

Disability and perceived stress in primary care patients with major depression

Jesús Santolaya Prego de Oliver¹, Roger Muñoz-Navarro^{2,3}, César González-Blanch⁴, Ana Llorca-Mestre³, Elisabeth Malonda³, Leonardo Medrano⁵, Paloma Ruiz-Rodríguez⁶, Juan Antonio Moriana⁷, and Antonio Cano-Vindel⁸

¹ Universidad Europea de Valencia, ² Universidad de Zaragoza, ³ Universitat de València, ⁴ Hospital Universitario "Marqués de Valdecillas" (Santander), ⁵ Pontificia Universidad Católica Madre y Maestra (República Dominicana), ⁶ Servicio Madrileño de Salud, ⁷ Universidad de Córdoba, and ⁸ Universidad Complutense de Madrid

Abstract

Background: Major depressive disorder (MDD) is highly prevalent in the Spanish primary care (PC) setting and the leading cause of disability in Spain. The aim of this study was to evaluate several key psychometric properties of the Sheehan Disability Scale (SDS) in patients with or without MDD and varying degrees of symptom severity using the Patient Health Questionnaire-9 (PHQ-9). **Method:** A total of 1,704 PC patients participating in the PsicAP clinical trial completed the SDS and PHQ-9. We evaluated the factor structure, measurement invariance across gender, internal consistency, and the discriminative and predictive validity. **Results:** Confirmatory factor analyses revealed a unifactorial model of the SDS containing 4 items (SDS-4) with the best model fit (CFI: .99; GFI: .99; TLI: 96; RMSEA: .10). This model contained the three life domain items (work, family, and social life) plus perceived stress (PS) with significant loadings. The internal consistency of the SDS-4 was acceptable in patients with or without MDD, regardless of symptom severity. The SDS-4 also showed good discriminative capacity and acceptable predictive validity in all subsamples. **Conclusions:** These findings support the use of the SDS-4 to assess depression-related disability in patients at Spanish primary care centres.

Keywords: Disability, perceived stress, primary care, major depression, depressive symptoms.

Resumen

Discapacidad y estrés percibido en pacientes de atención primaria con depresión mayor. Antecedentes: el Trastorno Depresivo Mayor (TDM) es muy prevalente en la atención primaria (AP) española y es la principal causa de discapacidad en España. En este trabajo se estudiaron algunas propiedades psicométricas de la Escala de Discapacidad de Sheehan (SDS) en pacientes con o sin TDM y con varios niveles de severidad de síntomas, usando el Patient Health Questionnaire-9 (PHQ-9). **Método:** 1.704 pacientes de AP que participaron en el ensayo clínico PsicAP cumplimentaron la SDS y el PHQ-9. Se estudió la estructura factorial, medida de invarianza de género, consistencia interna, validez discriminativa y la validez predictiva. **Resultados:** el análisis factorial confirmatorio mostró un modelo unifactorial con 4 ítems (SDS-4) con un buen ajuste (CFI: .99; GFI: .99; TLI: 96; RMSEA: .10). Este modelo contenía los tres ítems de esferas de la vida (trabajo, familia y vida social) más el ítem de estrés percibido con cargas significativas. La consistencia interna de la SDS-4 fue aceptable en pacientes con MDD y con diferente severidad de síntomas depresivos. También se halló una buena validez discriminativa y aceptable validez predictiva. **Conclusiones:** nuestros hallazgos respaldan el uso de la SDS-4 cuando se evalúe la discapacidad relacionada con la depresión en centros de AP españoles.

Palabras clave: discapacidad, estrés percibido, atención primaria, depresión mayor, síntomas depresivos.

Major depressive disorder (MDD) is among the most disabling mental disorders worldwide (World Health Organization, 2017), imposing an enormous social and financial burden on society (Friedrich, 2017). In Spain, in addition to the financial and societal costs, this is the leading cause of disability as measured in quality-adjusted life years (QALY) in (Ruiz-Rodríguez et al., 2017) and is highly prevalent in the Spanish primary care (PC) setting (Muñoz-Navarro et al., 2017a; Roca et al., 2009).

Moreover, greater presence of MDD results in higher levels of disability and less quality of life (González-Blanch et al., 2018a; 2018b). Unfortunately, MDD is underdiagnosed and undertreated in the Spanish national health care system (Fernández et al., 2010). Consequently, there is an urgent need to improve early detection and to shorten the time between detection and referral to appropriate care.

The use of brief screening tests in the PC setting is a cost-effective and efficient approach to detect patients with suspected depression. The Patient Health Questionnaire (PHQ-9) (Kroenke, Spitzer, & Williams, 2001) is the most widely-used screening tool to assess depression and depressive symptoms in primary care (Mitchell, Yeadarfar, Gill, & Stubbs, 2016). This tool was recently validated by Muñoz-Navarro et al. (2017a) for use in Spanish primary care centres. That validation study showed that a

cut-off score of 12 on the PHQ-9 yields an acceptable sensitivity and specificity (.84 and .78, respectively). Similarly, the PHQ-9 used as a diagnostic algorithm also showed an acceptable sensitivity (.88) and specificity (.80). The findings of another recent study supported the plausibility of a unifactorial, invariant model across gender and various sociodemographic variables in a large (n=1250) Spanish primary care sample (González-Blanch et al., 2018c). These studies confirm the reliability of the PHQ-9 in the PC setting. However, to detect the presence of clinical depressive symptoms, it is essential to also check for the presence of impairment and disability. Given that disability and depressive symptoms do not necessarily present simultaneously, it is important to detect and treat both components, especially in primary care (Luciano et al., 2010).

According to the World Health Organisation (WHO, 2019), disability is a physical or mental condition that limits an individual's ability to perform common activities of daily living. Although there are many domains of disability, depression-related disability has been mainly associated with functional and social domains. One of the most widely used measures to assess these disability domains in the primary care setting is the Sheehan Disability Scale (SDS) (Sheehan, Harnett-Sheehan, & Raj, 1996), due to its shortness. The SDS is a five-item instrument that measures the severity of disability as the degree of functional impairment on three inter-related life domains in one single factor—work, family, and social life—plus two additional items to assess perceived stress (PS) and social support (SS) that work independently. The majority of psychometric studies conducted in PC settings to examine depression (Leon, Shear, Portera, & Klerman, 1992; Leon, Olfson, Portera, Farber, & Sheehan, 1997; Luciano et al., 2010) and in other settings and health conditions (Arbuckle et al., 2009; Hambrick, Turk, Heimberg, Schneider, & Liebowitz, 2004; Hodgins, 2013) have consistently found a factor structure consisting of only three-items (work, family, and social life), while the other two items (PS and SS) provide additional information but are not included in the factor structure. The SDS was validated for administration in Spain by Bobes et al. (1999), who found an acceptable internal consistency for the life domain factor ($\alpha = 0.72$), convergent validity ($r = 0.35 - 0.52$), and a high test-retest reliability ($r = 0.87$). However, the single items (PS and SS) yielded worse results. To date, only one study has evaluated the psychometric properties of the SDS in the Spanish PC setting (Luciano et al., 2010). Consistent with previous studies, that study also found a one-dimensional factor containing the three life domains items with the PS and SS as independent items. The one-dimensional factor presented good internal consistency ($\alpha = 0.83$) and good discriminative validity for detecting MDD (sensitivity = 81.6%, specificity = 70.6%). To our knowledge, no studies have yet evaluated the psychometric properties of the five-item SDS. This is an important gap in the literature given previous research suggesting the need to evaluate different models that assess the dimensionality of items that assess latent factors (Muñiz & Fonseca-Pedrero, 2018). Therefore, the objective of the present study, carried out in primary care centres in Spain, was to evaluate the factor structure, measurement invariance across gender, internal consistency, and discriminative and predictive validity of the SDS in a sample of patients with or without depression and with varying levels of symptom severity.

Method

Participants

A total of 1,704 patients were recruited between January, 2014 and May, 2017 from 28 PC centres involved in the PsicAP study (Cano-Vindel et al., 2016). The PsicAP is a randomized controlled trial being conducted to test the efficacy of group transdiagnostic cognitive-behavioral therapy versus treatment as usual for emotional disorders in the PC setting in Spain. Of these 1704 patients, 78.6% were women and 21.4% men. The mean age was 43.6 (± 12.3) years (range, 16 - 80). Table 1 shows the sociodemographic data for the sample.

Instruments

Sheehan Disability Scale (SDS). The SDS is a five-item tool that measures disability severity using an analogue visuo-spatial scale with numeric and verbal descriptive anchors (Sheehan et al., 1996). It contains a single total disability factor containing the

Table 1
Demographic characteristics of sample

	Total sample (n = 1704)	
	N	%
Gender		
Female	1340	78.6
Male	364	21.4
Age group, years		
≤ 19	27	1.6
20 - 39	609	35.7
40 - 59	912	53.5
≥ 60	156	9.2
Marital status		
Married	790	46.4
Divorced	154	9.0
Widowed	55	3.2
Separated	87	5.1
Never married	358	21.0
Unmarried	260	15.3
Level of education		
No schooling	25	1.5
Basic education	432	25.4
Secondary education	371	21.8
High School	434	25.5
Bachelor	366	21.5
Master/doctorate	76	4.5
Employment status		
Employed part-time	639	14.7
Employed full time	250	37.5
Unemployed, in search of work	367	21.5
Unemployed, not looking for work	205	12.0
Temporary incapacity to work	129	7.6
Permanent incapacity to work	38	2.2
Retired	76	4.5
Income		
< 12,000 €	673	39.5
12,000 - 24,000€	697	40.9
24,000 - 36,000€	221	13.0
> 36,000€	113	6.6

first three items, which measure three inter-related life domains; a) working life; b) family life; and c) social life, plus two additional items that assess perceived stress and social support. For the first four items, patients rate the extent to which their symptoms impair the domain of interest using a 10-point Likert scale (0 = 'Not at all, 5 = 'Moderately, 10 = 'Extremely). The total disability total score ranges from 0 (unimpaired) to 30 (highly impaired). Scores for the PS range from 0-10 while the SS is scored from 0% to 100%. Higher values on the PS reflect greater stress while higher percentages on the SS indicate greater social support. The Spanish version validated by Bobes et al. (1999) was used for this study.

Patient Health Questionnaire-9 (PHQ-9). The PHQ-9 (Kroenke et al., 2001) is the depression module of the PHQ containing the nine DSM-IV items used to diagnose major depressive disorder (MDD). The scale is designed to detect the presence of the following symptoms over the previous two weeks: (a) anhedonia; (b) depressed mood; (c) sleep problems; (d) feelings of tiredness; (e) changes in appetite or weight; (f) feelings of guilt; (g) difficulty concentrating; (h) feelings of worry; and (i) suicidal ideation. Responses are given on a four-point Likert scale ranging from 0 - 3 as follows: 0 (never), 1 (several days), 2 (more than half of the days), and 3 (most days). The total score thus ranges from 0 to 27. The PHQ-9 can be used to diagnose MDD using a specific coding algorithm based on the DSM-IV criteria. For this diagnosis, at least one of the two first symptoms (i.e., items) must have a score ≥ 2 and four of the remaining items must also receive a score ≥ 2 (with the exception of item 9 [suicide], in which a rating of 1 is sufficient). The PHQ-9 also rates symptom severity, as follows: 0-4 (no symptoms); 5-9 (mild); 10-14 (moderate); 15-19 (moderately severe); and 20-27 (severe). The Spanish version of this module validated for use in Spanish primary care centres was used (Muñoz-Navarro et al., 2017a).

Procedure

General practitioners involved in the PsicAP trial invited patients with a suspected emotional disorder (anxiety, depression, or somatisation) to participate. All participants were provided with a patient information sheet containing full details about the study purpose. Signed informed consent was required for participation in the study. Patients who agreed to participate were scheduled to meet with a clinical psychologist, at which time they completed various computerized measures. The study was conducted in accordance with the principles of the Declaration of Helsinki and with the Spanish Law on Data Protection. It was approved by the Clinical Research Ethics Committee of primary care in Valencia, Spain (CEIC-APCV) and the Spanish Agency of Medicines and Medical Devices (AEMPS) (EUDRACT: 2013-001955-11 and Protocol Code: ISRCTN58437086).

Data analysis

Factor structure

The SPSS v.26 software program and the Amos module were used to perform the statistical analyses. A confirmatory factor analysis (CFA) was performed to test the fit of four different models, as follow: 1) Model 1 (SDS-3), a unidimensional factor containing the three items of the total disability factor; 2) Model 2 (SDS-4), a unidimensional factor containing the total disability

factor plus the PS item; 3) Model 3 (SDS-5), a unidimensional factor containing all five items; and 4) Model 4, a bidimensional model with 2 factors: factor 1 contained the three items of the total disability factor and factor 2 contained the other two items (PS and SS). Confirmation of the adequacy of these models was performed by determining absolute fit indices using the chi-square statistic (χ^2) and the ratio among degrees of freedom (values < 2 are recommended). We assessed the goodness-of-fit (GFI) indices and the comparative fit index (CFI) to assess incremental fit. CFI and GFI values range from 0 and 1, with a reference value of 0.90 (Hu & Bentler, 1998). Finally, the error of the root mean square approximation (RMSEA) was calculated. For this index, smaller values indicate better fit, ranging from 0.05 to 0.10. Differences between the models were assessed by comparing the χ^2 values to check for significance.

Measurement invariance for gender

Analyses of multiple-group invariance were conducted to determine the extent to which the factor structure was comparable across gender. Four levels of measurement invariance were sequentially tested (configural, weak, strong, and strict invariance), where each level introduced more equality constraints across groups. The fit of the restricted model (equal factor loadings across groups) and the free model were compared in terms of their χ^2 values. A non-significant increase in the χ^2 value (relative to df) in the constrained model relative to the unconstrained model indicated that the constraints across groups were possible. As an additional criterion, the change in the CFI coefficient was considered. If the drop in CFI of the constrained model relative to the unconstrained model did not exceed 0.01, the constrained model was accepted.

Internal consistency

A univariate descriptive data analysis was performed to explore how the variables were presented. We assessed the characteristics of the items, and calculated the means, standard deviations (SD), and item-total correlations for each item on the SDS and the PHQ-9. Internal consistency was evaluated using the ordinal Cronbach's omega (ω) for the SDS and Cronbach's alpha (α) for the PHQ-9. The average variance extracted (AVE) and composite reliability (CR) were also tested. Data from the SDS regarding the presence or not of major depression (based on the PHQ-9 diagnostic algorithm) are described, together with the depressive symptoms and severity (based on the following PHQ-9 range for symptom severity: no symptoms, mild, moderate, moderately severe, or severe symptoms).

Discriminative validity

A one-way analysis of covariance (ANCOVA) was implemented to determine whether or not differences in 1) the presence or not of depression and 2) the severity of depressive symptoms had an impact on the degree of disability and on the SS score. The post hoc comparisons were adjusted using Bonferroni correction (alpha level for the post hoc tests was set at 0.05). Effect sizes (Cohen's d) were also determined. For symptom severity, the effect sizes were compared with the previous symptoms sample to determine the effect on an increase in symptom severity.

Predictive validity of the SDS to detect major depression and depressive symptoms

We conducted a receiver operating characteristic (ROC) curve analysis to determine the area under the curve (AUC) and the range of cut-off scores for the SDS to assess the utility of the scale to discriminate between patients with and without MDD and to discriminate between patients with different levels of symptom severity. We also calculated the sensitivity, specificity, positive and negative predictive value, positive and negative likelihood ratios and the Youden’s Index (*J*).

Results

Factor structure analysis

Table 2 present the CFA results for all models. Model 2 presented a good fit ($\chi^2 = 40.06$; $p < 0.05$; CFI = 0.99; IFI = 0.99; and RMSEA = 0.10). Model 2 (SDS-4) is a unifactorial model containing four items: the three life domains (work, family, and social life) plus the PS. Model 3 (SDS-5) also presented a good model fit. However, the SS item did not present a significant factor loading, suggesting that model 2 (SDS-4) is the best model fit. No differences between models were found. See Table 2 and Figure 1 for more information about model fit.

Measurement invariance across gender

To determine whether gender affected the measurement model, the sample was divided into males and females and constraints were introduced to test if the parameters could be constrained across both groups. Both models were compared in terms of $\Delta\chi^2$ and by Δ CFI. All levels of invariance up to strict invariance could be assumed across gender, as evidenced by a non-significant drop in model fit (Δ CFI 0.01) for the successively more constrained models. See Table 3.

Internal consistency

After obtaining the one-factor solution (SDS-4) described above, we performed several descriptive analyses to evaluate the internal consistency of this SDS-4. Scores on this four-item scale range from 0-40, with a mean (SD) of 18.4 (10.4); corrected item–total correlations ranged from 0.56 to 0.67. The mean SS score was 53.23 (32.74), with a corrected item–total correlation of 0.042, indicating a low correlation with the other items. Internal consistency for the SDS-4 was acceptable ($\omega = 0.84$), with a good AVE (0.86) and CR (0.93). The mean score for the PHQ-9 items was 13.3 (8.2). The corrected item–total correlations ranged from

0.52 to 0.73 on this scale, with very good internal consistency ($\alpha = 0.87$).

Mean PHQ-9 scores as a function of depressive diagnosis

In patients who met PHQ-9 criteria for MDD, the mean (SD) scores on this tool were 19.0 (3.72), compared to 8.40 (3.72) for those without a diagnosis of MDD. This difference was significant on the one-way ANCOVA [$F(1,1) = 3438.95$; $p < 0.0001$]. On the post hoc analysis, patients with MDD scored significantly higher on the PHQ-9 than patients without MDD with a large effect size ($d = 2.84$). See Table 4 for more details.

Mean PHQ-9 scores according to depressive symptom severity

Mean PHQ-9 scores presented significant differences for all symptom severity categories [$F(1,4) = 6557.137$; $p < 0.0001$], with greater severity associated with higher PHQ-9 scores and very large effect sizes (> 3) when symptom severity was compared with previous symptoms. Table 4 shows all values.

Discriminative validity of the SDS-4 in patients with and without MDD

The mean scores on the SDS-4 were evaluated in patients with and without MDD using the PHQ-9 diagnostic algorithm. As Table 6 shows, the mean (SD) SDS-4 values were 23.4 (9.3) for patients with MDD versus 14.15 (9.3) in those without MDD. Post hoc analyses of the ANCOVA conducted to compare the two groups showed that these differences were significant [$F(1,1) = 417.313$; $p < 0.0001$]. Patients with MDD had significantly higher disability scores on the SDS-4 than patients without MDD, with a large effect size ($d = 0.99$). The internal consistency of the PHQ-9 for the SDS-4 was acceptable for patients with MDD (0.77) and good for patients without a diagnosis of depression (0.83). See Table 4 for more details.

Discriminative validity of the SDS-4 in patients with MDD and different degrees of symptom severity

We found that SDS-4 scores increased as a function of symptom severity [$F(1,4) = 173.516$; $p < 0.0001$]. Post hoc analyses showed that there were significant differences at all levels, indicating that patients with more severe depressive symptoms presented greater disability. In this case, medium effect sizes were found when disability scores were compared with previous symptoms. Internal consistency ranged from acceptable to good (0.74-0.81). All values for the SDS-4 are shown in Table 4.

Table 2
CFA adjustment indices

	χ^2	df	CFI	GFI	TLI	RMSEA (90% IC)	χ^2 dif	AIC
Model 1 (3 items)	174.70*	2	0.94	0.94	0.86	0.23 (.19 - .25)		182.70
Model 2 (4 items)	40.06*	2	0.99	0.99	0.96	0.10 (.08 - .14)	134.64	56.06
Model 3 (5 items)	58.93*	5	0.98	0.99	0.96	0.08 (.06 - .10)	115.77	78.93
Model 4 (2 factors)	39.08*	4	0.98	0.99	0.96	0.08 (.06 - .10)	135.62	61.08

* $p < 0.05$

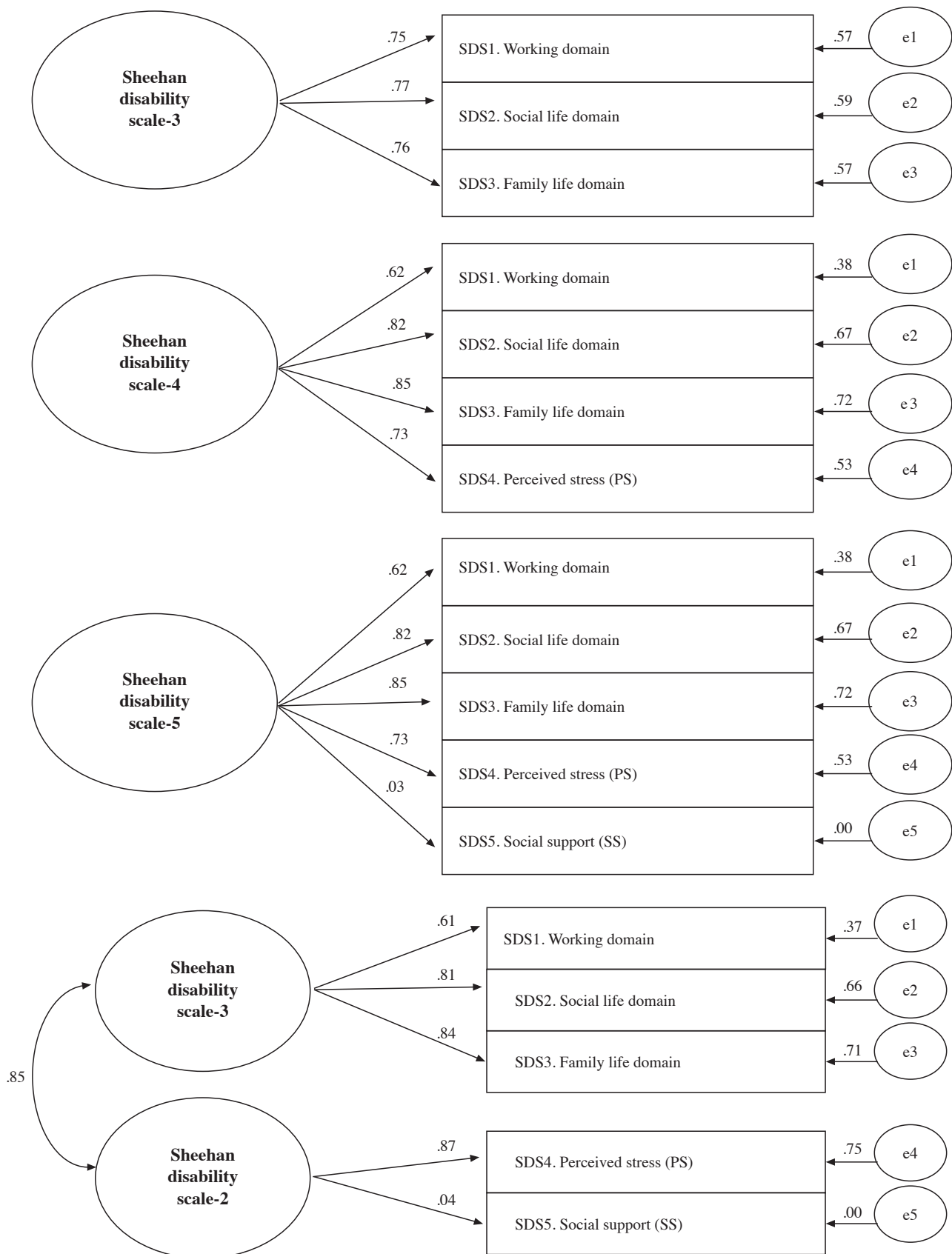


Figure 1. Models fitness and representation

Table 3
Gender Invariance of the SDS-4

	χ^2	df	CFI	TLI	RMSEA	Δ CFI	$\Delta \chi^2$	Δ df
Configural Invariance	41.94	4	0.98	0.96	0.075	–	–	–
Weak Invariance	42.79	7	0.98	0.97	0.055	0.00	0.85	3
Strong Invariance	42.80	8	0.98	0.98	0.051	0.00	0.01	1
Strict Invariance	59.95	12	0.98	0.98	0.048	0.00	17.15	4

Table 4
Values for discriminative validity according presence or not of MDD and different severity symptoms

PHQ-9 criteria	PHQ-9					SDS-4					SS item			
	n	M	SD	p	d	M	SD	α	p	d	M	SD	p	d
PHQ-9 without major depression	914	8.40	3.72			14.2	9.3	0.83			54.9	32.7		
PHQ-9 with major depression	790	19.00	3.72	0.0001	2.84	23.4	9.3	0.77	0.0001	0.99	51.3	32.7	.04	-0.11
PHQ-9 without symptoms (0-4)	150	2.7	1.32			7.6	7.6	0.81			56.8	35.8		
PHQ-9 mild symptoms (5-9)	394	7.2	1.35	0.0001	3.33	13.0	8.7	0.81	0.0001	0.64	56.4	32.7	n.s	-0.01
PHQ-9 moderate symptoms (10-14)	426	11.9	1.44	0.0001	3.41	17.9	8.6	0.75	0.0001	0.57	52.2	31.3	n.s	-0.13
PHQ-9 moderate severe symptoms (15-20)	513	17.7	1.93	0.0001	3.39	22.8	9.3	0.78	0.0001	0.55	52.0	32.0	n.s	-0.01
PHQ-9 severe symptoms (21-27)	221	23.8	1.62	0.0001	3.25	26.4	8.7	0.74	0.0001	0.40	49.9	34.4	n.s	-0.06

Relationship between the SS item, depression diagnosis, and symptom severity.

Patients without MDD on the PHQ-9 had a mean (SD) score on the SS item of 54.9 (32.7) versus 51.3 (32.7) in those who met PHQ-9 criteria for MDD, with a significant difference (one-way ANOVA: $[F(1,1) = 3.373; p < 0.05]$). Post hoc analyses showed that patients with MDD reported less social support than patients without MDD, but with a low effect size ($d = -0.11$). In terms of the association between SS item scores with different symptom severity levels, we found no significant differences within any level ($p > 0.05$). See Table 4 for more details.

Predictive validity of the SDS-4 to discriminate for major depression

The accuracy of the SDS to discriminate between cases with and without MDD was acceptable (AUC = 0.76, standard error (SE) = 0.012, $P < 0.0001$, 95% CI, 0.73–0.78). Based on the ROC analysis, a cut-off score of 19 yielded the best trade-off between sensitivity and specificity (Youden’s index) for the SDS-4 ($J = 0.39$). Also, a cut-off score of 16 also yielded a good Youden’s index for the SDS-4 ($J=0.38$). See in table 5 all the values of sensitivity, specificity, positive and negative predictive values, and positive and negative likelihood ratios.

Predictive validity of the SDS-4 to discriminate for mild depressive symptoms

On the ROC analysis, the AUC of the SDS-4 was 0.83 (SE = 0.017; $P < 0.0001$; 95% CI, 0.80–0.86), indicating a good capacity to detect mild depressive symptoms. The best Youden’s index score ($J = 0.52$) suggested a score of 12 as the optimal cut-off point but also a cut-off score of 9 yielded a good Youden’s index for the

SDS-4 ($J = 0.50$). Table 6 shows other possible cut-off points and CIs.

Predictive validity of the SDS-4 to discriminate for moderate depressive symptoms

The ROC analysis showed that the AUC of the SDS was 0.78 (SE = .012, $P < .0001$, 95% CI, 0.76–0.81), indicating a good capacity to detect moderate depressive symptoms. The best Youden’s index ($J = .44$) suggested a cut-off score of 16 as the optimal cut point but a cut-off score of 14 offered a good Youden Index for the SDS-4 ($J = .43$). All differences in this ROC values can be observed in Table 7.

Predictive validity of the SDS-4 to discriminate for moderately severe depressive symptoms

The AUC for the SDS was 0.77 (SE = 0.011, $P < 0.0001$, 95% CI 0.74-0.79), indicating a good capacity to detect moderately

Table 5
Operational characteristics of the SDS for major depression

Cut-off score	SE	SP	+ PV	- PV	+ LR	- LR	J
SDS-4 \geq 15	0.79	0.56	0.61	0.76	1.79	0.56	0.35
SDS-4 \geq 16	0.77	0.61	0.63	0.75	1.96	0.51	0.38
SDS-4 \geq 17	0.74	0.63	0.63	0.74	2.01	0.50	0.37
SDS-4 \geq 18	0.70	0.67	0.65	0.72	2.11	0.47	0.37
SDS-4 \geq 19	0.68	0.70	0.67	0.72	2.30	0.43	0.39
SDS-4 \geq 20	0.63	0.74	0.68	0.70	2.44	0.41	0.37

Abbreviations: Sensitivity (SE), specificity (SP), positive predictive value (+ PV) and negative predictive value (- NP), positive likelihood ratios (+ LR) and negative likelihood ratios (- LR) and the Youden’s Index (J)

severe depressive symptoms. The best Youden's index ($J=0.39$) suggested a cut-off score of 19 as the optimal cut-off point for the SDS-4, but a cut-off score of 16 on the SDS-4 provided an equally good Youden's index ($J=0.39$). See Table 8 for other possible cut points and CIs.

Predictive validity of the SDS-4 to discriminate for severe depressive symptoms

On the ROC analysis, the AUC for the SDS-4 was 0.75 (SE = 0.017, $P < 0.0001$, 95% CI 0.72–0.79), indicating a good capacity to detect severe depressive symptoms. The best Youden's Index ($J = 0.38$) suggested a cut-off score of 19 and 22 on the SDS-4. Table 9 shows the differences between ROC values in this provided scores and other possible cut-off points.

Discussion

This study shows that the best model fit for the SDS was model 2, which was comprised of a 4-item scale (SDS-4) that included the assessment of perceived stress. Although model 3 presented a slightly better fit, SS did not significantly load on the model. Consequently, SS was not included in the model but could be administered as an adjunct to assess perceived social support. The 4-item SDS model differs from most previous studies, which have

Table 6
Operational characteristics of the SDS for mild depressive symptoms

Cut-off score	SE	SP	+PV	-PV	+LR	-LR	J
SDS-4 ≥ 8	0.84	0.65	0.96	0.27	2.36	0.42	0.48
SDS-4 ≥ 9	0.81	0.69	0.96	0.26	2.58	0.39	0.50
SDS-4 ≥ 10	0.78	0.71	0.96	0.23	2.64	0.38	0.48
SDS-4 ≥ 11	0.75	0.76	0.97	0.23	3.14	0.32	0.51
SDS-4 ≥ 12	0.73	0.79	0.97	0.22	3.52	0.28	0.52
SDS-4 ≥ 13	0.71	0.81	0.97	0.21	3.65	0.27	0.51

Table 7
Operational characteristics of the SDS for moderate depressive symptoms

Cut-off score	SE	SP	+PV	-PV	+LR	-LR	J
SDS-4 ≥ 11	0.84	0.57	0.80	0.62	1.93	0.52	0.40
SDS-4 ≥ 12	0.81	0.60	0.81	0.60	2.04	0.49	0.42
SDS-4 ≥ 13	0.80	0.63	0.82	0.59	2.13	0.47	0.42
SDS-4 ≥ 14	0.76	0.66	0.83	0.57	2.27	0.44	0.43
SDS-4 ≥ 15	0.74	0.68	0.83	0.55	2.33	0.43	0.42
SDS-4 ≥ 16	0.71	0.73	0.85	0.54	2.63	0.38	0.44

Table 8
Operational characteristics of the SDS for moderate severe depressive symptoms

Cut-off score	SE	SP	+PV	-PV	+LR	-LR	J
SDS-4 ≥ 14	0.83	0.53	0.57	0.80	1.75	0.57	0.36
SDS-4 ≥ 15	0.81	0.55	0.57	0.79	1.79	0.56	0.36
SDS-4 ≥ 16	0.79	0.60	0.60	0.79	1.97	0.51	0.39
SDS-4 ≥ 17	0.76	0.63	0.61	0.77	2.02	0.49	0.38
SDS-4 ≥ 18	0.72	0.66	0.62	0.76	2.12	0.47	0.38
SDS-4 ≥ 19	0.70	0.69	0.63	0.75	2.26	0.44	0.39

Table 9
Operational characteristics of the SDS for moderate severe depressive symptoms

Cut-off score	SE	SP	+PV	-PV	+LR	-LR	J
SDS-4 ≥ 18	0.81	0.54	0.21	0.95	1.75	0.57	0.34
SDS-4 ≥ 19	0.81	0.57	0.22	0.95	1.89	0.53	0.38
SDS-4 ≥ 20	0.75	0.62	0.23	0.94	1.95	0.51	0.37
SDS-4 ≥ 21	0.72	0.64	0.23	0.94	2.00	0.50	0.36
SDS-4 ≥ 22	0.71	0.67	0.24	0.94	2.17	0.46	0.38
SDS-4 ≥ 23	0.67	0.71	0.26	0.94	2.30	0.44	0.38

consistently found a unifactorial structure containing only the three life domain items (work, family, and social life) while the other two items are considered independent (Arbuckle et al., 2009; Hambrick et al., 2004; Hodgins, 2013; Leon et al., 1992; Leon et al., 1997; Luciano et al., 2010). For instance, Luciano et al. (2010) found a good model fit for the three-item scale in Spanish primary care, except for the chi-squared GFI statistic, which was statistically significant. By contrast, we found that the addition of PS in this disability assessment tool yielded a good model fit; moreover, this model was invariant across gender, further supporting its value in the primary care setting in Spain and, potentially, other regions.

In the PC setting, the prevalence of depression increased due to the lingering effects of the global financial crisis that began in 2008/2009, especially among individuals who perceive social instability as a stressor (Navarro-Mateu et al., 2015). A large body of research has associated stress with depression and/or burnout (Bianchi, Schonfeld, & Laurent, 2015), suggesting that perceived stress should be evaluated and considered in the context of treating mental health disorders. This is especially important because an impaired cognitive capacity to manage stress could be a barrier to emotional wellbeing (Puigbó, Edo, Rovira, Limonero, & Fernández-Castro, 2019).

In terms of internal consistency, the SDS-4 presented a good Ordinal Cronbach's alpha ($\omega = 0.84$). In addition, when the sample was divided into groups according to the presence or not of MDD or categorized by symptom severity, the SDS-4 had a good internal consistency for all levels. The SDS-4 was acceptable with a good discriminative capacity for all symptom severity levels. All these findings further support the use of the SDS-4 version.

When we compared patients with and without MDD (PHQ-9 diagnostic algorithm), we found that patients with MDD scored significantly higher on the SDS-4 (medium effect size). Scores on the SDS-4 also increased as a function of greater symptom severity (medium effect sizes) when compared to previous symptoms. Moreover, the mean PHQ-9 scores were higher in patients with more severe depressive symptoms (see Table 4), a finding that is consistent with previous research (Luciano et al., 2010; Mitchell et al., 2016).

The mean scores on the SDS-4 were higher in patients with MDD versus those without, indicating a greater level of disability. This finding held true for all levels of symptom severity. Overall, these findings support previous research suggesting that patients with depression who have greater impairment and greater symptom severity have more disability (Friedrich, 2017; World Health Organization, 2017). By contrast, patients without depression reported more social support than patients with depression, but with a very low effect size. We found no association between symptom severity and the degree of social support, suggesting that perceived social support had no effect on symptom severity.

The predictive validity (ROC analysis) of the SDS-4 to discriminate for depression was good, with an optimal cut-off score (according to Youden's index) of 19. Although a cut-off of 16 provided better sensitivity (0.77), this cut point could lead to a higher rate of false positives (39% in our sample; specificity=0.61). Several cut-off scores yielded similar sensitivity and specificity results for mild, moderate, moderately severe, and severe depressive symptoms; however, the most sensitive cut-off points were, respectively, 9, 14, 16 and 19, although all of these cut points also resulted in a high rate of false positives (ranging from 43 to 31%, respectively). Although the SDS-4 is more accurate to detect mild depressive symptoms, its capacity to discriminate for the other levels of depression was acceptable. Consequently, clinicians should select the cut-off scores that most closely align with their clinical objectives (e.g., greater sensitivity versus higher specificity). In short, this new tool could be highly useful in the PC setting, allowing a range of different health care professionals to work together to accurately identify patients with depression-related disability to offer them the best treatment possible in a timely manner (Muñoz-Navarro et al., 2017b).

This study has several limitations, the most being the lack of a gold standard (i.e., structured clinical interview) to definitively diagnose depression. However, previous validity of the Spanish version of the PHQ-9 in PC patients in Spain has shown that the

PHQ-9 is reliable when compared to the gold standard (Muñoz-Navarro et al., 2017a). In this study, the diagnostic algorithm of the PHQ-9 yielded good sensitivity and specificity rates (0.88 and 0.80), which support the use of the PHQ-9 for the present work. Another study limitation is that the convergent validity of the SDS-4 was not assessed, which should be done in future studies.

To conclude, the present study shows that adding an additional item—perceived stress (PS)—to create the SDS-4 improves the psychometric properties of this very brief tool. Based on these findings, the SDS-4 appears to offer clinicians a reliable tool to assess disability in patients with major depression.

Acknowledgements

This work has been supported by the Fundación Española para la Promoción y el Desarrollo Científico y Profesional de la Psicología to Antonio Cano-Vindel; by the Agencia Estatal de Investigación (PSI2012-36589) to Antonio Cano-Vindel; by the IDIVAL, INNVAL (ref.: 16/08) and PRIMVAL (ref.: 18/03) to César González-Blanch; and by the Agencia Estatal de Investigación (SI2014-56368-R) to Juan A. Moriana

We thank all the PsicAP Research Group members who kindly participated in this large project.

References

- Arbuckle, R., Frye, M. A., Brecher, M., Paulsson, B., Rajagopalan, K., Palmer, S., & Degl'Innocenti, A. (2009). The psychometric validation of the Sheehan Disability Scale (SDS) in patients with bipolar disorder. *Psychiatry Research*, *165*(1-2), 163-174. <https://doi.org/10.1016/j.psychres.2007.11.018>
- Bianchi, R., Schonfeld, I. S., & Laurent, E. (2015). Burnout-depression overlap: A review. *Clinical Psychology Review*, *36*, 28-41. <https://doi.org/10.1016/j.cpr.2015.01.004>
- Bobes, J., Badía, X., Luque, A., García, M., González, M. P., & Dal-Ré, R. (1999). Validation of the Spanish version of the Liebowitz social anxiety scale, social anxiety and distress scale and Sheehan disability inventory for the evaluation of social phobia. *Medicina clínica*, *112*(14), 530-538.
- Cano-Vindel, A., Muñoz-Navarro, R., Wood, C. M., Limonero, J. T., Medrano, L. A., Ruiz-Rodríguez, P., ... Santolaya, F. (2016). Transdiagnostic cognitive behavioral therapy versus treatment as usual in adult patients with emotional disorders in the primary care setting (PsicAP Study): Protocol for a randomized controlled trial. *JMIR Research Protocols*, *5*(4), e246. <https://doi.org/10.2196/resprot.6351>
- Fernández, A., Pinto-Meza, A., Bellón, J. A., Roura-Poch, P., Haro, J. M., Autonell, J., ... Serrano-Blanco, A. (2010). Is major depression adequately diagnosed and treated by general practitioners? Results from an epidemiological study. *General Hospital Psychiatry*, *32*(2), 201-209. <https://doi.org/10.1016/j.genhosppsych.2009.11.015>
- Friedrich, M. J. (2017). Depression Is the Leading Cause of Disability Around the World. *Jama*, *317*(15), 1517-1517. <https://doi.org/10.1001/jama.2017.3826>
- González-Blanch, C., Fernando Hernández-de-Hita, Muñoz-Navarro, R., Ruiz-Rodríguez, P., Medrano, L. A., Moriana, J. A., & Cano-Vindel, A. (2018a). Domain-specific associations between disability and depression, anxiety, and somatization in primary care patients. *Psychiatry Research*, *269*, 596-601. <https://doi.org/10.1016/j.psychres.2018.09.007>
- González-Blanch, C., Hernández-de-Hita, F., Muñoz-Navarro, R., Ruiz-Rodríguez, P., Medrano, L. A., & Cano-Vindel, A. (2018b). The association between different domains of quality of life and symptoms in primary care patients with emotional disorders. *Scientific Reports*, *8*(1), 11180. <https://doi.org/10.1038/s41598-018-28995-6>
- González-Blanch, C., Medrano, L. A., Muñoz-Navarro, R., Ruiz-Rodríguez, P., Moriana, J. A., Limonero, J. T., ... Cano-Vindel, A. (2018c). Factor structure and measurement invariance across various demographic groups and over time for the PHQ-9 in primary care patients in Spain. *PLoS ONE*, *13*(2), e0193356. <https://doi.org/10.1371/journal.pone.0193356>
- Hambrick, J. P., Turk, C. L., Heimberg, R. G., Schneier, F. R., & Liebowitz, M. R. (2004). Psychometric properties of disability measures among patients with social anxiety disorder. *Journal of Anxiety Disorders*, *18*(6), 825-839. <https://doi.org/10.1016/j.janxdis.2003.10.004>
- Hodgins, D. C. (2013). Reliability and validity of the Sheehan disability scale modified for pathological gambling. *BMC Psychiatry*, *13*(1), 177-186. <https://doi.org/10.1186/1471-244X-13-177>
- Hu, L., & Bentler, P. M. (1998). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- 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*(9), 606-613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- Leon, A. C., Shear, M. K., Portera, L., & Klerman, G. L. (1992). Assessing impairment in patients with panic disorder: the Sheehan Disability Scale. *Social Psychiatry and Psychiatric Epidemiology*, *27*(2), 78-82. <https://doi.org/10.1007/BF00788510>
- Leon, Andrew 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. <https://doi.org/10.2190/T8EM-C8YH-373N-1UWD>
- Luciano, J. V., Bertsch, J., Salvador-Carulla, L., Tomás, J. M., Fernández, A., Pinto-Meza, A., ... Serrano-Blanco, A. (2010). Factor structure, internal consistency and construct validity of the Sheehan Disability Scale in a Spanish primary care sample. *Journal of Evaluation in*

- Clinical Practice*, 16(5), 895-901. <https://doi.org/10.1111/j.1365-2753.2009.01211.x>
- Mitchell, A. J., Yadegarfar, M., Gill, J., & Stubbs, B. (2016). Case finding and screening clinical utility of the Patient Health Questionnaire (PHQ-9 and PHQ-2) for depression in primary care: A diagnostic meta-analysis of 40 studies. *BJPsych Open*, 2(2), 127-138. <https://doi.org/10.1192/bjpo.bp.115.001685>
- Muñiz, J., & Fonseca-Pedrero, E. (2019). Ten steps for test development. *Psicothema*, 31(1), 7-16. doi:10.7334/psicothema2018.291
- Muñoz-Navarro, R., Cano-Vindel, A., Medrano, L. A., Schmitz, F., Ruiz-Rodríguez, P., Abellán-Maeso, C., ... Hermosilla-Pasamar, A. M. (2017a). Utility of the PHQ-9 to identify major depressive disorder in adult patients in Spanish primary care centres. *BMC Psychiatry*, 17(1), 291-300. <https://doi.org/10.1186/s12888-017-1450-8>
- Muñoz-Navarro, R., Cano-Vindel, A., Ruiz-Rodríguez, P., Adrián Medrano, L., González-Blanch, C., Moriana, J. A., ... Dongil-Collado, E. (2017b). Diagnostic and referral hierarchical model of common mental disorders in primary care centers. The PsicAP clinical trial approach. *Ansiedad y Estrés*, 23(2-3), 124-129. <https://doi.org/10.1016/j.anyes.2017.10.002>
- Navarro-Mateu, F., Tormo, M. J., Salmerón, D., Vilagut, G., Navarro, C., Ruiz-Merino, G., ... Alonso, J. (2015). Prevalence of mental disorders in the South-East of Spain, one of the European regions most affected by the economic crisis: The cross-sectional PEGASUS-Murcia project. *PLoS ONE*, 10(9), e0137293. <https://doi.org/10.1371/journal.pone.0137293>
- Puigbó, J., Edo, S., Rovira, T., Limonero, J. T., & Fernández-Castro, J. (2019). Influence of the perceived emotional intelligence on the coping of the daily stress. *Ansiedad y Estrés*, 25(1), 1-6. <https://doi.org/10.1016/j.anyes.2019.01.003>
- Roca, M., Gili, M., García-García, M., Salva, J., Vives, M., García Campayo, J., & Comas, A. (2009). Prevalence and comorbidity of common mental disorders in primary care. *Journal of Affective Disorders*, 119(1-3), 52-58. <https://doi.org/10.1016/j.jad.2009.03.014>
- Ruiz-Rodríguez, P., Cano-Vindel, A., Muñoz Navarro, R., Medrano, L., Moriana, J. A., Buiza Aguado, C., ... González-Blanch, C. (2017). A systematic review and critique of the economic impact and burden of common mental disorders in Spain. *Ansiedad y Estrés*, 23(2-3), 118-123. <https://doi.org/10.1016/j.anyes.2017.10.003>
- Sheehan, D. V., Harnett-Sheehan, K., & Raj, B. A. (1996). The measurement of disability. *International Clinical Psychopharmacology*, 11, 89-95.
- World Health Organization (2017). *Depression and other common mental disorders: Global health estimates* (No. WHO/MSD/MER/2017.2). World Health Organization. <https://doi.org/10.1002/mas>
- World Health Organization (2019). Disabilities. Retrieved from <https://www.who.int/topics/disabilities/en/>