

Mokken R. (1971) *Theory and Procedure of Scale Analysis*, Mouton & Co.

Molenaar A.W., Sijtsma K. (2000) *User's Manual MSP5 for Windows*, iecProGAMMA.

Murray C.J.L., Lopez A.D. (1996) *The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries, and Risk Factors in 1990 and Projected to 2020*, Harvard University Press.

Norton J., de Roquefeuil G., Benjamins A., Boulenger J.P., Mann A. (2004) Psychiatric morbidity, disability and service use amongst primary care attenders in France. *European Psychiatry*, **19**(3), 164–167, DOI: 10.1016/j.eurpsy.2003.11.003.

Perini S.J., Slade T., Andrews G. 2006. Generic effectiveness measures: Sensitivity to symptom change in anxiety disorders. *Journal of Affective Disorders*, **90**(2–3), 123–130.

Posl M., Cieza A., Stucki G. (2007) Psychometric properties of the WHODASII in rehabilitation patients. *Quality of Life Research*, **16**(9), 1521–1531, DOI: 10.1007/s11136-007-9259-4.

Prince M.J., de Rodriguez J.L., Noriega L., Lopez A., Acosta D., Albanese E., Arizaga R., Copeland J.R., Dewey M., Ferri C.P., Guerra M., Huang Y., Jacob K.S., Krishnamoorthy E.S., McKeigue P., Sousa R., Stewart R.J., Salas A., Sosa A.L., Uwakwa R., Dementia Research Group. (2008a) The 10/66 Dementia Research Group's fully operationalised DSM-IV dementia computerized diagnostic algorithm, compared with the 10/66 dementia algorithm and a clinician diagnosis:A population validation study. *BMC Public Health*, **8**, 219, DOI: 10.1186/1471-2458-8-219.

Prince M., Acosta D., Albanese E., Arizaga R., Ferri C.P., Guerra M., Huang Y., Jacob K.S., Jimenez-Velazquez I.Z., Rodriguez J.L., Salas A., Sosa A.L., Sousa R., Uwakwe R., van der Poel R., Williams J., Wortmann M. (2008b) Ageing and dementia in low and middle income countries – using research to engage with public and policy makers. *International Review of Psychiatry*, **20**(4), 332–343, DOI: 10.1080/09540260802094712.

R Development Core Team. (2007) R: A language and environment for statistical computing. In: *R Foundation for Statistical Computing*, Vienna.

Rehm J., Ustun T.B., Saxena S. (1999) On the development and psychometric testing of the WHO screening instrument to assess disablement in the general population. *International Journal of Methods in Psychiatric Research*, **8**, 110–122.

Sijtsma K., Emons W.H., Bouwmeester S., Nyklicek I., Roorda L.D. (2008) Nonparametric IRT analysis of Quality-of-Life Scales and its application to the World Health Organization Quality-of-Life Scale (WHOQOL-Bref) *Quality of Life Research*, **17**(2), 275–290, DOI: 10.1007/s11136-007-9281-6.

Sijtsma K., Hemker B.T. (2000) A taxonomy of IRT models for ordering persons and items using simple sum scores. *Journal of Educational and Behavioral Statistics*, **25**(4), 391–415, DOI: 10.3102/10769986025004391.

SPSS Inc. (2003) *Amos 5.0*, SPSS Inc.

SPSS Inc. (2005) *SPSS for Windows*, Rel. 15.0, SPSS Inc.

Stata Corporation. (2007) *Stata Statistical Software: Release 10.0*, Stata Corporation.

Tucker L., Lewis C. (1973) A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, **38**, 1–10.

United Nations. (2007) *World Population Prospects: The 2006 Revision, Highlights*, Department of Economic and Social Affairs, Population Division, United Nations.

Van der Ark L.A., Croon M.A., Sijtsma K. (2007) Possibilities and challenges in Mokken scale analysis using marginal models. In: *New Trends in Psychometrics* (eds Shigemasu K., Okada A., Imaizumi T., Hoshino T.), Universal Academic Press.

van Tubergen A., Landewe R., Heuft-Dorenbosch L., Spoorenberg A., van der Heijde D., van der Tempel H., van der Linden S. 2003. Assessment of disability with the World Health Organisation Disability Assessment Schedule II in patients with ankylosing spondylitis. *Annals of the Rheumatic Diseases*, **62**(2), 140–145, DOI: 10.1136/ard.62.2.140.

Von Korff M., Crane P.K., Alonso J., Vilagut G., Angermeyer M.C., Bruffaerts R., de Girolamo G., Gureje O., de Graaf R., Huang Y., Iwata N., Karam E.G., Kovess V., Lara C., Levinson D., Posada-Villa J., Scott K.M., Ormel J. (2008) Modified WHODAS-II provides valid measure of global disability but filter items increased skewness. *Journal of Clinical Epidemiology*, **61**(11), 1132–1143, DOI: 10.1016/j.jclinepi.2007.12.009.

World Health Organization (WHO). (2001a) *International Classification of Functioning, Disability and Health (ICF)*, WHO.

World Health Organization (WHO). (2001b) *WHODAS II Disability Assessment Schedule*. http://who.int/icidh/whodas/index.html, accessed 7 July 2008.

Appendix

Table A1 The 12-items WHODAS II

Question	Description					
H1	How do you rate your overall health in the past 30 days?	Very good	Good	Moderate	Bad	Very bad
In the last 30 days, how much difficulty did you have in:						
S1	Standing for long periods such as 30 minutes?	None	Mild	Moderate	Severe	Extreme/can't do
S2	Taking care of your household responsibilities?	None	Mild	Moderate	Severe	Extreme/can't do
S3	Learning a new task, for example, learning how to get to a new place?	None	Mild	Moderate	Severe	Extreme/can't do
S4	How much of a problem did you have joining in community activities (for example, festivities, religious or other activities) in the same way as anyone else can?	None	Mild	Moderate	Severe	Extreme/can't do
S5	How much have you been emotionally affected by your health problems?	Not at all	Mildly	Moderately	Severely	Extremely
S6	Concentrating on doing something for 10 minutes?	None	Mild	Moderate	Severe	Extreme/can't do
S7	Walking a long distance such as a kilometre (or equivalent)?	None	Mild	Moderate	Severe	Extreme/can't do
S8	Washing your whole body?	None	Mild	Moderate	Severe	Extreme/can't do
S9	Getting dressed?	None	Mild	Moderate	Severe	Extreme/can't do
S10	Dealing with people you do not know?	None	Mild	Moderate	Severe	Extreme/can't do
S11	Maintaining a friendship?	None	Mild	Moderate	Severe	Extreme/can't do
S12	Your day to day work?	None	Mild	Moderate	Severe	Extreme/can't do