Use of the long and short forms of the Depression in the Medically Ill questionnaire in Spanish population

https://doi.org/10.1177/1073191110397273

ABSTRACT

This study aimed to translate the original (DMI-18) and the short (DMI-10) version of the Depression in the Medically Ill into Spanish, validate them and determine the best cut-off points for detecting depression. A back-translation procedure was used. Patients with somatic disorders (N=366) completed the translated DMI-18 and another depression questionnaire. Of these, 167 were also assessed by a mental health professional. Reliability (Cronbach alpha > 0.90) and convergent validity (r > 0.74) were satisfactory. The CFA results supported the one factor model (depression). The best cut-off of the Spanish version was 15 for the DMI-18 and 9 for the DMI-10. Sensitivity and specificity were 93% and 73% for the DMI-18 and 87% and 74% for the DMI-10. Our data confirm the validity of the Spanish-language versions of both DMI versions. Their use in a clinical context may help non-psychiatric professionals to detect affective comorbidities in their patients.

Keywords: depression, patients, Spain, psychometrics, validation studies.

INTRODUCTION

Depression is common among patients with somatic disorders, with a prevalence of the order of 20% (Kilbourne, Daugherty, & Pincus, 2007; Parker et al., 2006), and is associated with a poorer prognosis and reduced quality of life (Adelman, Greene, Friedmann, & Cook, 2008; Herrero et al., 2003). However, depression often goes undiagnosed and untreated in these patients, in part because symptoms of depression such as apathy, fatigue and disturbed sleep are similar to those of many somatic conditions (Herrero et al., 2003). Overlooking depression is a lost opportunity for improving quality of life, reducing hospital stays, improving treatment adherence, and reducing the risk of suicide in patients with somatic disorders (Koening, 1997; Newport & Nemeroll, 1998).

An instrument for detecting depression that is simple to administer, easy to grade, and specifically designed for patients with somatic disorders would be very helpful for clinicians (Bambauer, Locke, Aupont, Mullan, & McLaughlin, 2006; Beck, Steer, Ball, Ciervo, & Kabat, 1997; Diez-Quevedo, Rangil, nchez-Planell, Kroenke, & Spitzer, 2001; Herrero et al., 2003; Kroenke, Spitzer, & Williams, 2001; Sharp & Lipsky, 2002; Zigmond & Snaith, 1983). Ideally, depression assessment in primary and secondary care should be restricted to variables and items that avoid confounding by medical illness (Parker, Hilton, Bains, & Hadzi-Pavlovic, 2002). Currently, three measures that exclude somatic items exist: the Beck Depression Inventory for Primary Care (BDI-PC) (Beck, Steer, & Brown, 2000); the Hospital Anxiety and Depression Scale (HADS-D) (Herrero et al., 2003) and the Depression in the Medically Ill questionnaire (DMI) (Parker et al., 2002).

The BDI-PC is a short version of the Beck Depression Inventory –II (BDI-II) (Beck et al., 1996), a questionnaire created for assessing the severity of depression in patients with mental disorders and for detecting possible depression in normal populations. Thus, this tool was not originally designed for the detection of affective conditions in

patients with somatic disorders. To the best of our knowledge, only HADS and DMI have been specifically designed for that aim. The HADS is the most commonly used tool for measuring depression in primary care. It is largely based on the symptom of anhedonia amongst all the symptoms of depressive disorders (Herrero et al., 2003). The DMI is a valid measure of depression in the medically ill focusing on cognitive symptoms, and includes not only anhedonia, but all areas central to depression (depressive humour, anhedonia, pessimism, low self esteem, etc.) (Parker et al., 2002; Parker, Hilton, Hadzi-Pavlovic, & Bains, 2001). The problematic nature of anhedonia in primary care patients is that it appears to link with somatic symptomatology (Parker et al., 2002). So far, only English and Chinese versions of the DMIs are available, while other tools, such as HADS, have been translated and validated into several languages.

In summary, the DMI provides a simple, easy and "cognitive" measure of depression, based on the belief that excluding somatic items overcomes the confounding effects of the medical illness.

The aim of this study was to translate and adapt into Spanish both versions of the DMI, as well as to test their psychometric characteristics, internal structure, sensitivity, specificity and optimal cut-off points. Their usefulness as screening instruments for depression was evaluated by comparing their diagnostic performance against the diagnosis made by the Primary Care Evaluation of Mental Disorders (PRIME-MD) (Spitzer et al., 1994) structured clinical interview. Confirmatory Factor Analyses and Known group comparisons were performed.

METHODS

Subjects

The study sample included 366 patients who were admitted to one of several services of Galdakao-Usansolo Hospital or came in for a consultation to our outpatient clinics between November 2007 and April 2008. Galdakao-Usansolo Hospital is a 400-bed general hospital with a coverage area of 300,000 inhabitants. The hospital covers an area, called Comarca Interior, situated in the Basque Country (northern Spain), and has a mixed urban, suburban and rural population of 300.299. The age distribution, education level, sources of employment, socioeconomic status, and health care services of the urban population are representative of the overall Basque Country (Instituto Vasco de Estadística, 2006). Health care in this region is provided by the public network of the Basque Health Care Service-Osakidetza, which provides free unrestricted care to nearly 100% of the population. Patients were drawn from the pain, obstetrics and gynaecology, endocrinology, gastroenterology, neurology, pneumology, and psychiatry units. The majority of the participants had a European background and 5 were from South America. Patients were included in the study if they were between 18 and 85 years of age, were evaluated at the hospital for a somatic illness, spoke Spanish, and agreed to collaborate in the study after being briefed about the study and its voluntary nature. Patients were excluded if they, at the physician's discretion, had a severe physical disease, cognitive deterioration, any brain disease, or a psychotic disorder that might have compromised their ability to participate in the study by filling in the questionnaires.

Instruments

Sociodemographic data obtained from the subjects included gender, age, marital status, education level, and employment status. Clinical data collected included length of time with the disease, and any psychiatric medication taken.

The study subjects completed a Spanish-language version of the DMI-18 questionnaire (Parker et al., 2001) that had been translated from English as described below. Items are ranged from "none" (rated 0) to "always" (rated 3). The study subjects also completed one of the following three short, easily administered depression scales, all applied in their Spanish versions:

The Beck Depression Inventory for Primary Care (BDI-PC) (Harcourt Assessment, 2006) consisted of 7 cognitive and affective items, extracted from the 21-item Beck Depression Inventory-II (BDI-II) (Beck et al., 1996; Sanz et al., 2005). It was developed for evaluating symptoms of depression in patients reporting somatic and behavioural symptoms that may be attributable to biological, medical, alcohol, and/or substance abuse problems. Its items evidenced good internal consistency (Cronbach α: 0.92) and showed strong correlation with the Milton Clinical Multiaxial Inventory II (MCMI II) major depression and dysthymia subscales ($r \geq 0.68$) (Penley, Wiebe, & Nwosu, 2003) (Sanz et al., 2005). A cut-off point of 4 is used to define presence of depression (Beck et al., 2000). In this study we applied the Spanish version of the BDI-PC (Harcourt Assessment, 2006).

The Hospital Anxiety and Depression Scale (HADS) (Herrero et al., 2003; Zigmond et al., 1983) was specially designed for identifying and quantifying depression and anxiety in physically ill patients. The HADS is a 14-item measure that includes a 7-item depression subscale (HADS-D) for measuring cognitive and emotional aspects of depression, predominately anhedonia, and a 7-item anxiety subscale (HADS-A) for measuring cognitive and emotional aspects of anxiety. Only the HADS-D subscale was used for the

purposes of the current study. This subscale was translated and validated in to Spanish, and it has a high level of internal consistency (Cronbach α: 0.84), and good psychometric characteristics (specificity: 87%; sensitivity: 72%) (Herrero et al., 2003). The authors of the original questionnaire recommended two cut-off scores: 7/8 for possible and 10/11 for probable depression (Zigmond et al., 1983).

The Patient Health Questionnaire-9 (PHQ-9) (Kroenke et al., 2001) is the mood module of the Patient Health Questionnaire (PHQ) (Spitzer, Kroenke, Williams, & the Patient Health Questionnaire Primary Care Study Group, 1999), a self-administered version of the Primary Care Evaluation of Mental Disorders (PRIME-MD) (Spitzer et al., 1994). The PHQ-9 consists of 9 items designed to correspond to the nine diagnostic criteria for major depressive disorder covered in the Diagnostic and Statistical Manual for Mental Disorders (American Psychiatry Association, 1994; Wilhelm, Kotze, Waterhouse, Hadzi-Pavlovic, & Parker, 2004) including somatic symptoms like fatigue, insomnia, and anorexia. Items are rated from 0 to 3 according to increased frequency of experiencing difficulties in each item. In this study we used the Spanish validation of the PHQ-9, which has a sensitivity of 89%, a specificity of 87% and it correlates strongly with the total BDI score (r=0.76) (Diez-Quevedo, Rangil, Sánchez-Planell, Kroenke, & Spitzer, 2001). The author of the questionnaire (Kroenke et al., 2001) specified 4 cut-off points: 5, 10, 15 and 20 that represented mild, moderate, moderately severe, and severe depression, respectively. Here we considered a cut-off points of 10 (>10 severe depression).

The mental health professional (6 psychiatrists and 4 psychologists) performed the psychiatric interview using the mood module of the PRIME-MD structured psychiatric interview in Spanish (Baca et al., 1999). The mood module of the PRIME-MD has nine items that represents the nine DSM-IV depression criteria with dichotomous response

categories (yes/no). The sensitivity of this instrument was found to be equal to 72% and its specificity 86% (Baca et al., 1999).

Procedures

Back translation procedure

Adaptation of the DMI for Spanish speakers was performed using the back translation procedure, which ensures conceptual equivalency (Aaronson et al., 1992; Brislin, 1970).

Translation from English to Spanish was carried out by two independent native Spanish-speaking translators who are proficient in English. The research team systematically reviewed and compared the two translations, arriving at a first consensus version. Two other independent translators, this time whose native language was English and who were fluent in Spanish, back translated the consensus version. After reaching consensus on a final translated version, it was sent to the DMI's original author (Dr. Parker) who compared it to the original version and gave his approval. The final version was administered to two small groups, one made up of patients, the other of clinical experts. After evaluating the results of this intelligibility test, small modifications were made, resulting in a final version of the Spanish DMI-18.

Validation of the translated questionnaire

All patients approached to take part in the study were informed about its nature by their physicians and advised of their informed consent to take part in the study. Patients who chose to participate were recruited by one of our investigators (M.O. or C.L.H.), who emphasized that their participation in the study was voluntary.

Each participant was asked to complete a series of sociodemographic questions, the DMI-18 questionnaire, and one of the other three depression screening questionnaires: HADS, BDI-PC or PHQ-9. Patients did not complete the DMI-10 questionnaire anew. Instead, a DMI-10 score was extracted for each participant from the respective items in the DMI-18. Regarding HADS, the participants completed all 14 items, but for the purpose of this article only the depression subscale (HADS-D) items were taken into account. Given that completing all battery of tests would be tiring for the patients, we originally aimed for the third of the sample to complete the HADS, a third the BDI-PC and another third the PHQ-9. Questionnaires were handed to consecutive patients, until the intended quota was approximately achieved.

After completing these materials, a mental health professional who was blinded to the results of the DMI-18, conducted on our behalf a psychiatric interview of a subset of the participants. This subset was determined by the mental health professional's and the patients' availability. The mental health professionals were based on the PRIME-MD, their expertise and the DSM-IV to determine a diagnosis of depression. The inter-rater reliability of the interviewers was assessed by reviewing videotapes of 10 patient interviews originally performed by a clinical professional with wide experience in diagnosing depressive disorders (JAP). These patients did not take part in the validation study. Inter-rater reliability was estimated with the kappa coefficient considering two classification categories (major depression yes or not). The minimum kappa value accepted was set at 0.61 (Fayers & Machin, 2007a; Viera & Garrett, 2005). A total of 10 evaluators met this requirement. Six of them obtained a κ value of 0.67, for 2 κ was equal to 0.83 and for the rest 2 κ was equal to 1.

All study procedures were approved by the institutional review board of our hospital.

Statistical analyses

In order to estimate the sample size for the predictive precision study, we assumed a depression rate of 30%. Interviewing 170 patients with the PRIME-MD we would expect to estimate a sensitivity of 85% with a 95% CI of ± 10% and a specificity of 70% with a 95% CI of ± 8% (Mulrow et al., 1995).

In order to validate the questionnaire in Spanish, we used the same procedures that were used to validate the original DMI-18, examining the reliability, construct validity and criteria validity.

Reliability: Cronbach's α and average inter-item correlation were calculated for examining the internal consistency of the DMI-18 and the DMI-10. A coefficient α greater than or equal to 0.7 is considered acceptable (Hair, Anderson, Tatham, & Black, 1998; Nunnally & Bernstein, 1994).

Structural validity: Confirmatory Factor Analysis (CFA) for the DMI-18 and DMI-10 were performed, using the Unweighted Least Squares estimation method (ULS). The following fit indexes are reported: Goodness of Fit-index (GFI), GFI Adjusted for degrees of freedom (AGFI), Root Mean Square Residual (RMR) and Standardized Root Mean Square Residual (SRMR). Values of GFI > 0.9, AGFI > 0.8, and a RMR and SRMR < 0.1 indicate a good model fit (Cole, 1987). Finally, only items with a statistically significant factor loading ($p<0.05$) and with loading values >0.40 were accepted.

Convergent validity: convergence between the new instruments scores (DMI-18 and DMI-10) and the established validated questionnaires scores (HADS-D, BDI-PC and PHQ-9), was examined implementing Pearson's correlation coefficient. 95% CI

were also calculated. Correlations of < 0.50, 0.50-0.80, > 0.80 were defined as weak, moderate and strong, respectively (Hatcher, 1994).

Known-groups validity: DMI mean score differences between the different categories established for the HADS-D, BDI-PC and PHQ-9 according to depression severity were evaluated with a one-way analysis of variance (ANOVA). The t-test was used for two group comparisons. The non parametric Jonckheere-Terpstra (JT) test for trend was also implemented for HADS-D and PHQ-9, and the Wilcoxon Rank-Sum test for BDI-PC.

Criterion validity: the primary objective of the DMI-18 and DMI-10 is to detect depression in secondary and tertiary care patients. To this end, the PRIME-MD structured clinical interview was used as a gold standard for depression. The accuracy and predictive validity of the DMI were analyzed using sensitivity, specificity and positive and negative predictive values. The estimates were presented with a 95% Confident Interval (CI). In order to determine the optimal cut-off point for sensitivity and specificity, two ROC curves were created, one for the DMI-18 and the other for the DMI-10 (Murphy et al., 1987). The most appropriate cut-off point was considered to be that which minimized the sum of false positives and false negatives (Herrero et al., 2003). The area under the curve (AUC) represents the ability of the questionnaire to discriminate between cases and non-cases. AUC values between 0.5 and 0.7 indicate poor discrimination, values between 0.7 and 0.9 indicate average discrimination and values greater than 0.9 indicate a high discrimination (McDowel & Newell, 1996).

Finally, the chi square test was used for comparisons of categorical variables.

Statistical analyses were carried out using SAS version 9.2 for Windows. The receiver operating characteristic (ROC) curves were plotted in SPSS version 16.

RESULTS

Evaluation of the first version of the questionnaire

A total of 11 native Spanish speaking patients who were diagnosed with depression completed the DMI-18. The average time needed was 4 minutes (SD: 2 minutes). The length of the questionnaire was considered to be adequate by 7 of the participants, 3 considered it to be short, and 1 response was lost. All patients considered the questions to be adequate for detecting depression. There also was a meeting with three mental health professionals, experts in diagnosing affective disorders who, after recommending some small modifications, considered the questionnaire to be simple and sufficient both in its format, length, and objective of detecting depression.

Validation of the translated questionnaire

Sample description

A total of 366 patients were recruited for the study. Twelve were excluded for not answering more than 9 items of the DMI, leaving a cohort of 354 patients. Of them, 125 patients completed the BDI-PC, 97 patients completed the HADS, and 130 completed the PHQ-9. Of the total sample, 167 were interviewed by the PRIME-MD structured clinical interview. Statistically significant differences between interviewed and non interviewed patients were found only for the variables of gender and psychiatric drug use. Sociodemographic data of the entire sample are shown in Table 1.

----------Table 1 here----------

Missing data treatment

Missing values were imputed using the mean imputation method. This consists of substituting the missing response in an item for the mean of the responses that the subject provided on the rest of his or her items (Fayers & Machin, 2007b). The mean imputation was performed whenever more than 50% of the items had been sufficiently answered. All of the 46 missing values found were imputed. More specifically: 41 subjects did not answer 1 item, 4 subjects did not answer 2 items, and 1 subject did not answer 3 items. No missing values pattern was observed.

Reliability of the DMI-18 and the DMI-10

The Cronbach's α and the average inter-item correlation for the DMI-18 were 0.95 and 0.54, and for the DMI-10 0.91 and 0.52 respectively.

Validity of the DMI-18 and the DMI-10

Based on the original DMI study (Parker et al., 2002), two hypothetical models were tested: a) a two-factor solution model (Model 1) in which the items were divided into two intercorrelated dimensions (cognitive and affective); and b) a one factor solution model (Model 2), with depression being the single latent factor.

Both models were supported by our data. The derived goodness-of-fit indexes were highly satisfactory (Table 2) and factor loadings were elevated and statistically significant (p <0.001). In model 1 the correlation between two dimensions was very high (r = 0.91). Factor loadings and error variances are presented in Table 2.

----------Table 2----------

The convergent validity results between the different outcomes measures used in the study are shown in Table 3. Correlations between both DMI questionnaires and the rest of the questionnaires were moderate to strong, being greater than 0.70.

Mean (standard deviation: SD) scores of the DMI-18 and DMI-10 in three ordered HADS-D and BDI-PC and two PHQ-9 categories according to depression severity were calculated (Table 3). In all cases the DMI scores increased along the ordered categories of the other questionnaires and their values differed significantly among groups ($p<0.0001$).

----------Table 3 here----------

The criterion validity of both questionnaires was assessed by examining the DMI scores of the 167 patients evaluated by the mental health professionals. At this stage, the predictive precision of the DMI-18 and DMI-10 were tested with the ROC curves (Figure 1). Both versions of the DMI questionnaire demonstrated a high discriminatory ability in distinguishing between depressed and non depressed patients, with the AUC of DMI-18 being 0.90 (95%CI: 0.85-0.94) and of DMI-10 being 0.89 (95%CI: 0.84-0.94). For the DMI-18 three different cut-off points, among which is found the cut-off value of 20 from the original validation study, are presented in Table 4. These data suggest that a cut-off point of 15 in the DMI-18 minimizes the sum of false positives and false negatives, correctly classifying 83.8% of the sample. For the DMI-10, among 3 cut-off points evaluated, the value with the best balances between sensitivity and specificity is 9, coinciding with the original. Therefore, the DMI-10 had an accuracy of 80.8%. These data, together with the psychometric values from the questionnaire, are shown in Table 4.

----------Figure 1 here----------

308 ----------Table 4 here----------

309

DISCUSSION

Several studies have reported that depression is a leading comorbidity which affects many patients. Nevertheless, to detect it is not an easy task. In a clinical context, it is important to have effective tools that may allow clinicians, especially family physicians but also other clinicians apart from psychologists and psychiatrists, to quickly and safely screen and refer to a mental health specialist those patients for whom depression is suspected. This study, with a large sample of patients with different somatic diseases, using several others mental health questionnaires and counting with the final evaluation of mental health professionals as a gold standard, tries to show the value of the DMI-18 and DMI-10. The DMI-18, and further the DMI-10, provide with an easy to use and quickly to respond tool to asses depression in all kind of patients.

The translation method used ensures conceptual equivalence, which is supported by the high level of concordance between the original version and the one obtained using back-translation.

The internal consistency of the translated DMI-18 and DMI-10 was satisfactory, with a Cronbach's α value greater than 0.9 and an average inter-item correlation coefficient greater than 0.5. This indicate that both questionnaires can be used reliably for individual patients (McHorney & Tarlov, 1995).

Correlations between both DMI questionnaires and HADS-D, BDI-PC and PHQ-9 were as expected, and indicate satisfactory convergent validity for the two new instruments. Increasing HADS-D, PHQ-9 and BDI-PC scores coincided with increasing DMI scores.

With regards to known groups validity, both questionnaires (DMI-18 and DMI-10) detected statistically significant differences between the categories established for the HADS-D, BDI-PC and PHQ-9 questionnaires to detect depression severity.

With respect to criteria validity, both versions of the translated DMI adequately discriminated between depressed and non-depressed patients. Both the sensitivity (DMI-18: 93%; DMI-10: 87%) and the specificity (DMI-18: 73%; DMI-10: 74%) surpassed the minimally acceptable levels (84% for the sensitivity; 72% for the specificity) found by Mulrow et al. (Mulrow et al., 1995) in analysis of 9 widely-used instruments for the detection of depression in primary care settings.

In the original DMI study, Parker et al. (Parker et al., 2002) recommend using a cut-off point of 20 for the DMI-18 and 9 for the DMI-10. In our study, the best cut-off point for detecting depression was 15 for the DMI-18 and 9 for the DMI-10. The choice of given a cut-off point represents a compromise between sensitivity and specificity. Increasing the cut-off point generally achieves greater specificity while reducing sensitivity. If our aim is to identify the greatest number of individuals with depression, even at the cost of obtaining a higher number of false positives, we should choose a lower cut-off point. But if we can provide the intervention only to a smaller group of patients with serious disorders, a higher cut-off point would be more appropriate (Herman, 1997).

To evaluate the DMI's internal structure, the original authors performed an exploratory factor analysis (EFA) of 23 items, from which three factors were extracted, grouped into cognitive, somatic, and anxiety items. In order to develop the final questionnaire, the authors eliminated the somatic items, after which no structural analysis was performed. Even though no CFA was performed, Parker, seemed to hypothesize that all items load on two factors (cognitive and anxiety), but even so, in

the end he opted for having just one total score, summing up all 18 items. We performed both a two-factor and a single factor CFA on the DMI-18, following Parker´s final hypothesis. Both hypothetical models turned out to be equally acceptable. However, we decided to retain the one-factor structure, because of the high correlation among the two factors and also for being the most parsimonious and theoretically meaningfully solution. Only the single factor solution was fitted to the DMI-10. The CFA models have demonstrated their ability both, when validating a new questionnaire as well as when adapting a questionnaire developed in another language. Currently, CFA is known to correct the deficiencies that are inherent to EFA (Batista-Foguet, Coenders, & Alonso, 2004; Hatcher, 1994).

The factor analyses results confirmed the uni-dimensional structure that the original authors seemed to have hypothesized for both questionnaires (DMI-18 and DMI-10), showing a good fit (Parker et al., 2002).

We agree with the original authors (Parker et al., 2002) in recommending the use of the abridged version of the DMI (DMI-10), given that the analysis shows very little loss of the ability to discriminate depression between the DMI-10 and the DMI-18.

A possible limitation of this work is the fact that, in the study population, patients with major depression were underrepresented and the majority of the sample is made up of patients who scored near the cut-off for depression. In other words, the study included a number of patients for whom we could be less certain about whether or not they were depressed. Although this allowed us to test the psychometric characteristics of the translated version of the DMI, it interferes with its discriminatory ability and possibly affects the sensitivity and specificity analyses. A second possible limitation is that the interviewed sample was composed mainly by women many of who reported psychiatric drug use. As this is not a prevalence study and we were not interested in

obtaining clinical results, we consider that this limitation has not interfered with the aim of our study. Finally, CFA of the DMI-10 was made using the same sample as the CFA of the DMI-18.

Our study was conducted in Spain. All subjects participated in it were Spanish-speaking and a Spanish DMI translation was used. Future studies need to test how do these tools perform in Spanish speaking populations from diverse cultural backgrounds and countries.

In conclusion, our data shows that the psychometric properties of the Spanish-language versions of the DMI are comparable with the original instrument, and are thus reliable and valid questionnaires for the study of depression in patients with somatic disorders. In addition, the short length and its easy acceptance by patients make them attractive for this purpose. Their incorporation in a clinical contest routine may facilitate the diagnosis of depression, ensuring that patients are treated not only for their physical disease but also for their depression.

ACKNOWLEDGMENTS

The adaptation and validation of the DMI questionnaire was made possible by a grant from the Instituto de Salud Carlos III (file. PI06/90421), and there are no conflicts of interest. First of all, we thank Dr. Parker for his help with this article. We are very grateful to the participants who voluntarily took part in this study, to the mental health professionals of the Galdakao-Usansolo Hospital (Aranzazu Madrazo, Edurne Aristegui, Mónica Galván, Mª Natividad Silva, Mª Jose Horcajo, Belén Berjano, Patricia Fernández, Javier García, Estela Saez y Elida Belén Grande) for conducting the interviews and to the services of this hospital that participated in the study (pain unit, obstetrics and gynaecology, endocrinology, gastroenterology, neurology, pneumology, and psychiatry units) for helping with the patient screening.

409 **REFERENCES**
410

411 Aaronson, N. K., Acquadro, C., Alonso, J., Apolone, G., Bucquet, D., &
412 Bullinger, M. (1992). Internacional quality of life assessmente (IQOLA) project.
413 *Quality of Life Research, 1,* 349-351.

414 Adelman, R. D., Greene, M. G., Friedmann, E., & Cook, M. A. (2008).
415 Discussion of depression in follow-up medical visits with older patients. *The American*
416 *Geriatrics Society, 56,* 16-22.

417 American Psychiatry Association (1994). *Diagnostic and Statistical Manual of*
418 *Mental Disorders*. (4ª ed ed.) Washington D.C.: American Psychiatric Association.

419 Baca, E., Saiz, J., Aguera, L., Caballlero, L., Fernández-Liria, A., Ramos, J. et
420 al. (1999). Validación de la versión española del PRIME-MD: un procedimiento para el
421 diagnóstico de trastornos mentales en atención primaria. *Actas Españolas de*
422 *Psiquiatría, 27,* 375-383.

423 Bambauer, K. Z., Locke, S. E., Aupont, O., Mullan, M. G., & McLaughlin, T. J.
424 (2006). Using the Hospital Anxiety and Depression Scale to screen for depression in
425 cardiac patients. *General Hospital Psychiatry, 27,* 275-284.

426 Batista-Foguet, M., Coenders, G., & Alonso, J. (2004). Análisis factorial
427 confirmatorio. Su utilidad en la validación de cuestionarios relacionados con la salud.
428 *Medicina Clínica, 122,* 21-27.

429 Beck, A. T., Steer, R. A., Ball, R., Ciervo, C. A., & Kabat, M. (1997). Use of the
430 Beck anxiety and depression inventories for primary care with medical outpatients.
431 *Assessment, 4,* 211-219.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck Depression Inventory-II Manual. 2nd. ed. San Antonio, The Psychological Corporation.
Ref Type: Catalog

Beck, A. T., Steer, R. A., & Brown, G. K. (2000). BDI-Fast Screen for Medical Patients Manual. San Antonio, The Psychological Corporation.
Ref Type: Catalog

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1,* 185-216.

Cole, D. A. (1987). Utility of confirmatory factor analysis in thest validation research. *Journal of Consulting and Clinical Psychology, 55,* 584-594.

Diez-Quevedo, C., Rangil, T., nchez-Planell, L., Kroenke, K., & Spitzer, R. L. (2001). Validation and utility of the patient health questionnaire in diagnosing mental disorders in 1003 general hospital spanis inpatients. *Psychosomatic Medicine, 63,* 679-686.

Diez-Quevedo, C., Rangil, T., Sánchez-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.

Fayers, P. M. & Machin, D. (2007a). Scores and measurements: validity, reliability, sensitivity. In *Quality of life. The assessment, analysis and interpretation of patient-reported outcomes* (pp. 77-108). England: Wiley.

453 Fayers, P. M. & Machin, D. (2007b). Missing data. In *Quality of Life. The*
454 *assessment, analysis and interpretation of patient-reported outcomes* (2ª ed., pp. 355-
455 384). England: Wiley.

456 Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). Factor
457 Analysis. In *Multivariate data analysis* (5ª ed., pp. 87-135). New Jersey: Prentice-Hall
458 International.

459 Harcourt Assessment, I. (2006). Spanish translation of the BDI-Fast Screen from
460 medical patients. United States of America.
461 Ref Type: Pamphlet

462 Hatcher, L. (1994). Developing measurement models with Confirmatory Factor
463 Analysis. In *A step by step approach to using the SAS system for factor analysis and*
464 *structural equation modelling* (pp. 249-342). North Carolina: SAS Institute.

465 Herman, C. (1997). Internacional experineces with the hospital anxiety and
466 depression scale-a review of validation data and clinical results. *Journal of*
467 *Psychosomatic Research, 42,* 17-41.

468 Herrero, M. J., Blanch, J., Peri, J. M., De Pablo, J., Pintor, L., & Bulbena, A.
469 (2003). A validation study of the hospital anxiety and depression scale (HADS) in
470 Spanish population. *General Hospital Psychiatry, 25,* 277-283.

471 Instituto Vasco de Estadística (2006). Poblational Study. www.eustat.es [On-
472 line].

Kilbourne, A. M., Daugherty, B., & Pincus, H. A. (2007). What do general medical guidelines say about depression care? Depression treatment recommendations in general medical practice guidelines. *Current Opinion in Psychiatry, 20,* 626-631.

Koening, H. G. (1997). Differences in psychosocial and health correlates of major and mino depression in medically ill older adults. *Journal of the American Geriatrics Society, 45,* 1487-1495.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine, 16,* 606-613.

McDowel, I. & Newell, C. (1996). The Theoretical and technical foundations of health measurement. In *Measuring Health. A guide to rating scales and questionnaires* (2ª ed., pp. 10-46). New York: Oxford University Press.

McHorney, C. A. & Tarlov, A. (1995). Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Quality of Life Research, 4,* 293-307.

Mulrow, C. D., Williams, J. W., Gerety, J. W., Ramírez, G., Montiel, O. M., & Kerber, C. (1995). Case-finding instruments for depression in primary care settings. *Annals of Internal Medicine, 122,* 913-921.

Murphy, J. M., Berwick, D. M., Weinstein, M. C., Borus, J. F., Budman, S. H., & Klerman, F. L. (1987). Performance of screening and diagnostic tests. Application of receiver operating characteristic Analysis. *Archives of General Psychiatry, 44,* 550-555.

Newport, D. J. & Nemeroll, C. B. (1998). Assessment and treatment of depression in the cancer patient. *Journal of Psychosomatic Research, 45,* 215-237.

Nunnally, J. & Bernstein, I. (1994). *Psychometric Theory*. (3ª ed.) New York: McGraw Hill.

Parker, G., Hilton, T., Bains, J., & Hadzi-Pavlovic, D. (2002). Cognitive-based measures screening for depression in the medically ill: the DMI-10 and the DMI-18. *Acta Psychiatrica Scandinavica, 105,* 419-426.

Parker, G., Hilton, T., Hadzi-Pavlovic, D., & Bains, J. (2001). Screening for derpession in the medically ill: the suggested utility of a cognitive-based approach. *Australian and New Zealand Journal of Psychiatry, 35,* 474-480.

Parker, G., McDonald, S., Heruc, G. A., ley, A., otchie, H., r et al. (2006). A validation study of two brief measures of depression in the cardiac population: the DMI-10 and DMI-18. *Psychosomatics, 47,* 129-135.

Penley, J. A., Wiebe, J. S., & Nwosu, A. (2003). Psychometric properties of the Spanish Beck Depression Inventory-II in a medical sample. *Psychological Assessment, 15,* 569-577.

Sanz, J., Garcia-Vera, M. P., Espinosa, R., Fortun, M., & Vazquez, C. (2005). Spanish adaptation of the Beck Depression Inventory-II (BDI-II): psychometric features inpatients with psychological disorders. *Clinica y Salud, 16,* 121-142.

Sharp, L. K. & Lipsky, M. S. (2002). Screening for depression across the lifespan: a review of measures for use in primary care settings. *American family physician, 66,* 1001-1008.

Spitzer, R. L., Kroenke, K., Williams, J. B., & the Patient Health Questionnaire Primary Care Study Group (1999). Validation and Utility of a Self-report Version of PRIME-MD: The PHQ Primary Care Study. *JAMA, 282,* 1737-1744.

Spitzer, R. L., Williams, J. B., Kroenke, K., Linzer, M., deGruy, F. V. I., Hahn, S. R. et al. (1994). Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. *The Journal of the American Medical Association (JAMA), 272,* 1749-1756.

Viera, A. J. & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Family Medicine, 37,* 360-363.

Wilhelm, K., Kotze, B., Waterhouse, M., Hadzi-Pavlovic, D., & Parker, G. (2004). Screening for Depression in the Medically Ill: a comparison of self-report Measures, Clinician Judgment, and DSM-IV Diagnoses. *Psychosomatics, 45,* 461-469.

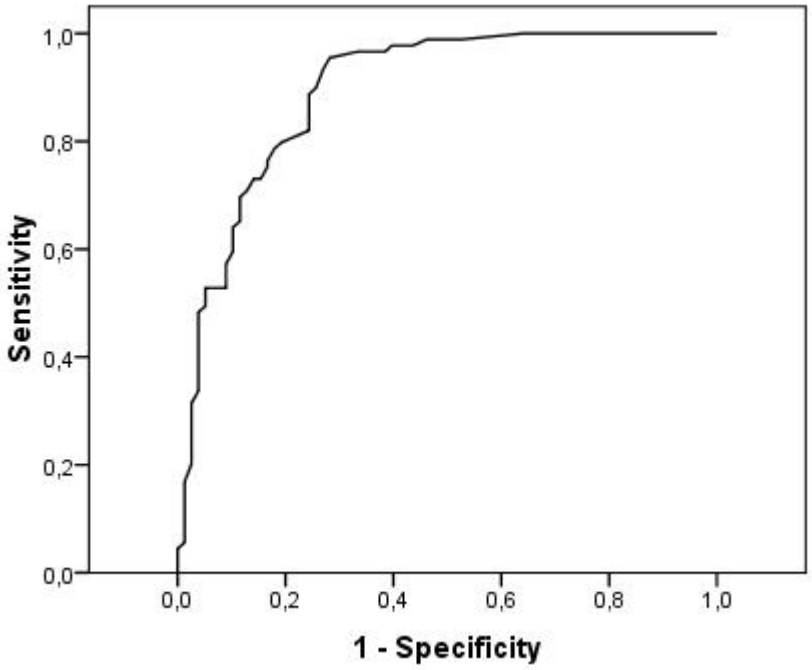
Zigmond, A. S. & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, 67,* 361-370.

FIGURE LEGENDS

Figure 1. Receiver Operating Characteristics (ROC) curve graph

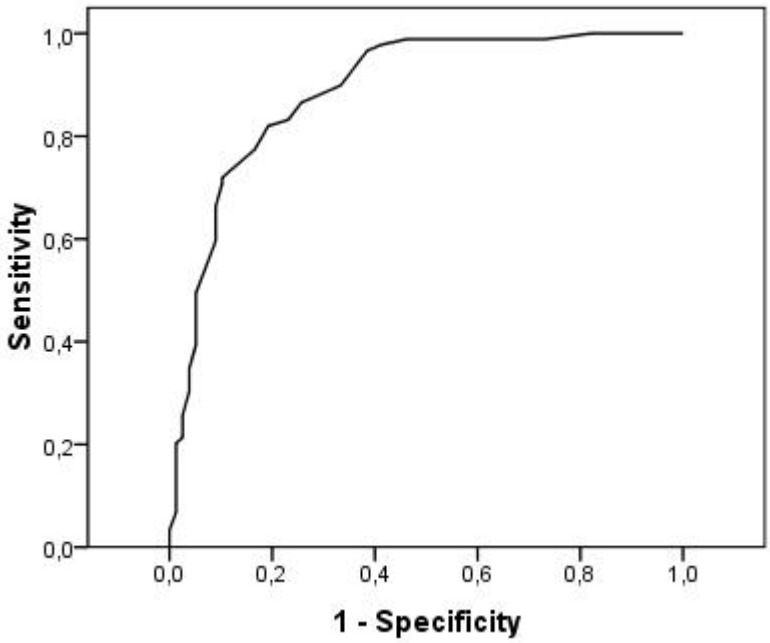
FIGURES

a) DMI-18



Area under the curve: 0.90 (95%CI: 0.85, 0.94)

b) DMI-10



Area under the curve: 0.89 (95%CI: 0.84, 0.94)

TABLES

Table 1. Sociodemographic data of the patient sample (n = 354).

Variables	*n (%)*
Gender	
Female	236 (66.7)
Age median (SD)	53.1 (15.3)
Marital Status	
Married or cohabitating	245 (69.2)
Divorced or separated	16 (4.5)
Single	57 (16.1)
Widow	35 (9.9)
Level of Education	
No schooling	29 (8.2)
Primary education	150 (42.4)
Secondary education	116 (32.8)
University studies	29 (8.2)
Posgraduate studies	8 (2.3)
Graduate/Professional	20 (5.6)
Employment Situation	
Employed	102 (28.8)
Homemaker	59 (16.7)
Unemployed	9 (2.5)
On strike	9 (2.5)
Retired	83 (23.4)
Disability payments	36 (10.2)
Study	6 (1.7)
On leave	38 (10.7)
Departments from which patients were recruited	
Pain unit	84 (23.7)
Psychiatry	17 (4.8)
OB/GYN[a]	16 (4.5)
Endocrinology	64 (18.1)
Gastroenterology	53 (15.0)
Neurology	39 (11.0)
Pneumology	29 (8.2)
Nephrology	29 (8.2)
Other Specialties	23 (6.5)
Time with disease	
Less than one year	36 (10.2)
More than one year	305 (86.2)
Current psychiatric drug use	
Yes	159 (44.9)
No	192 (54.2)
Do not know	2 (0.6)

The number of answers does not coincide with the number of participants because some subjects did not answer all of the questions. [a]OB/GYN: obstetrics and gynaecology

Table 2. Standarized estimates of unweighted least squares and fit indexes for the two-factor solution (Model 1) and the one-factor solution (Model 2) (N= 354)

	Model 1 DMI-18		*Model 2 DMI-18*		*Model 2 DMI-10*	
Items	*Factor Loading*	*Error Variance*	*Factor Loading*	*Error Variance*	*Factor Loading*	*Error Variance*
*1**	.65	.69	.61	.76	.59	.79
*2**	.73	.49	.68	.57	.67	.59
*3**	.69	.46	.65	.51	-	-
*4**	.78	.39	.73	.47	-	-
5	.74	.57	.74	.58	-	-
6	.79	.41	.78	.41	-	-
7	.78	.39	.78	.39	-	-
8	.74	.52	.69	.52	-	-
9	.70	.56	.70	.57	.72	.54
10	.76	.44	.75	.45	-	-
11	.61	.47	.61	.47	.62	.46
12	.76	.54	.76	.54	.72	.41
13	.82	.37	.82	.38	.75	.56
14	.70	.44	.70	.44	.82	.38
15	.72	.39	.72	.39	.72	.42
16	.72	.42	.72	.42	.75	.36
17	.80	.36	.80	.37	-	-
18	.73	.40	.72	.41	.70	.45
Fit indexes						
GFI	.994		.993		.992	
AGFI	.992		.991		.988	
RMR	.044		.047		.048	
SRMR	.045		.047		.049	

DMI-18: Depression in the Medically Ill questionnaire long version; DMI-10: Depression in the Medically Ill questionnaire abridge version; GFI: Goodness of Fit-index; AGFI: Adjusted for degrees of freedom; RMR: Root Mean Square Residual; SRMR: standarized root mean square residual.

* Items that saturate with Factor 1 in Model 1. Rest of the items saturate with factor 2. All items in the questionnaires showed a significant (p <0.001) saturation with the factors.

Table 3. Mean (SD) depression scores of DMI-18 and DMI-10 in HADS-D, PHQ-9 and BDI-PC categories according cut-off points established by the literature and Pearson's correlation coefficients (r) with 95% CI

	DMI-18 *Mean (SD)*	*DMI-10* *Mean (SD)*	*n*
HADS-D*			
0-7	11,7 (9,1)	6,2 (4,9)	48
8-10	26,0 (10,2)	14,3 (5,4)	21
≥11	30,8 (12.2)	16,7 (6,8)	28
Total			97
r (95% CI)	0.76 (0.65, 0.83)	0.74 (0.64, 0.82)	
PHQ-9*			
0-4	5,9 (5,7)	3,4 (3,3)	55
5-9	12,3 (6,4)	7,1 (4,0)	37
≥10	30,4 (11.7)	16,9 (6,6)	38
Total			130
r (95% CI)	0.88 (0.84, 0.92)	0.86 (0.81, 0.90)	
BDI-PC**			
0-3	6,7 (6,4)	3,7 (3,5)	62
≥4	24,2 (12.4)	13,4 (6,7)	63
Total			125
r (95% CI)	0.86 (0.80, 0.90)	0.85 (0.80, 0.89)	

DMI: Depression in the Medically Ill questionnaire; HADS-D: depression subscale of the Hospital Anxiety and Depression Scale; PHQ-9: the mood module of the Patient Health Questionnaire; BDI-PC: Beck Depression Inventory for Primary Care; r: Pearson' correlation coefficient; 95% CI: 95% Coefficient Intervals.

* ANOVA and J.T.

** T-test and Wilcoxon Rank-Sum test

In all cases the comparisons resulted in highly significant *p* values (i.e. $p<0.0001$)

Table 4. Validity and predictive precision of the DMI-18 and DMI-10 compared to PRIME-MD (N=167)

DMI-18						
Cut-off points	*TP/total +*	*TN/total -*	*Sensitivity (95%CI)*	*Specificity (95%CI)*	*PPV (95%CI)*	*NPV (95%CI)*
15	83/89	57/78	93 (88-98)	73 (63-83)	80 (72-88)	91 (83-98)
20	70/89	64/78	79 (70-87)	82 (74-91)	83 (75-91)	77 (68-86)
25	56/89	70/78	63 (53-73)	90 (83-96)	88 (79-96)	68 (59-77)
DMI-10						
Cut-off points	*TP/total +*	*TN/total -*	*Sensitivity (95%CI)*	*Specificity (95%CI)*	*PPV (95%CI)*	*NPV (95%CI)*
6	87/89	46/78	98 (95-100)	59 (48-70)	73 (65-81)	96 (90-100)
9	77/89	58/78	87 (79-94)	74 (65-84)	79 (71-87)	83 (74-92)
12	69/89	65/78	78 (69-86)	83 (75-92)	84 (76-92)	77 (67-85)

TP/total+: True positive/total positive; TN/total-: True negative/total negative; 95% CI: 95% confidence intervals; PPV: positive predictive value; NPV: negative predictive value.

679 **APPENDIX 1. SPANISH VERSION OF THE DEPRESSION IN THE**

680 **MEDICALLY ILL QUESTIONNAIRE**

681

Por favor responda, ¿Cómo se ha sentido en los 2 o 3 últimos días comparado con cómo se suele sentir normalmente?

	Marque con una X la opción más adecuada			
	Nunca	**Alguna vez**	**A menudo**	**Siempre**
1. ¿Le da demasiadas vueltas a las cosas?	□	□	□	□
2. ¿Se ha sentido más sensible emocionalmente (más vulnerable) que de costumbre?	□	□	□	□
3. ¿Se ha sentido más solo/a de lo habitual?	□	□	□	□
4. ¿Ha tenido más ganas de llorar que de costumbre?	□	□	□	□
5. ¿Ha sentido que ya no disfruta de las cosas que antes disfrutaba?	□	□	□	□
6. ¿Se ha sentido pesimista?	□	□	□	□
7. ¿Se ha sentido mal consigo mismo/a?	□	□	□	□
8. ¿Se ha sentido más inseguro/a que de costumbre?	□	□	□	□
9. ¿Está siendo duro/a y crítico/a consigo mismo/a?	□	□	□	□
10. ¿Se ha sentido desmoralizado/a (es decir, con el ánimo bajo)?	□	□	□	□
11. ¿Se siente culpable por algo ocurrido en su vida?	□	□	□	□
12. ¿Ha sentido como si ya no fuese el/la mismo/a (ha cambiado su forma de ser habitual)?	□	□	□	□
13. ¿Se ha sentido deprimido/a?	□	□	□	□
14. ¿Ha sentido que usted no vale la pena?	□	□	□	□
15. ¿Se siente desesperanzado/a o falto/a de ayuda?	□	□	□	□
16. ¿Se ha sentido más distante del resto de la gente?	□	□	□	□
17. ¿Ha perdido el interés en sus actividades habituales?	□	□	□	□
18. ¿Ha sentido que nada le puede levantar el ánimo?	□	□	□	□