This is the accepted manuscript of the article that appeared in final form in **Journal of Psychosomatic Research** 69(4) : 399-406 (2010), which has been published in final form at https://doi.org/10.1016/j.jpsychores.2010.04.007.

Detecting depression in medically ill patients: comparative accuracy of four screening questionnaires and physicians' diagnoses in Spanish population

Miren Orive, MPH[a*], Jesús A. Padierna, MD[b], Jose M. Quintana, MD, PhD[a], Carlota Las Hayas PhD[a], Kalliopi Vrotsou, MSc[a], Urko Aguirre, MSc[a]

[a]*Research Unit-CIBER Epidemiología y Salud Pública (CIBERESP). Hospital Galdakao-Usansolo.*

[b]*Department of Psychiatry, Hospital Galdakao-Usansolo, Bº Labeaga s/n, Bizkaia, Spain, 48960.*

*** Corresponding author:** Miren Orive

Research Unit-CIBER Epidemiología y Salud Pública (CIBERESP), Hospital Galdakao-Usansolo, Bº Labeaga s/n, Bizkaia, Spain, 48960.

Tel: +34 94 4007105; fax: +34 94 4007132

E-mail address: miren.orivecalzada@osakidetza.net

Running head: Depression screening in medically ill patients.

25

26

27

28

Abstract

Objective: To compare the diagnostic accuracy of four depression screening tools commonly used in patients with medical disorders, relative to a reference diagnostic standard – a structured psychiatric interview.

Methods: The Depression in the Medically Ill-18 (DMI-18) questionnaire was administered to 167 patients with medical disorders, of those 53 completed the Beck Depression Inventory for Primary Care (BDI-PC), 67 the Hospital Anxiety and Depression Scale (HADS) and 46 the Patient Health questionnaire-9 (PHQ-9). The entire sample was also interviewed with a structured psychiatric interview conducted by a mental health professional. Sensitivity, specificity, likelihood ratios (LR) and area under the curve (AUC) were calculated and compared for the different measures.

Results: At their respective recommended cut-off points, sensitivities (95% CI) were 86% (70-95), 82% (63-94), 93% (86-97) and 68% (47-85) for the HADS-D, BDI-PC, DMI-18, and PHQ-9 respectively, while specificities ranged from 72% (47-90) for BDI-PC to 89% (72-98) for PHQ-9. The sensitivities of DMI-18 were significantly higher compared to those of HADS-D (p= 0.045) and PHQ-9 (p= 0.01). The PHQ-9 questionnaire obtained the most favourable positive LR (6.35; 95% CI: 2.48-18.36). In contrast, the DMI-18 showed the best negative LR (0.09; 95% CI: 0.04-0.18). AUCs (95% CI) ranged from 0.92 (0.83-1.02) to 0.84 (0.74-0.94). Staistically significant differences were found between the AUCs of the DMI-10 and the BDI-PC.

Conclusion: Our results suggest that all evaluated scales have acceptable abilities and can be used as screening instruments for depression in patients with medical disorders. The DMI stands out for its sensitivity.

. **Keywords:** depression, screening, medical disorder, psychometrics

Introduction

Depression is the second most common chronic disorder seen in general practice and in primary care [1]. Approximately 12 percent of patients seen in primary care settings have major depression [1;2], a rate exceeding that in the general population (5%-10%) [3]. Depression is a major cause of psychological and physical comorbidity, and is associated with greater suffering and disability compared to other chronic medical conditions [4;5].

Since many patients with depression can be effectively treated with medication and psychotherapy, early diagnosis and treatment can significantly reduce the impact of the depression [6]. However, symptoms of depression are not recognised in up to half of patients with depressive disorders in general practice, in primary care and in general hospital settings [3;7;8].

Providing primary care practitioners and other generalists with short, reliable questionnaires can help them identify and manage patients with depression. The use of such screening instruments for improving the quality of care for depression has been supported by different institutions [3;9].

Selecting the appropriate screening instrument is an important first step. The characteristics of the target population, the psychometric properties of the questionnaire (i.e. validity, sensitivity and specificity), the time required to complete it and its comprehensiveness are some of the issues that must be considered [1;10].

The purpose of this report was to determine the comparative validity of the depression subscale of the Hospital Anxiety and Depression Scale (HADS-D), the Beck Depression Inventory for Primary Care (BDI-PC), the Patient Health Questionnaire-9 (PHQ-9), the Depression in the Medically Ill-18 (DMI-18) and the abridged version of the DMI-18 (DMI-10) in diagnosing depression, using as gold standard a structured

interview, performed by a mental health professional, the Primary Care Evaluation of Mental Disorders (PRIME-MD).

Methods

Subjects

This is a cross-sectional study carried out at the Galdakao-Usansolo Hospital (Bizkaia, Spain) between November 2007 and April 2008. Galdakao-Usansolo is a 400-bed teaching hospital, which serves a population of 300,000 inhabitants. It belongs to the network of public hospitals of the Basque Health Care Service, which provides free unrestricted care to nearly 100% of the population.

In order to have heterogeneity of medical disorders, the target population included patients with a medical disorder recruited from the waiting rooms of various services at the Galdakao-Usansolo Hospital: pain unit, obstetrics and gynaecology, endocrinology, gastroenterology, neurology, pneumology, nephrology, otorhinolaryngology and psychiatry units. To be eligible to participate, patients had to be adult (over age 18 years) and attending one of the collaborating services for a medical disorder. Patients were excluded if they, at the physician's discretion, had a severe physical disease, cognitive deterioration, any neurological disease, a psychotic disorder that might compromised their ability to complete the questionnaires, or if they declined to participate after informed consent. Also we excluded those patients who did not answer more than 50% of the assigned questionnaires.

The study protocol was approved by the ethics committee of the Galdakao-Usansolo Hospital.

Procedure

Two of the authors (M.O. and C.L.H.) approached patients about participating in the study. They emphasized to patients that participation was voluntary and explained that the objective of the study was to evaluate “the emotional reactions associated with the

fact of suffering from a disease." Patients were also told about the study by their physicians. Informed consent had to be provided before patients took part in the study.

Each participant was asked to complete a set of sociodemographic questions, the DMI-18 questionnaire, and one of the other three depression screening questionnaires: HADS, BDI-PC or PHQ-9. Patients were not asked to complete both forms of the DMI to avoid redundancy. Responses to the abridged version were collected from the responses of each patient to the long DMI version. Regarding HADS, the participants completed all 14 items, but for the analyses only the depression subscale (HADS-D) items were taken into account.

All the patients were interviewed by a group of mental health professionals the same day they completed the questionnaires. The corresponding mental health professional (either a psychiatrist or a clinical psychologist with broad experience in interdisciplinary consultation) was blinded to the results of the questionnaires.

To assure that the mental health collaborators evaluated patients in a consistent way, all of them undertook an inter-rater study. They had to obtain a Kappa (κ) score of at least 0.60 when comparing their assessments (presence or absence of depression) with those of a gold standard. The gold standard consisted of a list of diagnoses of "case or no case" of depression performed by a psychiatric expert in diagnosing depressive disorders who offered 10 of his own patients to be re-evaluated by the mental health professionals [11]. A total of 10 mental health collaborators obtained the kappa level requirement. Six of them obtained a κ value of 0.67, for two κ was equal to 0.83 and for the rest two κ was equal to 1.

Materials

Screening questionnaires

Four short easily administered depression scales were applied in their Spanish versions:

The Depression in the Medically Ill (DMI) questionnaire was specifically designed to detect depression in patients with medical disorders. There are two versions: a complete version with 18 questions (DMI-18) and an abridged version with 10 questions (DMI-10). For each question there is a 4 point ordinal response, with options ranging from "none" (scored as 0) to "always" (scored as 3). The reliability and construct validity of the Spanish-language version is satisfactory. Scores above a cut-off point of 15 for the DMI-18 and above a cut-off point of 9 for the DMI-10 were considered to indicate depression.

The Beck Depression Inventory for Primary Care (BDI-PC) [12] consists of 7 cognitive and affective items extracted from the 21-item Beck Depression Inventory-II (BDI-II) [13]. 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. The manual recommends a cut-off point of 4 to identify depression [14].

The Hospital Anxiety and Depression Scale (HADS-D) was specially designed for identifying and quantifying depression and anxiety in physically ill patients [15;16]. 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. For the present study we only used the HADS-D subscale. Originally, a cut-off point of 8 indicated a possible case of anxiety or depression [17].

The Patient Health Questionnaire-9 (PHQ-9) is the mood module of the Patient Health Questionnaire, a self-administered version of the PRIME-MD [18;19]. 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 [5;20], 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. Values equally or greater than 10 were considered indicative of depression [21].

Psychiatric interview

Collaborating mental health professionals used the mood module of the Primary Care Evaluation of Mental Disorders (PRIME-MD) structured psychiatric interview in Spanish [7] to help themselves in screening for depression. Thus, mental health professionals' expertise along with their scores in the PRIME-MD structured interview was used as the gold standard for the presence of depressive disorder. The PRIME-MD has nine items that represent the nine DSM-IV depression criteria with dichotomous response categories (yes/no).

Statistical analysis

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% [22].

Missing values were treated using the mean imputation method [23]. 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 whenever more than 50% of the items have been sufficiently answered.

Associations between categorical variables were examined with the chi-square test. Significance of score differences was tested with the Wilcoxon Rank-Sum or the Kruskall-Wallis test. The internal consistency of the different questionnaires was examined with Cronbach's alpha. Convergent validity of the scales was tested with Pearson's correlation coefficient (r) and the intraclass correlation coefficient (ICC) for the degree of agreement between the different measurements.

Sensitivity and specificity along with their exact binomial 95% Confidence Interval (CI) [24] were calculated to assess the ability of the screening instruments to render a clinically validated diagnosis of depression. The McNemar test was used for comparing these quantities between DMI-18 and the other 3 screening tools, as well as with DMI-10 [25]. Standard cut-off points for each instrument were used following the corresponding literature. Furthermore, positive and negative likelihood ratios (LR) of the tests with 95% CI were calculated [24;26;27].

Receiver operating characteristic (ROC) [28] curves were designed and the areas under those curves (AUC) were calculated. Finally, pairwise comparisons of the obtained AUCs were performed [28]. Statistical analyses were performed with SAS for Windows, version 9.1.

Results

Sample description

Of the 167 patients who agreed to participate in the study, a patient did not answer 50% of the HADS-D items, thus leaving a cohort of 166 patients for some analyses. Given that completing all battery of tests would be tiring for the patients, we originally aimed for a 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. Of the total sample, 67 patients had completed the HADS, 46 patients the BDI-PC, and 53 the PHQ-9.

Baseline characteristics of the patients are shown in Table 1. No statistically significant differences were found among the 3 subgroups, except for the variables of gender and department. Medians scores for all categories of the baseline characteristics in the 3 subgroups are presented in table 2.

----------Table 1----------

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

Internal consistency and intercorrelations

The internal consistency of all screening questionnaires turned out to be good with Cronbach alpha values exceeding 0.80 in all cases: 0.83 for the HADS-D; 0.86 for the BDI-PC; 0.90 for the PHQ-9; 0.96 for the DMI-18; and 0.92 for the DMI-10. Table 3 shows the values for Pearson´s r and ICC values. The total DMI-18 score and total DMI-10 score correlated strongly with BDI-PC, HADS-D and PHQ-9 scores. The ICC ranged from 0.65 (95% CI: 0.45-0.79) (BDI-PC vs. DMI-18) to 0.87 (95% CI: 0.79-0.92) (PHQ-9 vs. DMI-10).

225 ----------Table 3----------

226

227 *Operating characteristics of the screening questionnaires*

228 Table 4 reports the sensitivity, specificity, and positive and negative likelihood ratios
229 for the cut-off points recommended in the literature for each questionnaire. The ability
230 of the PHQ-9 in assessing a clinically validated diagnosis of depression was slightly
231 low: 68% (95% CI: 47-85) of the patients with depression were correctly identified by
232 this test. Statistically significant differences were found when comparing the
233 sensitivities: DMI-18 and HADS-D (p= 0.045); DMI-18 and PHQ-9 (p= 0.01); and also
234 DMI-18 and DMI-10 (p= 0.01) within the same samples. In all cases the sensitivity of
235 DMI-18 was higher. The PHQ-9 questionnaire obtained the most favourable positive
236 LR (6.35; 95% CI: 2.48-18.36). In contrast, the DMI-18 showed the best negative LR
237 (0.09; 95% CI: 0.04-0.18).

238 ----------Table 4----------

239 Receiver Operating Characteristic curves (ROC) for DMI-18 along with the
240 other 3 questionnaires and DMI-10 are presented in Figure 1. In all cases the estimated
241 AUC were quite similar. A statistically significant difference was found between the
242 AUCs of the BDI-PC and the DMI-18 (p=0.02).

243 ----------Figure 1----------

244

Discussion

The goal of this article was to determine the operating characteristics of four self-report depression screening instruments relative to a reference diagnostic standard in patients with medical disorders.

The results demonstrated excellent internal consistencies for all the instruments. The substantial positive correlations between both DMI questionnaires and the three older instruments showed the extent to which the scales measure the same construct. Patients with high depression scores on the DMI-18 or DMI-10 also had high depression scores on the HADS-D, BDI-PC and PHQ-9.

In a review of 9 widely used instruments for the detection of depression in primary care settings, Mulrow, Williams, Gerety, Ramírez, Montiel and Kerber [22] found minimum sensitivity and specificity values of 84% and 72% respectively. If we considered these to represent the minimally acceptable levels of sensitivity and specificity, the HADS-D, BDI-PC, DMI-18 and DMI-10 reached the minimum in our study. The sensitivity of the PHQ-9 (68%) did not reach this minimum, but its specificity did (89%). The DMI-18 appeared to be statistically significantly more sensitive than the HADS-D, PHQ-9 and DMI-10. On the other hand, the ROC analyses suggested that apart from the BDI-PC, all other questionnaires had the same overall screening accuracy with that of DMI-18. The importance of these differences should also be tested in a clinical setting. Finally, all positive LR were greater than 1, and all negative LR were less than 1, indicating that the positive test result is associated with presence of the disease and a negative test results with its absence [26]. The PHQ-9 had the highest positive LR (6.35; 95% CI: 2.48-18.36). The DMI-18 (whole sample) had the lowest negative LR (0.09; 95% CI: 0.04-0.18), meaning that for a negative result (i.e. no depression) the probability of depression is very low.

The HADS is the most commonly used screening tool for depression in patients with medical disorders. Its good operating characteristics have been demonstrated in several validation studies [15;29] and our results are in line with most of them. The concept of anhedonia is predominant in the scale and 5 of the 7 depression subscale items are related with this feature. Its authors [16] considered anhedonia the "central pathological feature of that form of depression that corresponds well to antidepressant drug treatment". Parker, Hilton, Bains and Hadzi-Pavlovic [30] do not agree with them. The latter suggest that "the problematic nature of anhedonia in medically ill patients is that it is strongly related with the somatic symptomatology which may hinder the detection of depression in such patients".

The BDI-PC was found to have acceptable psychometric characteristics. Even though this tool is consisted of only 7 items, it takes quite a long time to be administered. This may be attributed to the fact that the alternative responses change from question to question, posing a cognitive processing burden on the respondents. Our experience is consistent with that of Shumway, Sentell, Unick and Bamberg [31] who said that the BDI is among the more cognitively complex measures evaluated in their study, and with those of Sentell and Ratcliff-Baird [32], who explored the difficulties involved in comprehending the BDI. In our study some of the participants commented that they found the content of the items a "bit aggressive" or "too direct", mainly referring to the last response options, as these are ranked from less to more severe alternatives. Many of these participants were not familiar with the symptoms of depression and were thus surprised with the content of those response options.

In our study, the PHQ-9 had higher specificity rather than sensitivity. This could suggest that the specific questionnaire might be more appropriated when higher

294 specificity levels are preferred. This questionnaire includes somatic symptoms like
295 fatigue, insomnia and anorexia.

296 The DMI is a relatively new instrument for screening depression in patients with
297 medical disorders. Its diagnostic validity is comparable to the other 3 older and
298 commonly used instruments examined in this article. Its high negative LR and its good
299 sensitivity value are very positive results, worth highlighting. This questionnaire is
300 based on affective symptoms that are purely cognitive, including all areas central to
301 depression. It is brief, user-friendly and easy to grade.

302 The statistically significant differences in gender and department seen between
303 the samples may be a limitation of this study. However, we found homogeneous scores
304 in the groups in terms of gender. Secondly, the screening questionnaires were
305 administered verbally by the researchers. Thus, their characteristics and subsequently
306 their ability to discriminate between depressed and non-depressed patients may differ
307 from those administered as self-reported questionnaires. To compensate for this
308 limitation, the researchers who administered the questionnaires followed a systematic
309 procedure with all patients, were experts in the field and familiar with these kinds of
310 tools. We preferred collecting the data in this way in order to reduce missing data.

311 With respect to our results, we may conclude that, for epidemiological purposes,
312 all tools can be equally recommended as valid and practicable screening instruments for
313 depression in patients with medical disorders. In contrast, for screening purposes, where
314 a high sensitivity is more desirable than a high specificity [33], we encourage using the
315 DMI-18 since it presents the highest levels in this attribute. Nevertheless, we also hold
316 in mind that more studies with larger sample are needed for confirming these results.

317

Acknowledgments

This study was supported by a grant from the Salud Carlos III Institute (PI06/90421). There are no conflicts of interests. We thank Dr. Parker for permitting us to adapt and validate the DMI-18 questionnaire into Spanish. We are also grateful for his availability and valuable suggestions. We would also like to express our thanks 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, nephrology, otorhinolaryngology, and psychiatry units) for helping with the patient screening.

Reference List

[1] Sharp LK, Lipsky MS. Screening for depression across the lifespan: a review of measures for use in primary care settings. Am Fam Physician 2002; 66: 1001-1008.

[2] Kilbourne AM, Daugherty B, Pincus HA. What do general medical guidelines say about depression care? Depression treatment recommendations in general medical practice guidelines. Curr Opin Psychiatr 2007; 20: 626-631.

[3] Gilbody S, House AO, Sheldon TA. Instrumentos de cribaje (screening) y detección de casos para la depresión (Revisión Cochrane traducida). In: La Biblioteca Cochrane Plus, 2007 Número 1. Oxford: Update Software Ltd. Disponible en: http://www.update-software.com.

[4] Watnick S, Pei-Li W, Demadura T, Ganzini L. Validation of depression screening tools in dialysis patients. Am J Kidney Dis 2005; 46: 919-924.

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

[6] Eack SM, Singer JB, Greeno CG. Screening for Anxiety and Depression in Community Mental Health: The Beck Anxiety and Depression Inventories. Community Ment Health J 2008 (in press).

[7] Baca E, Saiz J, Aguera L, Caballlero L, Fernández-Liria A, Ramos J, Gil A, Madrigal M, Porras A. Validación de la versión española del PRIME-MD: un procedimiento para el diagnóstico de trastornos mentales en atención primaria. Actas Esp Psiquiatr 1999; 27: 375-383.

[8] Pignone M, Gaynes BN, Rushton JL, Burchell CM, Orleans CT, Mulrow CD. Screening for depression in adults: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2002; 136: 765-776.

[9] United states Preventive Services Task Force. Screening for depression: recommendations and rationale. Ann Intern Med 2002; 136: 760-764.

[10] Parker G, Gladstone G. Capacity of the 10-item Depression in the Medically Ill screening measure to detect depression "caseness" in psychiatric out-patients. Psychiatry Res 2004; 127: 283-287.

[11] Hair JF, Anderson RE, Tatham RL, Black WC. Factor Analysis. In Multivariate data analysis (5ª ed). New Jersey: Prentice-Hall International 1998: 87-135.

[12] Harcourt Assessment I. Spanish translation of the BDI-Fast Screen from medical patients. United States of America 2006.

[13] Beck AT, Steer RA, Brown GK. Beck Depression Inventory-II Manual (2nd ed), San Antonio: The Psychological Corporation 1996.

[14] Beck AT, Steer RA, Brown GK. BDI-Fast Screen for Medical Patients Manual. San Antonio: The Psychological Corporation 2000.

[15] Herrero MJ, Blanch J, Peri JM, De Pablo J, Pintor L, Bulbena A. A validation study of the hospital anxiety and depression scale (HADS) in Spanish population. Gen Hosp Psychiatry 2003; 25: 277-283.

[16] Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatri Scand 1983; 67: 361-370.

[17] Herman C. International experiences with the hospital anxiety and depression scale-a review of validation data and clinical results. J Psychosomc Res 997; 42: 17-41.

[18] Diez-Quevedo C, Rangil T, Sánchez-Planell L, Kroenke K, Spitzer RL. Validation and utility of the patient health questionnaire in diagnosing mental disorders in 1003 general hospital Spanish inpatients. Psychosom Med 2001; 63: 679-686.

[19] Spitzer RL, Williams JB, Kroenke K, Linzer M, deGruy FVI, Hahn SR, Brody D, Johnson JG. Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. JAMA 1994; 272: 1749-1756.

[20] American Psychiatry Association. Diagnostic and Statistical Manual of Mental Disorders, (4th ed). Washington D.C.: American Psychiatric Association; 1994.

[21] Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001; 16: 606-613.

[22] Mulrow CD, Williams JW, Gerety JW, Ramirez G, Montiel OM, Kerber, C. Case-finding instruments for depression in primary care settings. Ann Intern Med 1995; 122: 913-921.

[23] Fayers PM, Machin D. Missing data. In: Quality of Life. The assessment, analysis and interpretation of patient-reported outcomes (2nd ed). England: Wiley 2007: 355-384.

[24] Bergstralh E. Binomial confidence interval. In. In: http://mayoresearch.mayo.edu/mayo/research/biostat/upload/bnmlci.sas 2004.

[25] Durkalski VL, Palesch YY, Lipsitz SR, Rust PF. Analysis of clustered matched-pair data. Stat Med 2003; 22: 2417-2428.

[26] Deeks JJ, Altman DG. Diagnostic tests 4: likelyhood ratios. BMJ 2004; 329: 168-169.

[27] Grimes AG, Schulz KF. Refining clinical diagnosis with likelihood ratios. Lancet 2005; 365: 1500-1505.

[28] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143: 29-36.

[29] Herrmann C. Internacional experineces with the hospital anxiety and depression scale-a review of validation data and clinical results. J Psychosom Res 1997; 42: 17-41.

[30] Parker G, Hilton T, Bains J, Hadzi-Pavlovic D. Cognitive-based measures screening for depression in the medically ill: the DMI-10 and the DMI-18. Acta Psychiatr Scand 2002; 105: 419-426.

[31] Shumway M, Sentell T, Unick G, Bamberg W. Cognitive complexety of self-administered depression measures. J Affect Disord 2004; 83: 191-198.

[32] Sentell T, Ratcliff-Baird B. Literacy and comprehension of Beck Depression Inventory response alternatives. Community Ment Health J 2003; 39: 323-331.

[33] Henkel V, Mergl R, Kohnen R, Maier W, Möler HJ, Hegerl U. Identifying depression in primary care: a comparison of different methods in a prospective cohor study. BMJ 2003; 326: 200-201.