

# Utility of the twelve-item World Health Organization Disability Assessment Schedule II (WHO-DAS II) for discriminating depression “caseness” and severity in Spanish primary care patients

Juan V. Luciano · José L. Ayuso-Mateos · Ana Fernandez ·  
Jaume Aguado · Antoni Serrano-Blanco · Miquel Roca ·  
Josep M. Haro

Accepted: 2 December 2009 / Published online: 18 December 2009  
© Springer Science+Business Media B.V. 2009

## Abstract

**Purpose** The 12-item WHO-DAS II was developed to assess the activity limitations and participation restrictions experienced by individuals irrespective of medical diagnosis. In this paper we examine the known-groups’ validity of the instrument by evaluating its ability to discriminate between patients with/without major depression, patients

with depression with/without medical comorbidity, and patients with depression with different depression severity. **Method** The participants were 3,615 PC patients from 17 regions of Spain, with a first-time diagnosis of major depressive episode according to the general practitioner. The 12-item WHO-DAS II, the PHQ-9, and a chronic medical conditions checklist were administered during the consultation.

**Results** The statistical analyses indicated that the 12-item WHO-DAS II was able to discriminate between patients with/without depression and between those with different depression severity. The ROC analysis revealed that with a cutoff score  $\geq 50$ , the instrument correctly classified 70.4% of the sample (area under the ROC curve = .76; sensitivity = 71.4%; specificity = 67.6%).

**Conclusions** Overall, our results support the discriminant validity of the 12-item WHO-DAS II for major depression, being quite recommendable its use in epidemiological research.

**Keywords** 12-item WHO-DAS II · Primary health care · Depression · Disability

---

J. V. Luciano (✉) · A. Fernandez · J. Aguado ·  
A. Serrano-Blanco · J. M. Haro  
Parc Sanitari Sant Joan de Déu, Fundació Sant Joan de Déu,  
Sant Boi de Llobregat, Barcelona, Spain  
e-mail: jvluciano@sjd-ssm.com

J. V. Luciano · A. Fernandez · A. Serrano-Blanco · M. Roca  
RedIAPP  
URL: <http://www.rediapp.org/>

J. L. Ayuso-Mateos  
Departamento de Psiquiatría, Universidad Autónoma de Madrid,  
Madrid, Spain

J. L. Ayuso-Mateos  
Servicio de Psiquiatría, Hospital Universitario de la Princesa,  
Madrid, Spain

J. L. Ayuso-Mateos · J. M. Haro  
CIBERSAM  
URL: <http://www.cibersam.es/>

J. Aguado  
Departament de Salut Pública, Universitat de Barcelona,  
Barcelona, Spain

M. Roca  
IUNICS, Universitat de les Illes Balears, Palma de Mallorca,  
Spain

M. Roca  
Unidad de Psiquiatría y Psicología Clínica – Hospital Juan  
March, Palma de Mallorca, Spain

## Introduction

Mental disorders are among the leading causes of disability worldwide and will be among the most burdensome conditions by the year 2020 [1]. Among mental disorders, major depression [2–4] is associated with greatest disability in high, middle, and low-income countries.

The prevalence of major depression in primary health care (PC) is considerable [5–7]; however, fewer than half of the patients with depression are correctly identified and adequately treated by general practitioners (GPs) [8, 9].

Moreover, we now know that, if used alone, screening instruments do not distinguish well between those patients who are disabled by their symptoms and those who are not [10] and have little or no impact on the detection, management, and outcome of depression [11].

The World Health Organization Disability Assessment Schedule II (WHO-DAS II) [12] was designed to assess the activity limitations and participation restrictions experienced by an individual irrespective of medical diagnosis. The main advantages of this instrument over other disability measures are as follows: it was cross-culturally developed and field tested in 16 languages in 14 different countries, it is compatible with an international classification system (the International Classification of Functioning, Disability and Health) [13], and it treats all disorders at parity when establishing the level of functioning. Several studies have extensively analyzed the dimensionality, internal consistency, test–retest reliability, and construct validity of the 36- and 12-item version of the WHO-DAS II in patients with diverse physical and mental conditions [14–17], demonstrating that the instrument possesses sound psychometric properties.

However, to the best of our knowledge, no study has addressed the extent to which the 12-item WHO-DAS II can discriminate between the following clinical groups in the context of primary health care: patients with/without major depression, patients with depression with/without medical comorbidity, and patients with depression with different depression severity. These known-groups validity analyses were carried out in the present work.

## Methods

For this study, we utilized the ERASMAP data set. The ERASMAP was a cross-sectional observational study carried out in 874 PC centers in Spain designed to identify the sociodemographic and clinical factors associated with diagnostic delay in a first diagnosed major depressive episode. The study was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and was approved by the Clinical Research and Ethics Committee of the University Hospital La Princesa (Madrid, Spain).

### Participants

The sample consisted of 3,615 adult (18 years or older) PC patients from 17 regions of Spain, with a first-time diagnosis of major depressive episode. PC patients with a previously diagnosed major depressive episode, bipolar disorder, schizophrenia or delusional disorder, and those

who were receiving treatment with any psychotropic medication were not included in the study.

### Measures

The *12-item interviewer administered version of the World Health Organization Disability Assessment Schedule II* (12-item WHO-DAS II) [14, 17]. In each item, individuals have to estimate the magnitude of the disability during the previous 30 days from none = 1 to extreme/cannot do = 5. The total score may vary from 0 to 100 with higher scores reflecting greater disability.

The *Patient Health Questionnaire nine-item depression module* (PHQ-9) [18, 19]. A nine-item scale that assesses the nine DSM-IV [20] depression symptoms. Each of the nine items is scored from 0, not at all, to 3, nearly every day. The PHQ-9 can be used as a screening tool, with summed score ranging from 0 (no depressive symptoms) to 27 (all symptoms occurring daily). Summed scores of 0–4 represent a minimal level of depression; 5–9, mild; 10–14, moderate; 15–19, moderately severe; and 20–27, severe. The PHQ-9 can also be used as a diagnostic tool using a “diagnostic algorithm”; major depression is diagnosed if 5 or more of the 9 symptoms have been present at least more than half the days of the past 2 weeks, and 1 of these symptoms is either depressed mood or anhedonia.

### Chronic medical conditions checklist

The presence of comorbid medical conditions was assessed using a yes-or-no checklist developed by the authors for the present study. It included questions about a wide range of conditions (e.g. migraine, arthritis, heart attack, hypertension, asthma, tuberculosis, diabetes, etc.). Respondents were asked whether they had experienced any of the symptom-based conditions in the checklist during the previous year.

### Procedure

During the consultation, the participating GPs assessed the patients meeting the inclusion criteria using a paper-and-pencil interview. Prior to the assessment, all patients had provided written informed consent.

### Data analyses

The known-groups’ validity approach is founded on the basis that certain specified groups of patients might be expected to score differently from others. In the present work, we carried out the following analyses to examine the known-groups’ validity of the 12-item WHO-DAS II: First, a Student’s *t* test for independent samples (with unequal variances) was performed to assess the validity of the

instrument for discriminating between the patients with major depression and those without (according to the PHQ-9 diagnostic algorithm).

We then conducted a ROC analysis to examine the sensitivity and specificity of the instrument for major depression, using the PHQ-9 as a “gold standard”. The area under the curve (interpretation: .50 to .75 = fair, .75 to .92 = good, .92 to .97 = very good, .97 to 1.00 = excellent), positive and negative predictive value, and the positive and negative likelihood ratio were all calculated.

Finally, to examine the differences in disability among patients with depression with and without comorbid medical conditions, as well as among those reporting different degrees of depression severity (depression groups using PHQ-9: 10–14 = moderate; 15–19 = moderately severe; 20–27 = severe depression), a Student’s *t* test for independent samples (with unequal variances) and one-way Analysis of Variance (ANOVA; using Games-Howell for post hoc comparisons) were performed, respectively. The overall alpha level for was set at .05.

## Results

Patient characteristics and scores on study measures are described in Table 1 using means and standard deviations for continuous variables and percentages for categorical variables. Means and standard deviations for the PHQ-9 and the 12-item WHO-DAS II by group are displayed in Table 2.

### Discriminating depression “caseness”

The analysis revealed a significant group difference in disability,  $t_{(1684.57)} = 26.86$ ,  $P < .001$ . The 2,612 PC patients with major depression (according to PHQ-9) obtained significantly higher scores on the WHO-DAS II ( $M = 58.02$ ,  $SD = 16.39$ ) than the 913 without depression ( $M = 41.84$ ,  $SD = 15.41$ ). We computed Cohen’s *d* from the value of the *t* test of the differences between the two groups (Rule of thumb: .20 = small; .50 = medium; .80 = large). The effect size was large ( $d = 1.31$ ).

Subsequently, the ROC analysis revealed that the accuracy of the WHO-DAS II with respect to discriminating depression “caseness” was good (see Fig. 1;  $AUC = .76$ ,  $SE = .0088$ ,  $P < .001$ ,  $95\%CI .75—.78$ ,  $LR + = 2.20$ ,  $LR - = .42$ ). The point of maximum curvature of the ROC analysis suggested that a cutoff score  $\geq 50\%$  yielded the best trade-off between sensitivity (71.4%) and specificity (67.6%) for the 12-item WHO-DAS II, correctly classifying 70.4% of the sample and producing a positive predictive value of 86.3% and a negative predictive value of 45.2%.

**Table 1** Characteristics of the study sample

Sociodemographic variables	
Gender %	
Male	32.67
Female	67.33
Age, <i>M</i> ( <i>SD</i> )	50.01 (13.89)
Marital status %	
Married	54.67
Living with a partner	9.48
Single	11.76
Separated/divorced	12.91
Widowed	11.18
Education level %	
Did not graduate from primary school	21.44
Primary school	15.12
Secondary school	43.14
University	20.30
Work status %	
Student	1.31
Homemaker	23.36
Paid employment	57.70
Paid employment but in sick leave	39.52
Unemployed	2.34
Permanent disability	1.78
Retired/pensioner	13.37
Others	0.14
Study measures, <i>M</i> ( <i>SD</i> )	
12-item WHO-DAS II	53.83 (17.63)
PHQ-9	17.15 (4.64)
No medical conditions	2.04 (1.72)

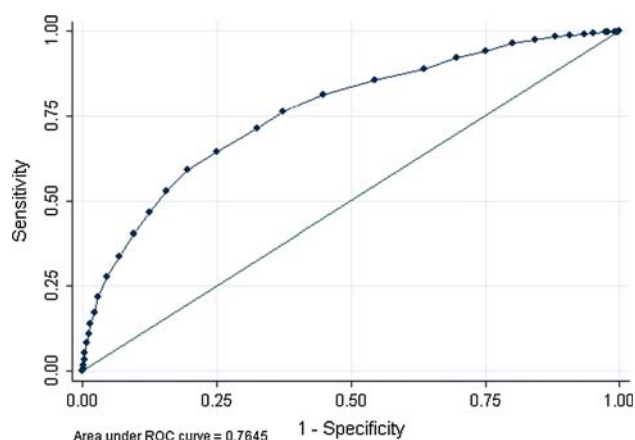
12-item WHO-DAS II = World Health Organization Disability Assessment Schedule II (12-item version); PHQ-9 = Patient Health Questionnaire

**Table 2** Means and standard deviations (SD) for the PHQ-9 and the 12-item WHO-DAS II by group

Group*	PHQ-9	12-item WHO-DAS II
Depressed	18.91 (3.47)	58.02 (16.39)
Nondepressed	11.73 (3.57)	41.84 (15.41)
Depressed with comorbidity	17.06 (4.70)	54.10 (17.15)
Depressed without comorbidity	16.98 (4.72)	52.82 (19.30)
Moderate depression	12.47 (1.32)	43.12 (13.23)
Moderately severe depression	17.05 (1.38)	54.56 (15.08)
Severe depression	22.30 (2.03)	64.74 (15.51)

\* Groups established using the PHQ-9

12-item WHO-DAS II = World Health Organization Disability Assessment Schedule II (12-item version); PHQ-9 = Patient Health Questionnaire



**Fig. 1** Receiver operating characteristic curve (ROC) for the 12-item WHO-DAS II versus PHQ-9 (gold standard) for major depression diagnosis

#### Discriminating depression with/without medical comorbidity

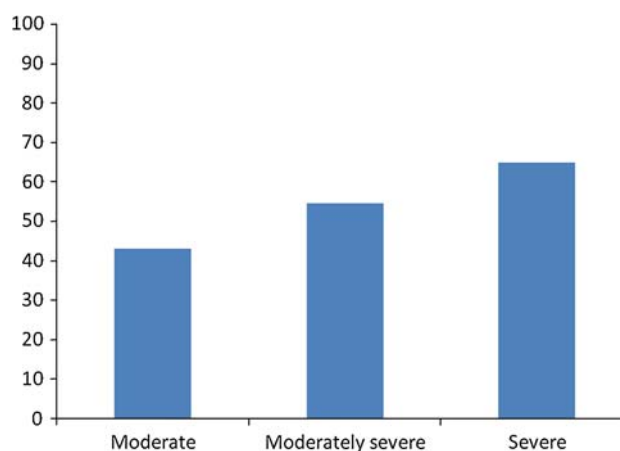
The 744 patients with depression without medical comorbidity presented lower scores on the 12-item WHO-DAS II ( $M = 52.82$ ,  $SD = 19.30$ ) than the 2,781 depressed participants that were suffering one or various comorbid medical conditions ( $M = 54.10$ ,  $SD = 17.15$ ), but this difference was not statistically significant,  $t_{(1077.03)} = 1.63$ ,  $P = .10$ .

#### Discriminating depression severity

Mean scores (standard deviations) on the 12-item WHO-DAS II were 43.12 (13.23), 54.56 (15.08), and 64.74 (15.51) for moderate ( $n = 793$ ), moderately severe ( $n = 1,414$ ), and patients with severe depression ( $n = 1,105$ ), respectively. The ANOVA yielded significant group differences in disability,  $F_{(2, 3309)} = 494.55$ ,  $P < .001$  ( $n = 3,312$  after listwise deletion). The effect size analysis based on partial eta-squared ( $\eta_p^2$  rule of thumb: .01 = small; .06 = medium; .14 = large) indicated that the difference was large ( $\eta_p^2 = .23$ ). The Games–Howell post hoc test indicated that all pairwise comparisons were statistically significant Fig. 2.

#### Discussion

The known-groups validity analyses reported here support the utility of the 12-item WHO-DAS II for discriminating depression “caseness” and severity among PC patients with a first diagnosed major depressive episode. However, the instrument was not able to discriminate the presence/absence of medical comorbidity among PC patients with depression.



**Fig. 2** 12-item WHO-DAS II scores of patients with depression with different levels of depression severity (according to PHQ-9)

Our results are in line with those recently obtained by Baron and collaborators [21] with the 36-item WHO-DAS II. These authors divided the patients with inflammatory arthritis into two subsets according to their scores on the Center for Epidemiologic Studies Depression Scale (CES-D) and found that the instrument was able to discriminate between patients with low ( $CES-D < 19$ ) and high ( $CES-D \geq 19$ ) depressive symptoms.

Although our objective was not to examine the validity of the instrument as a diagnostic tool, we found in the ROC analysis that with a cutoff score  $\geq 50\%$ , depression “caseness” was detected with an acceptable sensitivity and specificity. Notwithstanding, it would not be reasonable to use the instrument as a substitute for available screening tools (e.g. the PHQ-9). In the clinical context, positive screening results are usually followed by a further diagnostic interview. Therefore, the sensitivity of screening instruments should be above specificity and be as high as possible (at least 90%) in order to avoid the presence of excessive false-negative results. At the same time it is also necessary to avoid the presence of too many false-positive results, requiring a specificity of at least 75% [22]. In the ROC analysis, none of the cutoff points yielded sensitivity and specificity values that met both criteria (data not presented). Therefore, we find appropriate to use the 12-item WHO-DAS II as a complementary tool, its administration being recommended in combination with a depression-screening or case-finding instrument.

The discriminative validity reported here was quite similar to that obtained with other disability instruments used in PC [23, 24]. Luciano and collaborators [23] found that the Sheehan disability scale (SDS) had good sensitivity (82%) and specificity (71%) for major depression. Similarly, Leon et al. [24] examined the utility of the SDS for identifying PC patients with any of six mental disorders (alcohol dependence, drug dependence, generalized anxiety disorder, major

depression, OCD and panic disorder) and found adequate sensitivity (83%) and specificity (69%). Given that the 12-item WHO-DAS II and the SDS seem to have similar psychometric properties using classical test theory, it might be interesting to analyze in a future study the ability of their individual items to discriminate across varying levels of disability using methods based on item-response theory [25].

**Acknowledgments** H. Lundbeck A/S funded this study.

## References

- Mathers, C. D., & Loncar, D. (2005). *Updated projections of global mortality and burden of disease, 2002–2030: data sources, methods and results*. Geneva: World Health Organization (Evidence and Information for Policy Working Paper).
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., et al. (2003). The epidemiology of major depressive disorder. Results from the National Comorbidity Survey Replication (NCS-R). *Journal of the American Medical Association*, *289*, 3095–3105.
- Üstün, T. B., Ayuso-Mateos, J. L., Chatterji, S., Mathers, C., & Murray, C. J. (2004). Global burden of depressive disorders in the year 2000. *British Journal of Psychiatry*, *184*, 386–392.
- Ormel, J., Petukhova, M., Chatterji, S., Aguilar-Gaxiola, S., Alonso, J., Angermeyer, M. C., et al. (2008). Disability and treatment of specific mental and physical disorders across the world. *British Journal of Psychiatry*, *192*, 368–375.
- Wittchen, H. U., & Pittrow, D. (2002). Prevalence, recognition and management of depression in primary care in Germany: The depression 2000 study. *Human Psychopharmacology*, *17*(Suppl. 1), S1–S11.
- Ansseau, M., Fischler, B., Dierick, M., Albert, A., Leyman, S., & Mignon, A. (2007). Socioeconomic correlates of generalized anxiety disorder and major depression in primary care: the GADIS II study (generalized anxiety and depression impact survey II). *Depression and Anxiety*, *26*, 1–8.
- Serrano-Blanco, A., Palao, D.J., Luciano, J.V., Pinto-Meza, A., Luján, L., Fernández, A., Roura, P., Bertsch, J., Mercader, M., Haro, J.M. (2009). Prevalence of mental disorders in primary care: Results from the diagnosis and treatment of mental disorders in primary care study (DASMAP). *Social Psychiatry and Psychiatric Epidemiology*. May 19. [Epub ahead of print].
- Cepoiu, M., McCusker, J., Cole, M. G., Sewitch, M., Belzile, E., & Ciampi, A. (2007). Recognition of depression by non-psychiatric physicians—A systematic literature review and meta-analysis. *Journal of General Internal Medicine*, *23*, 25–36.
- Leclercq, Y. (2007). Widespread underrecognition and undertreatment of anxiety and mood disorders: Results from 3 European studies. *Journal of Clinical Psychiatry*, *68*(Suppl 2), 36–41.
- Collings, S. (2005). Disability and the detection of mental disorder in primary care. *Social Psychiatry and Psychiatric Epidemiology*, *40*, 994–1002. The MaGPIe Research Group.
- Gilbody, S., Sheldon, T., & House, A. (2008). Screening and case-finding instruments for depression: A meta-analysis. *Canadian Medical Association Journal*, *178*, 997–1003.
- World Health Organization. (2000). Disability assessment schedule II (WHO-DAS II). Geneva: WHO. <http://www.who.int/icidh/whodas/whodasversions/12int.pdf>. Accessed 15 May 2008.
- World Health Organization. (2001). International classification of functioning, disability and health. Geneva: WHO. <http://www.who.int/classifications/icf/en/>. Accessed 9 May 2008.
- 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*, 507–514.
- Pösl, M., Cieza, A., & Stucki, G. (2007). Psychometric properties of the WHO-DAS II in rehabilitation patients. *Quality of Life Research*, *16*, 1521–1531.
- Luciano, J.V., Ayuso-Mateos, J.L., Fernández, A., Serrano-Blanco, A., Roca, M., Haro, J.M. (2009). Psychometric properties of the twelve item World Health Organization Disability Assessment Schedule II (WHO-DAS II) in Spanish primary care patients with a first major depressive episode. *Journal of Affective Disorders*. May 21. [Epub ahead of print].
- Vázquez-Barquero, J. L., Vázquez Bourgón, E., Herrera Saiz, J., Uriarte, M., Morales, F., Gaité, L., et al. (2000). Spanish version of the new World Health Organization Disability Assessment Schedule II (WHO-DAS-II): Initial phase of development and pilot study. Cantabria disability work group. *Actas Españolas de Psiquiatria*, *28*, 77–87.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*, 606–613.
- Diez-Quevedo, C., Rangil, T., Sanchez-Planell, L., Kroenke, K., & Spitzer, R. L. (2001). Validation and utility of the patient health questionnaire in diagnosing mental disorders in 1003 general hospital Spanish inpatients. *Psychosomatic Medicine*, *63*, 679–686.
- Association, American Psychiatric. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.
- Baron, M., Schieir, O., Hudson, M., Steele, R., Kolahi, S., Berkson, L., et al. (2008). The clinimetric properties of the World Health Organization Disability Assessment Schedule II in early inflammatory arthritis. *Arthritis and Rheumatism*, *59*, 382–390.
- Löwe, B., Spitzer, R. L., Grafe, K., Kroenke, K., Quenter, A., Zipfel, S., et al. (2004). Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physician's diagnoses. *Journal of Affective Disorders*, *78*, 131–140.
- Luciano, J.V., Bertsch, J., Salvador-Carulla, S., Fernandez, A., Pinto-Meza, A., Haro, J.M., Palao, D.J., Serrano-Blanco, A. Factor structure, internal consistency and construct validity of the Sheehan disability scale in a spanish primary care sample. *Journal of Evaluation in Clinical Practice* (In press).
- 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*, 93–105.
- Rabe-Hesketh, S., & Skrondal, A. (2008). Classical latent variable models for medical research. *Statistical Methods in Medical Research*, *17*, 5–32.