Factorial and Diagnostic Validity of the Beck Depression Inventory-II (BDI-II) in Croatian Primary Health Care

**Nenad Jakšić · Ena Ivezić · Nataša Jokić-Begić ·
Zsuzsanna Surányi · Stanislava Stojanović-Špehar**

Abstract The aim of this study was to examine the factorial and diagnostic validity of the Beck Depression Inventory-Second Edition (BDI-II) in Croatian primary health care. Data were collected using a medical outpatient sample (N = 314). Reliability measured by internal consistency proved to be high. While the Velicer MAP Test showed that extraction of only one factor is satisfactory, confirmatory factor analysis indicated the best fit for a 3-factor structure model consisting of cognitive, affective and somatic dimensions. Receiver operating characteristics (ROC) analysis demonstrated the BDI-II to have a satisfactory diagnostic validity in differentiating between healthy and depressed individuals in this setting. The area under the curve (AUC), sensitivity and specificity were high with an optimal cut-off score of 15/16. The implications of these findings are discussed regarding the use of the BDI-II as a screening instrument in primary health care settings.

Introduction

Primary health care (PHC) plays a key role in the detection and management of depression and represents the entry point to psychiatric care for people with depression (WHO, 2001). While the reported prevalence of depression is dependent on diagnostic methods used, it usually ranges between 6 and 15 % (Bauer et al., 2007; King et al., 2008), with some studies reporting even higher occurrence rates (e.g., Al-Windi, 2005) and a large variability across different countries (Tarricone et al., 2012). Also, the majority of depressed patients seek help from and are treated by general practitioners (GPs) (Gaynes et al., 2008; Wancata & Friedrich, 2011), while the severity of symptoms in these individuals is similar to those of psychiatric patients (Gaynes et al., 2005). Unfortunately, it is estimated that over 50 % of depressed patients often remain undiagnosed by GPs, and are therefore left untreated (Cepoiu et al., 2008; Jackson, Passamonti, & Kroenke, 2007). These particular individuals have increased levels of mortality and morbidity (Katon, 2003; Licht-Strunk, Beekman, de Haan, & van Marwijk, 2009) and seek medical help more often (Weissman et al., 2010). Arguably, the early detection and treatment of depression is a worthy aim in order to reduce these negative consequences and to promote remission and prevent relapse. Indeed, in the last decade, the use of screening instruments for the early detection of depression has increased (O'Connor, Whitlock, Beil, & Gaynes, 2009).

N. Jakšić
Department of Psychiatry, University Hospital Center Zagreb, Zagreb, Croatia

E. Ivezić
Psychiatric Hospital "Sveti Ivan", Zagreb, Croatia

N. Jokić-Begić (✉)
Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia, I. Lucica 3, 10 000 Zagreb, Croatia
e-mail: njbegic@ffzg.hr

Z. Surányi
Department of General Psychology, Institute of Psychology, Károli Gáspár University of the Reformed Church, Budapest, Hungary

S. Stojanović-Špehar
Department of Family Medicine, Andrija Štampar School of Public Health, University of Zagreb, Zagreb, Croatia

In Croatian PHC, the use of depression screening instruments is not common practice, and for this reason there is currently no accurate and valid data on the prevalence of depression in primary care (Stojanovic-Spehar et al., 2009). In contrast, there is frequent use of screening instruments for measuring depression in patients suffering from different somatic illnesses, such as diabetic patients (Pibernik-Okanovic, Peros, Szabo, Begic, & Metelko, 2005), patients infected with HIV (Kolaric, Tesic, Ivankovic, & Begovac, 2006), epilepsy patients (Hecimovic, Bosnjak, & Demarin, 2008), and various other somatic disorders (Filipcic et al., 2007).

Croatia is an Eastern European country currently in a post-transitional period in which society has evolved from a traditional socialist to a liberal capitalist model. Although Croatia is currently at the end of this transition, the conclusion of these important multilevel changes was postponed by war and the later impact of this war on Croatian society. These turbulent conditions over the past two decades became a source of stress for all Croatian citizens. Recent research in a community sample of citizens directly exposed to war in Croatia and other ex-Yugoslavian countries (Priebe et al., 2010) documented higher prevalence rates of depressive disorders (25.9 %) than in Western countries (Richards, 2011), a finding that clearly demands both research and clinical attention. Arguably, it is important to first establish a psychometrically valid and "user-friendly" screening instrument for the early detection of depression that can be used in both PHC practice and in research.

One of the most widely used instruments for depression is the Beck Depression Inventory-Second Edition (BDI-II; Beck, Steer, & Brown, 1996), a self-rating scale for the assessment of the severity of depression in adults and adolescents older than 13 years. Validation studies have shown high test–retest reliability and internal consistency (e.g., Arnau, Meagher, Norris, & Bramson, 2001; Steer, Rissmiller, & Beck, 2000) and moderate to high convergent and divergent validity (e.g., Beck et al., 1996; Kapci, Uslu, Turkcapar, & Karaoglan, 2008). The BDI-II has been translated into many languages and has shown good qualities as a screening method for identifying the possible presence and severity of depressive symptoms (e.g., Campos & Goncalves, 2011; Segal, Coolidge, Cahill, & O'Riley, 2008). The BDI-II is also one of the most commonly used screening measures for adults in PHC settings (Sharp & Lipsky, 2002) and has been recommended specifically for this purpose, together with the Patient Health Questionnaire 9 (PHQ-9) and the Hospital Anxiety and Depression Scale (HADS), by the UK Quality and Outcomes Framework (QOF) (NHS Employers and the General Practitioners' Committee, 2009). All three recommended screening measures have shown generally strong and comparable psychometric properties in terms of their internal consistency, factor structure and convergent validity (Applied Health Sciences, 2011).

The factor structure of the BDI-II is inconsistent across studies and still remains somewhat controversial. Some studies using non-clinical samples have often found a structure comprising two factors, usually called *Cognitive-affective* and *Somatic* (Beck et al., 1996; Dozois, Dobson, & Ahnberg, 1998; Whisman, Perez, & Ramel, 2000; Wiebe & Penley, 2005), while other studies have failed to confirm this particular structure (e.g., Kojima et al., 2002; Uslu, Kapci, Oncu, Ugurlu, & Turkcapar, 2008). In clinical samples, two factors have also been frequently obtained but differed from those found in nonclinical samples, typically labelled as *Somatic-affective* and *Cognitive* (Beck et al., 1996; Bedi, Koopman, Thompson, 2001; Steer et al., 2000). In two studies, a three-factor model was obtained that distinguished between cognitive, somatic and affective factors (Beck, Steer, Brown, & Van der Does, 2002; Buckley, Parker, & Heggie, 2001). However, there were some differences between these studies in the partition of items into factors. To date, there has been little research examining the factor structure of the BDI-II with general PHC samples. However, existing studies have shown the two-factor structure obtained in the original validation of the clinical sample, thus confirming the existence of somatic-affective and cognitive factors (e.g., Arnau et al., 2001; Viljoen, Iverson, Griffiths, & Woodward, 2003).

In clinical practice, the diagnostic validity of cut-off scores empirically derived from psychometric instruments is of critical importance. These scores are assessed using Receiver Operating Characteristics (ROC) Analysis, which includes the value of the area under the ROC curve (AUC), sensitivity, specificity, positive (PPP) and negative predictive power (NPP), as well as the optimal cut-off score in the differentiation of healthy and depressed individuals. For the BDI-II, Beck et al. (1996) suggested the following cut-off scores for patients suffering from depression: minimal (0–13), mild (14–19), moderate (20–28), and severe depression (29–63). While some studies have obtained very similar values for these levels of depression (e.g., Kapci et al., 2008), most studies have found that cut-off scores between 14 and 20 best discriminated healthy individuals from depressed ones (e.g., Arnarson, Olason, Smar, & Sigurethsson, 2008; Bunevicius, Staniute, Brozaitiene, & Bunevicius, 2012; Huffman et al., 2010). These studies were carried out using various samples, with the particular values depending on the specific characteristics of the examined population. To date, only a handful of studies investigating the diagnostic validity parameters of the BDI-II have been conducted with general PHC samples (e.g., Arnau et al., 2001; Dutton et al., 2004). Furthermore, these cut-off scores are not culturally independent. The

results of one pan-European study using the BDI (an earlier, and very similar, version of the BDI-II) suggest that the experience of depression may differ cross-culturally (Nuevo et al., 2009). The authors of this study concluded that, while the BDI can be used across cultures in European settings, one must take into account that the probability of responding low or high to several items might be biased by cultural or language issues.

A comparison of the BDI-II structure across different cultures has yet to receive sufficient attention (Nuevo et al., 2009). In order to ensure that a scale is consistent across different versions or with different groups, and that any interpretation based on differences in scores is valid, it is necessary to first establish evidence of measurement invariance. In light of this prerequisite, the aim of this study was to examine the factorial and diagnostic validity of BDI-II in a general PHC sample in Croatia. In doing so, the findings of the present study aim to contribute to further development and understanding of a valid and easily applicable instrument that allows rapid and accurate screening of depressed individuals. In light of previous findings from the limited number of studies conducted in PHC settings with the BDI-II thus far, a two-factor structure (with somatic-affective and cognitive factors) was expected. Further, the diagnostic validity of this questionnaire was expected to meet criteria consistent with high-precision screening instruments.

Method

The sample consisted of 314 adult participants recruited from four primary health care offices in Zagreb, of which 204 were women (65 %). The age of participants ranged from 25 to 87 years, with an average age of 55.01 years (SD = 12.99). The level of educational attainment amongst participants was also assessed, with 12 % of participants (n = 38) having finished primary school only, 63 % (n = 198) having finished secondary school, and 25 % (n = 78) holding a degree in post-secondary education. All of the participants were Caucasian. Patients participating in this study were recruited during GP visits, the majority of whom were visiting their GP for various acute physical complaints and related social concerns, such as work-related sick leave. Exclusion criteria for participation included being below 18 or over 90 years of age, an inability to read and write Croatian, or the presence of mental retardation, dementia or other severe cognitive disabilities.

The Beck Depression Inventory-Second Edition (BDI-II; Beck et al., 1996) is a self-report measure used for the examination of the presence and severity of depressive symptoms. It contains 21 items in the form of statements that describe these symptoms. For each item, respondents are offered 4 statements and are asked to select the

statement that best describes their mood in the last 2 weeks. It is measured on a 4-point scale, with higher scores indicating greater severity of depressive symptoms and a maximum possible score of 63. The administration of the questionnaire is very simple and usually takes between 5 and 10 min.

In addition to the BDI-II, two further instruments were used in the present study. The Major Depression Inventory (MDI) and Doctor's Interview (DI) are two measures that are a part of the World Health Organization (WHO) Information Pack (WHO, 1998), a tool intended for the formal diagnosis of major depressive disorder in PHC settings. The MDI is a self-report questionnaire comprising 10 items, with two items (numbers 8 and 10) divided into two subitems, a and b. For these items, only the highest scores (either a or b) are included in statistical analysis. On each item, respondents are asked to report the amount of time various depressive symptoms were present during the past 14 days. Responses are made using a 6-point Likert scale, ranging from 0 (the symptom was not present at all) to 5 (the symptom was present all of the time). The maximum possible score is 50. The MDI includes a specific scoring algorithm that indicates the absence or presence of depression and, in the latter case, the severity of depression according to ICD-10 and/or DSM-IV criteria. Validation studies were conducted with psychiatric patients and have shown high reliability and diagnostic validity for this instrument (Bech et al., 2001; Olsen, Jensen, Noerholm, Martiny, & Bech, 2003). The DI is an interview conducted by the GP, consisting of 7 questions designed to help the physician confirm the diagnosis of major depressive disorder indicated by the MDI questionnaire. The DI is structured according to three symptom criteria (A, B and C) derived from the MDI, which are explained in detail on the instruments' official website (WHO, 2006).

Finally, a demographic data sheet that included information about age, gender, and educational status was completed by each participant.

Procedure

All data were gathered by the authors of the study in four PHC offices in Zagreb, Croatia. Two researchers were assigned to a single office, and each was present for 4 h on a given day. During this time, primary care physicians identified eligible patients from those visiting the office and invited these patients to participate in the study. Information regarding the purpose and procedure of the study was provided to patients by the GP according to a predetermined standardized protocol. Those who provided informed consent were taken by the researchers to a private room where they completed the demographic data sheet,

the BDI-II and the MDI. Those who met the MDI criteria for the presence of major depressive disorder, according to the DSM-IV, were questioned with the DI to confirm the diagnosis. Data were collected over a time period of 2 months. Primary care physicians were informed of the assessment results for these patients. All study procedures were conducted in accordance with ethical standards on human experimentation (World Medical Association Helsinki Declaration) and were approved by the ethics committee of the Faculty of Medicine at the University of Zagreb.

Data Analyses

Data analysis was performed using the statistical software package SPSS (Statistical Package for the Social Sciences) version 17, the AMOS 7 and MedCalc Software version 12. Visual inspection of all assessment measures following completion by the participants revealed there to be no missing data. This was confirmed by the descriptive statistics of the data.

The reliability of the BDI-II was analyzed using the internal consistency coefficient and corrected item-total correlations. The association between BDI-II scores and the socio-demographic characteristics of participants was determined by t test, Pearson r and Spearman rho correlations. T test was also used to test the significance of differences in total BDI-II scores between healthy and depressed participants. The Cohen-d value was calculated for determining effect size. In order to examine the factor structure of the BDI-II, the optimal number of factors was explored using: (a) Velicer's minimum average partial (MAP) test with a parallel test, (b) a criterion of eigenvalues over 1, and (c) model fit indices of confirmatory factor analysis (CFA). In addition, previously proposed models and the model revealed by our sample through explorative factor analysis (EFA) were compared using CFAs. The diagnostic validity of the BDI-II was investigated using MDI and DI as criterion measures for the presence of depression followed by ROC analysis to calculate the AUC, sensitivity, specificity, positive (PPP) and negative predictive power (NPP) as well as the optimal cut-off score.

Results

Depression Severity and Sociodemographic Characteristics

Based on the results of the MDI and subsequent DI as diagnostic criterion measures, 52 participants (16 % of the total 314 participants) were found to be currently suffering from depression. In contrast, a BDI-II score of 14 and

higher was achieved by 26.8 % of participants, which, according to the authors of the questionnaire, corresponds to at least a mild level of depression. The average BDI-II score for participants in this study was 10.35 (SD = 10.27), indicative of a minimal level of depression. The lowest score recorded was 0 and the highest was 51. Across BDI-II items, the lowest score was achieved on the 9th item (Suicidal Thoughts) while the highest was achieved on the 15th item (Loss of Energy).

While there was no statistically significant gender difference (t = 1.319, $p > 0.05$), the BDI-II score was positively associated with age (r = .12, $p < 0.05$) and negatively associated with level of educational attainment (rho = −.20, $p < 0.01$).

Reliability

Internal consistency of the BDI-II was analyzed using the Cronbach α, the indicator of the average intercorrelation among test items. The Cronbach α of the BDI-II achieved in the present study was .94, indicating very high reliability. Corrected BDI-II item-total correlations ranged from .44 (Loss of Interest in Sex) to .75 (Tiredness) with most correlations exceeding .60, a finding also suggestive of high internal consistency.

Factor Analysis

Velicer's minimum average partial (MAP) test compares the relative amount of systematic and unsystematic variance remaining in a correlation matrix after the extraction of an increasing number of components (O'Connor, 2000). The smallest average squared partial correlation that indicates the appropriate number of components was .0131, which was found in the case of the first factor (Table 1). Consistent with O'Connor's (2000) recommendation, parallel analyses were also conducted in order to compare the eigenvalues derived from the actual data in the MAP test and the eigenvalues of 1,000 random datasets. Factors are usually retained as long as the eigenvalue from the actual data is greater than the mean (or 95th percentile) eigenvalue of the random data. In the present study, this assumption was satisfied for one factor only (Table 1). Together with the results of the MAP test, this finding supports the one-factor model of depression.

However, principal axis factoring revealed the hypothesized two-factor solution with a total accounted variance of 47.34 % (F-I: 46.39 %; F-II: 5.76 %). The appropriateness of this factor analysis was confirmed by Bartlett's test of sphericity ($\chi^2 = 3,568.44$, $p = 0.000$), as well as by the high value of the Kaiser–Meyer–Olkin (KMO) coefficient (0.95). Consistent with our expectations, the present study obtained somatic-affective and cognitive factors

Table 1 Velicer's minimum average partial (MAP, 1976) test and parallel analysis (Horn, 1965) test for determining the BDI-II number of components

Factor	MAP test		Parallel test	
	Eigenvalues	Squared partial correlation	Mean of eigenvalues	95th percentile of eigenvalues
0		0.1947		
1	**9.7417**	**0.0131**	**1.4901**	**1.5728**
2	1.2191	0.0149	1.4040	1.4660
3	0.9906	0.0173	1.3376	1.3904
4	0.9288	0.0216	1.2814	1.3274
5	0.8751	0.0267	1.2309	1.2701
6	0.7898	0.0312	1.1834	1.2224
7	0.7563	0.0362	1.1395	1.1758
8	0.6654	0.0424	1.0971	1.1321
9	0.5863	0.0518	1.0563	1.0884
10	0.5654	0.0595	1.0182	1.0509
11	0.5235	0.0697	0.9811	1.0110
12	0.4874	0.0843	0.9443	0.9753
13	0.4362	0.1015	0.9083	0.9395
14	0.4212	0.1221	0.8719	0.9028
15	0.3937	0.1481	0.8364	0.8693
16	0.3404	0.1914	0.7996	0.8325
17	0.3004	0.252	0.7644	0.7983
18	0.2911	0.3284	0.7274	0.7632
19	0.2663	0.4899	0.6891	0.7255
20	0.2219	1	0.6459	0.6837
21	0.1993		0.5933	0.6385

Note In the MAP test, the smallest average squared partial correlation is .0131, indicating one as the optimal factor number (values highlited in bold). A one-factor solution is also supported by the fact that eigenvalues in the MAP test are larger than the eigenvalues of parallel analyses for only one factor

(Table 2). These two factors were found to correlate strongly ($r = .647$, $p < 0.01$).

The single-factor model, five 2-factor models and two 3-factor models were all tested using CFAs. Model fit indices for all models are shown in Table 3.

The 3-factor model proposed by Beck et al. (2002), consisting of cognitive, somatic and affective factors, achieved the best fit indices overall, while the 3-factor model obtained by Buckley et al. (2001) achieved similar values. When CMIN/df (χ^2) statistics are considered, all models demonstrated acceptable fit (CMIN/df was lower than 5 in all of models). For all tested models, GFI and TLI scores were lower than the minimally expected .90, but were moderately acceptable in both 3-factor models. In addition, these latter two were the only models with moderately acceptable CFI scores (>.90). For all models, RMSEA significantly differed from the expected interval (it was significantly larger than .05, $p = .001$). Figure 1 depicts factor loadings and factor

Table 2 Pattern matrix of principal axis factoring

	Somatic-affective	Cognitive
Suicidal thoughts	**0.79**	−0.14
Agitation	**0.79**	−0.09
Sadness	**0.75**	0.00
Loss of energy	**0.75**	0.01
Concentration difficulty	**0.72**	0.08
Tiredness	**0.70**	0.10
Loss of pleasure	**0.67**	0.13
Irritability	**0.64**	0.00
Changes in sleeping	**0.64**	−0.08
Pessimism	**0.58**	0.19
Indecisiveness	**0.53**	0.25
Changes in appetite	**0.50**	0.06
Self-dislike	**0.43**	0.33
Loss of interest	**0.43**	0.37
Crying	**0.40**	0.12
Loss of interest in sex	**0.32**	0.15
Guilty feelings	−0.21	**0.86**
Punishment feelings	−0.09	**0.73**
Worthlessness	0.22	**0.64**
Self-criticalness	0.13	**0.43**
Past failure	0.33	**0.40**

Note Promax with Kaiser normalization was used as a rotation method. Bold highlited are the salient (>.30) item loadings

intercorrelations for Beck et al.'s (2002) model, which achieved the best fit indices in relation to all other models. All factor loadings are significant and the intercorrelations between the three factors are high.

Interestingly, the 3-factor solution is not consistent with our expectations, which were based on previous studies conducted in PHC settings (e.g., Arnau et al., 2001; Viljoen et al., 2003). This is perhaps due to the fact that previous studies used exploratory factor analysis instead of CFA, which allows for the comparison of different factor models and the determination of the best fitting model.

Diagnostic Validity

Healthy and depressed participants, categorized according to the results of the MDI and DI as criterion measures, differed significantly ($t = 13.13$, $p < 0.001$) on the BDI-II scores, with a large effect size (Cohen $d = −3.04$, $r = −.84$). In other words, healthy participants ($M = 6.92$, $SD = 5.590$) had, on average, lower BDI-II scores than depressed participants ($M = 27.62$, $SD = 11.090$). According to the cut-off values proposed by Beck et al. (1996), participants with depressive disorder, diagnosed according to the criterion measures, had average BDI-II scores that indicated moderate levels of depression.

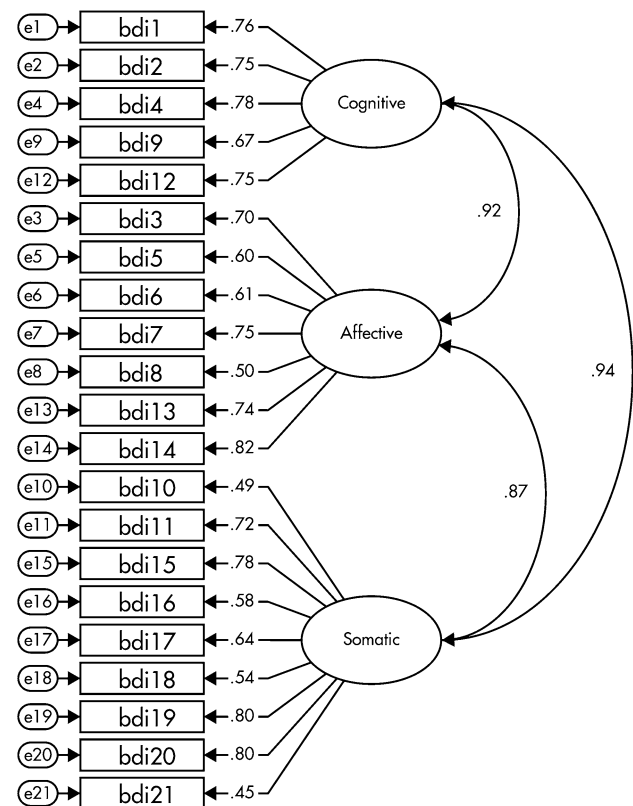**Table 3** Goodness of fit indices for various BDI-II models

	χ^{2a}	df	χ^2/df	GFI	TLI	CFI	RMSEA	(90 % CI)
Three-factor model (Beck et al., 2002)	526.163	186	2.829	.864	.889	.901	.076	(.069–.084, $p = .000$)
Three-factor model (Buckley et al., 2001)	531.201	186	2.856	.861	.887	.900	.077	(.069–.085, $p = .000$)
Two-factor clinical model (Beck et al., 1996)	559.598	188	2.977	.855	.880	.892	.079	(.072–.087, $p = .000$)
Two-factor non-clinical model (Beck et al., 1996)	572.662	188	3.046	.852	.875	.889	.081	(.073–.088, $p = .000$)
Two-factor model (Arnau et al., 2001)	572.311	187	3.060	.855	.875	.888	.081	(.074–.089, $p = .000$)
Two-factor model (Viljoen et al., 2003)	570.035	186	3.065	.855	.874	.889	.081	(.074–.089, $p = .000$)
One factor model	608.944	189	3.222	.842	.865	.878	.084	(.077–.092, $p = .000$)
Two-factor model of principal axis factoring	606.927	188	3.228	.842	.864	.879	.084	(.077–.092, $p = .000$)

df Degree of freedom, χ^2/df CMIN/df values (Bentler, 1980), *GFI* Goodness of fit index (Jöreskog & Sörbom, 1993), *TLI* Tucker-Lewis index (Bentler, 1980; Tucker & Lewis, 1973), *CFI* Comparative Fit Index (Bentler, 1980), *RMSEA* root mean square error of approximation and its 90 % confidence interval (Bollen & Long, 1993; Steiger, 1990)

[a] The associated *p* values were all lower than .001

A Receiver Operating Characteristics (ROC) analysis was conducted to determine the diagnostic characteristics of the BDI-II as a screening instrument that discriminates participants with and without depressive disorder diagnoses obtained on the basis of the previously described criterion measures. Here, the AUC value was 0.96 (95 % CI .94 to .98) and there was a statistically significant difference from the area under the curve of the instrument that would discriminate the participants by chance (z = 46.92, $p < 0.001$). The parameters of sensitivity, specificity, positive predictive power (PPP) and negative predictive power (NPP) for all possible total scores were also considered in order to identify the optimal critical value for discriminating non-depressed and depressed individuals in this sample. Using the results of this analysis, shown in Table 4, the maximum Youden index value (sensitivity + specificity − 100) of 79.68 is achieved at a critical value of 15/16. This result demonstrates that the highest diagnostic accuracy of the BDI-II in differentiating non-depressed and depressed individuals was achieved at the critical value of 15/16, where the best balance between sensitivity (88.46) and specificity (91.22) is indicated. In addition, the PPP at this critical value was 66.7 % while the amount of NPP was very high (97.6 %). These results confirm our expectation that the BDI-II is a screening instrument with high levels of diagnostic validity. A score of 16 or higher, indicative of at least a mild level of depression, was achieved by 22 % of the participants.

Discussion

The results of the present study indicated that 16 % of the sample met criteria for depression, according to the MDI and DI. A BDI-II score of 14 or higher (recommended in the BDI-II manual as a cut-off score for mild depression) was achieved by 26.8 % of subjects, while a score of 16 or higher (the cut-off score based on the ROC curve) was achieved by 22 % of subjects. Together, these results indicate somewhat higher prevalence rates for depressive disorder in this study than those found in PHC patients in research from other national and cultural communities (e.g., Arnau et al., 2001; Dutton et al., 2004), particularly in the case of the BDI-II scores. There are several potential explanations for these results. The first potential reason for

Fig. 1 Factor loadings and factor intercorrelations for the 3-factor model based on Beck et al. (2002)

Table 4 Critical values, sensitivity, specificity, positive predictive power (PPP) and negative predictive power (NPP) in distinguishing between non-depressed and depressed subjects

Critical value	Sensitivity	Specificity	PPP	NPP
≥0	100.00	0.00	16.6	
>0	100.00	14.89	18.9	100.0
>1	100.00	19.08	19.7	100.0
>2	100.00	25.57	21.1	100.0
>3	100.00	32.06	22.6	100.0
>4	100.00	37.02	24.0	100.0
>5	100.00	43.51	26.0	100.0
>6	100.00	51.53	29.1	100.0
>7	100.00	59.16	32.7	100.0
>8	100.00	68.70	38.8	100.0
>9	98.08	74.43	43.2	99.5
>10	98.08	78.63	47.7	99.5
>11	96.15	80.92	50.0	99.1
>12	94.23	83.21	52.7	98.6
>13	92.31	86.26	57.1	98.3
>14	88.46	88.93	61.3	97.5
>15	**88.46**	**91.22**	**66.7**	**97.6**
>16	86.54	91.60	67.2	97.2
>17	78.85	95.04	75.9	95.8
>18	76.92	95.42	76.9	95.4
>19	73.08	96.18	79.2	94.7
>20	63.46	96.95	80.5	93.0
>21	57.69	98.09	85.7	92.1
>22	55.77	99.24	93.5	91.9
>23	53.85	100.00	100.0	91.6
>26	46.15	100.00	100.0	90.3
>30	40.38	100.00	100.0	89.4
>31	34.62	100.00	100.0	88.5
>33	30.77	100.00	100.0	87.9
>34	26.92	100.00	100.0	87.3
>35	25.00	100.00	100.0	87.0
>37	21.15	100.00	100.0	86.5
>40	19.23	100.00	100.0	86.2
>41	17.31	100.00	100.0	85.9
>42	11.54	100.00	100.0	85.1
>43	9.62	100.00	100.0	84.8
>46	1.92	100.00	100.0	83.7
>51	0.00	100.00		

Note Bold highlited values indicate the optimal cut-off score

this finding relates to environmental factors specific to the Croatian context that might produce stressful circumstances. As previously mentioned, Croatia is an Eastern European country in a post-transitional period with recent experience in war and post-war consequences and currently experiencing an economic recession. However, an alternative explanation for these results might be found in a consideration of

the structure of the sample in the present study. For the most part, the participants in this study were PHC patients who were visiting their GP with acute medical complaints or conditions, a factor that might have elevated the prevalence of depressive disorder. Indeed, the achieved prevalence rates in the present study are comparable to those from a study conducted with a sample of Croatian patients suffering from a chronic somatic illness, in which 29 % of patients achieved BDI-II results above the cut-off score (Filipcic et al., 2007). Furthermore, the observed discrepancy, in relation to the prevalence of depression, between the criterion measures (MDI and DI) and the BDI-II might be due to the fact that the BDI-II is more highly saturated with items describing somatic complaints and thus slightly over-estimates the number of depressed individuals, particularly amongst a sample of PHC patients. This criticism of the BDI-II has similarly been raised in a recent study examining the BDI-II as a screening measure for mood disorders in pregnant women (Curzik & Jokic Begic, 2012).

The highest score on the BDI-II was achieved on the Loss of Energy item, which is not surprising considering that energy loss is a common symptom of numerous health problems. More specifically, this finding might be the result of the fact that participants, PHC patients visiting their GP with an acute medical complaint, were interviewed in a situation in which they were experiencing some form of physical or mental discomfort. It might also be argued that the over-expression of this symptom can be explained by the collectivist culture to which Croatia belongs, a context in which the reporting of physical symptoms such as energy loss might be more prevalent. An extensive meta-analysis (Oyserman, Coon, & Kemmelmeier, 2002) demonstrated that Southern European cultures presented higher levels of collectivism than northern European countries (e.g., Norway) or English-speaking countries (e.g., Australia). In a similar examination on the differences in social support amongst European countries, the ODIN study also reported results pointing in a similar direction (Lehtinen et al., 2003). In such a cultural context, it seems logical that symptoms demonstrating a help-seeking message are more dominant. Conversely, the lowest score on the BDI-II was achieved on the Suicidal Thoughts item. This is perhaps a similarly unsurprising finding, where suicide is generally a final choice when people do not find other solutions to their problems. This symptom is present at similarly low levels in northern European countries (Nuevo et al., 2009), in which a more highly individualist culture encourages the independent resolution of one's problems.

Although depression is more common among women than men, a significant gender difference in total BDI-II scores was not obtained in the present sample. Findings from other studies in this regard are inconsistent (Dozois et al.,

1998; Arnau et al., 2001; Osman, Kopper, Barrios, Gutierrez, & Bagge, 2004; VanVoorhis & Blumentritt, 2007).

In the present study, the correlation between age and BDI-II score was low and positive, while the correlation between level of educational attainment and total score was low and negative. Similar results were obtained in Arnau et al.'s (2001) study with a PHC sample, while other studies conducted with different samples provided inconsistent results concerning the association between age and BDI-II scores (Steer, Kumar, Ranieri, & Beck, 1998; Kojima et al., 2002; VanVoorhis & Blumentritt, 2007).

The BDI-II proved to be a highly reliable instrument, consistent with the estimates of reliability in other studies conducted with general PHC samples, where reliability estimates are generally around 0.90 (e.g., Arnau et al., 2001; Dutton et al., 2004).

In this and other studies, the findings with regards to the structure of the scale and to the existence of subscales are somewhat contradictory. While high internal reliability and Velicer's MAP test (along with a parallel test, as recommended by O'Connor (2000)) both confirmed the adequacy of computing a total depression score as a reliable indicator of depression, the eigenvalue criterion of principal axis factoring suggested that a two-factor solution should be used. Consistent with the results from the present study, a two-factor structure with somatic-affective and cognitive dimensions was obtained in studies using comparable general PHC samples (Arnau et al., 2001; Viljoen et al., 2003), as well as in studies using other non-clinical samples (e.g., Kojima et al., 2002; Uslu et al., 2008). However, in the present study, confirmatory factor analysis demonstrated that the two 3-factor models have the best adequate fit indices, with the model proposed by Beck et al. (2002) achieving the best overall values of the various fit indices. A similar model obtained by Buckley et al. (2001) was also found to be quite satisfactory. The existence of three factors has also been found in other BDI-II studies. In one study using a sample of adolescent psychiatric outpatients, negative attitude, performance difficulty and somatic elements were found as factors (Osman et al., 2004), a result confirmed by Carmody (2005) using a student sample. Finally, a recent large study using confirmatory factor analysis also yielded a three-factor structure (with cognitive, affective and somatic factors), although the authors also suggested the need for a shortened version of the BDI-II (Vanheule, Desmet, Groenvynck, Rosseel, & Fontaine, 2008).

The discrepancies between factor solutions observed using different analytical methods are partly a consequence of high inter-item correlations and high associations between cognitive, somatic and affective factors. These findings support the use of the total BDI-II score in clinical settings. In addition, there are two important advantages to using this total score in primary care specifically. First, the total score covers the broad clinical picture of major depressive disorder based on DSM-IV criteria. Second, a number of studies, including the present study, have empirically derived critical values using the whole instrument but not its subscales. However, the findings of the present study also suggest that three subscales with a good CFA fit can be used by researchers and clinicians interested in a straightforward and unambiguous assessment of the cognitive, affective, and somatic symptoms of depression measured by the BDI-II. This becomes especially relevant in light of the fact that these subscales have shown unique relations with different psychological variables, such as alexithymia (Vanheule, Desment, Verhaeghe, & Bogaerts, 2007) and autobiographical memory (Mackinger & Svaldi, 2004). Indeed, more research is needed to develop a clearer picture of each subdimension of depression as well as their potential sensitivity to different antidepressant treatment options used in primary care.

In order to verify the usefulness of the BDI-II as a screening instrument for depression in a PHC population, various parameters of diagnostic validity were analyzed. When differentiating healthy and depressed patients, the AUC value was very high, thus locating the BDI-II in the category of high-precision diagnostic instruments (Streiner & Cairney, 2007). Specifically, the findings indicated a 96 % chance for a randomly selected individual from the depressed group to have a higher overall score than a randomly selected individual from the healthy group. Other studies conducted with different populations have also demonstrated relatively high levels of general diagnostic informativeness for this instrument, which usually ranges between 0.86 and 0.96 (e.g., Arnau et al., 2001; Kumar, Steer, Teitelman, & Villacis, 2002; Uslu et al., 2008). The results also indicated that the largest possible sum of the amount of sensitivity and specificity was obtained with a critical value of 15/16, meaning that the total BDI-II scores equal to and above 16 were indicative of an increased probability of depressive disorder in patients from PHC. Furthermore, the levels of both sensitivity and specificity were approximately equal and very satisfactory at this critical value. Despite a certain number of false positive results, these findings suggest that, if a patients' score is below the critical value, one can infer with very high certainty that he or she is not clinically depressed, a conclusion reflected in the very high value of NPP. Arguably, this is a more critical feature for instruments used for screening purposes in PHC, where the goal is to avoid a failure to detect patients who suffer from this disorder.

It should be kept in mind that the decision to use a certain critical value must be based on the unique characteristics of the population being served and the purpose for which the instrument is used (Pintea and Moldovan, 2009). To date, only a few studies have examined the parameters

of diagnostic validity of the BDI-II in the context of PHC, reporting critical values of 14 (Dutton et al., 2004) and 18 (Arnau et al., 2001). In Dutton et al.'s (2004) study, the percentage of depressed PHC patients was significantly higher than the average prevalence of depression generally reported in PHC. Therefore, the authors proposed the use of a higher critical value, similar to that found in the present study, where the prevalence of depressive patients (16 %) was similar to the expected prevalence. Determining the most appropriate critical value also depends on the purpose of the instrument being examined. In the case of screening measures for use in PHC, it is perhaps most critical that such instruments have a high amount of sensitivity in order to minimize the number of false negative outcomes while at the same time retaining a reasonable level of specificity. For example, in the present study, a critical value of 12 would result in very high sensitivity and NPP, making it nearly impossible to fail to identify patients with depressive disorder. However, the application of this critical value would mean reducing specificity and PPP so that 50 % of patients labeled as depressed would actually be healthy, thus resulting in unnecessary further treatment for these individuals.

The critical importance of the implementation of effective screening for depression in PHC is reflected in research demonstrating that GPs sometimes fail to detect up to more than half of depressed patients (Cepoiu et al., 2008; Jackson et al., 2007). Research has also shown that the use of screening measures increases the detection of depression in PHC (Pignone et al., 2002; Gilbody, Whitty, Grimshaw, & Thomas, 2003), but that systematic screening should also be an integral part of depression support programs in order to achieve a significant reduction of depressive symptoms, an improvement in social and work-specific functioning, and a reduction in mortality rates of PHC patients (Gilbody, House, and Sheldon, 2008; O'Connor et al., 2009; Mitchell, 2012; Sikorski, Luppa, König, van den Bussche, & Riedel-Heller, 2012). These programs include various aspects of depression care such as systematic screening, office staff training, individualized evidence-based treatment, closer monitoring of the patient, and available mental health referral. It has also been shown that depression-specific instruments such as the BDI-II and high-risk screening instruments that use empirically derived critical values have a greater effect in depression care (Gilbody et al., 2008). In the context of the present study, Croatian PHC, the BDI-II could be easily used due to the small amount of time required to complete the measure. In this setting, the instrument might be administered to patients waiting to see the GP, thus avoiding unnecessary extra time with the GP. Those individuals who scored above the optimal critical value of 15 (i.e., 16 and above) would be next examined using a diagnostic interview for final confirmation of the presence of depressive disorders. Furthermore, specific results from each of the three BDI-II factors could help the GP to better determine the specific nature of each patient's depressive disorder by providing information about which of the three types of symptoms (affective, cognitive, and/or somatic) are predominant in the clinical picture. This information could potentially allow the GP to provide the most appropriate antidepressant treatment, although more research is needed before such conclusions can be made. An additional advantage of the BDI-II is that it is often used as an antidepressant treatment outcome measure in various populations, including PHC patients (e.g., de Graaf et al., 2009; Reeves, Rohan, Langenberg, Snitker, & Postolache, 2012).

A number of limitations to the present study should be considered alongside the presented findings. As previously mentioned, a self-report instrument (MDI) was used as a criterion measure against which the diagnostic validity of the BDI-II was examined. Due to their shared method variance, the diagnostic usefulness of the BDI-II might have been somewhat inflated. It is perhaps more methodologically sound to use validated diagnostic interviews such as the Structured Clinical Interview for DSM-IV (SCID-IV; Spitzer, Williams, Gibbon, & First, 1994) as the criterion measure. In addition, the participation refusal rate was not recorded for practical reasons. Although the vast majority of patients approached agreed to participate in the study, the number and depression-related characteristics of those who refused to participate is unknown and thus limits the generalization of the findings on the prevalence of depression and the diagnostic parameters of the BDI-II, including the size of the critical value. Furthermore, the use of a convenience sampling method in the present study may have influenced the response rate and style of the participants. Indeed, one might hypothesize that individuals suffering from depression and related mental disorders might be more prone to agreeing to take part in such a study, perhaps because they expect some form of psychosocial benefit from participation. If this was the case, the prevalence of depression in the whole sample would have been inflated. Finally, in light of the confirmation of the 3-factor structure of BDI-II achieved in the present study, it might be useful to conduct future ROC analyses using these three subscales separately.

On the whole, the findings of the present study suggest that the Croatian version of the BDI-II is a highly reliable instrument, with satisfying structural validity and high diagnostic accuracy. As such, it supports the possibility of using the BDI-II as a screening instrument in Croatian primary health care settings.

Aaron T. Beck with permission of publisher, Harcourt Assessment, Inc. Reproduced with permission. All rights reserved. "Beck Depression Inventory" and "BDI" are trademarks of Harcourt Assessment, Inc. registered in the United States of America and/or other jurisdictions.

References

Al-Windi, A. (2005). Depression in general practice. *Nordic Journal of Psychiatry, 59*, 272–277.

Applied Health Sciences (Mental Health), University of Aberdeen. (2011). *Assessing the validity of the PHQ-9, HADS, BDI-II and QIDS-SR$_{16}$ in measuring severity of depression in a UK sample of primary care patients with a diagnosis of depression.* Edinburgh: NHS Quality Improvement Scotland (NHS QIS).

Arnarson, T. O., Olason, D. T., Smári, J., & Sigurethsson, J. F. (2008). The Beck Depression Inventory Second Edition (BDI-II): Psychometric properties in Icelandic student and patient populations. *Nordic Journal of Psychiatry, 62*, 360–365.

Arnau, R. C., Meagher, M. W., Norris, M. P., & Bramson, R. (2001). Psychometric evaluation of the Beck Depression Inventory-II with primary care medical patients. *Health Psychology, 20*, 112–119.

Bauer, M., Bschor, T., Pfennig, A., Whybrow, P. C., Angst, J., & Versiani, M. (2007). World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for biological treatment of unipolar depressive disorders in primary care. *The World Journal of Biological Psychiatry, 8*, 67–104.

Bech, P., Rasmussen, N. A., Raabaek Olsen, L., Noerholm, V., & Abildgaard, W. (2001). The sensitivity and specificity of the Major Depression Inventory, using the Present State Examination as the index of diagnostic validity. *Journal of Affective Disorders, 66*, 159–164.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II.* San Antonio: Psychological Corporation.

Beck, A. T., Steer, R. A., Brown, G. K., & Van der Does, A. J. W. (2002). *BDI-II-NL Handleiding [BDI-II-Dutch Manual].* Lisse: Psychological Corporation.

Bedi, R. P., Koopman, R. F., & Thompson, J. M. (2001). The dimensionality of the Beck Depression Inventory-II and its relevance for tailoring the psychological treatment of women with depression. *Psychotherapy: Theory, Research, Practice, Training, 38*, 306–318.

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modelling. *Annual Review of Psychology, 31*, 419–456.

Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models.* New York: Sage.

Buckley, T. C., Parker, J. D., & Heggie, J. (2001). A psychometric evaluation of the BDI-II in treatment-seeking substance abusers. *Journal of Substance Abuse Treatment, 20*, 197–204.

Bunevicius, A., Staniute, M., Brozaitiene, J., & Bunevicius, R. (2012). Diagnostic accuracy of self-rating scales for screening of depression in coronary artery disease patients. *Journal of Psychosomatic Research, 72*, 5–22.

Campos, R. C., & Goncalves, B. (2011). The Portuguese version of the Beck Depression Inventory-II (BDI-II)—preliminary psychometric data with two nonclinical samples. *European Journal of Psychological Assessment, 27*, 258–264.

Carmody, D. P. (2005). Psychometric characteristics of the Beck Depression Inventory-II with college students of diverse ethnicity. *International Journal of Psychiatry in Clinical Practice, 9*, 22–28.

Cepoiu, M., McCusker, J., Cole, M. G., Sewitch, M., Belzile, E., & Ciampi, A. (2008). Recognition of depression by non-psychiatric physicians—a systematic literature review and meta-analysis. *Journal of General Internal Medicine, 23*, 25–36.

Curzik, D., & Jokic Begic, N. (2012). The utility of BDI-II in assessment of pre- and postpartum depression symptoms and their relation to labor pain. *Psychiatria Danubina, 24*, 167–174.

de Graaf, L. E., Gerhards, S. A., Arntz, A., Riper, H., Metsemakers, J. F., Evers, S. M., et al. (2009). Clinical effectiveness of online computerised cognitive-behavioural therapy without support for depression in primary care: Randomised trial. *British Journal of Psychiatry, 195*, 73–80.

Dozois, D. J. A., Dobson, K. S., & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck Depression Inventory-II. *Psychological Assessment, 10*, 83–89.

Dutton, G. R., Grothe, K. B., Jones, G. N., Whitehead, D., Kendra, K., & Brantley, P. J. (2004). Use of the Beck Depression Inventory-II with African American primary care patients. *General Hospital Psychiatry, 26*, 437–442.

Filipcic, I., Popovic-Grle, S., Marcinko, D., Basic, S., Hotujac, L., Pavicic, F., et al. (2007). Screening for depression disorders in patients with chronic somatic illness. *Collegium Antropologicum 31*, 139–143.

Gaynes, B. N., Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Balasubramani, G. K., McGrath, P. J., et al. (2008). Primary versus specialty care outcomes for depressed outpatients managed with measurement-based care: Results from STAR*D. *Journal of General Internal Medicine, 23*, 551–560.

Gaynes, B.N., Rush, A.J., Trivedi, M., Wisniewski, S.R., Balasubramani, G.K., Spencer, D.C., et al. (2005). A direct comparison of presenting characteristics of depressed outpatients from primary vs. specialty care settings: preliminary findings from the STAR*D clinical trial. *General Hospital Psychiatry, 27*, 87–96.

Gilbody, S., Sheldon, T., & House, A. (2008). Screening and case-finding instruments for depression: A meta-analysis. *Canadian Medical Association Journal, 178*, 997–1003.

Gilbody, S. M., Whitty, P. M., Grimshaw, J. M., & Thomas, R. E. (2003). Improving the detection and management of depression in primary care. *Quality and Safety in Health Care, 12*, 149–155.

Hecimovic, H., Bosnjak, J., & Demarin, V. (2008). Prevalence of mood dysfunction in epilepsy patients in Croatia. *Collegium Antropologicum, 32*, 65–68.

Huffman, J. C., Doughty, C. T., Januzzi, J. L., Pirl, W. F., Smith, F. A., & Fricchione, G. L. (2010). Screening for major depression in post-myocardial infarction patients: Operating characteristics of the Beck Depression Inventory-II. *International Journal of Psychiatry in Medicine, 40*, 187–197.

Jackson, J. L., Passamonti, M., & Kroenke, K. (2007). Outcome and impact of mental disorders in primary care at 5 years. *Psychosomatic Medicine, 69*, 270–276.

Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command LANGUAGE.* Chicago: Scientific Software International Inc.

Kapci, E. G., Uslu, R., Turkcapar, H., & Karaoglan, A. (2008). Beck Depression Inventory II: Evaluation of the psychometric properties and cut-off points in a Turkish adult population. *Depression and Anxiety, 25*, 104–110.

Katon, W. J. (2003). Clinical and health services relationships between major depression, depressive symptoms, and general medical illness. *Biological Psychiatry, 54*, 216–226.

King, M., Nazareth, I., Levy, G., Walker, C., Morris, R., Weich, S., et al. (2008). Prevalence of common mental disorders in general practice attendees across Europe. *The British Journal of Psychiatry, 192*, 362–367.

Kojima, M., Furukawa, T. A., Takahashi, H., Kawai, M., Nagaya, T., & Tokudome, S. (2002). Cross-cultural validation of the Beck Depression Inventory-II in Japan. *Psychiatry Research, 110*, 291–299.

Kolaric, B., Tesic, V., Ivankovic, D., & Begovac, J. (2006). Prevalence of moderate and severe depression among Croatian

patients infected with human immunodeficiency virus. *Collegium Antropologicum, 30*, 85–88.

Kumar, G., Steer, R. A., Teitelman, K. B., & Villacis, L. (2002). Effectiveness of Beck Depression Inventory–II subscales in screening for major depressive disorders in adolescent psychiatric inpatients. *Assessment, 9*, 164–170.

Lehtinen, V., Michalak, E., Wilkinson, C., Dowrick, C., Ayuso-Mateos, J. L., Dalgard, O. S., et al. (2003). Urban-rural differences in the occurrence of female depressive disorder in Europe—evidence from the ODIN study. *Social Psychiatry and Psychiatric Epidemiology, 38*, 283–289.

Licht-Strunk, E., Beekman, A. T., de Haan, M., & van Marwijk, H. W. (2009). The prognosis of undetected depression in older general practice patients. A one year follow-up study. *Journal of Affective Disorders, 114*, 310–315.

Mackinger, H. F., & Svaldi, J. J. (2004). Autobiographical memory predicts cognitive but not somatic change in sleep apnea patients vulnerable for affective disorder. *Journal of Affective Disorders, 81*, 17–22.

Mitchell, A. J. (2012). Clinical utility of screening for clinical depression and bipolar disorder. *Current Opinion in Psychiatry, 25*, 24–31.

NHS Employers and the General Practitioners' Committee. (2009). Quality and outcome frameworks: Guidance for GMS contract 2009/10.

Nuevo, R., Dunn, G., Dowrick, C., Vazquez-Barquero, J. L., Casey, P., Dalgard, O. S., et al. (2009). Cross-cultural equivalence of the Beck Depression Inventory: A five-country analysis from the ODIN study. *Journal of Affective Disorders, 114*, 156–162.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers, 32*, 396–402.

O'Connor, E. A., Whitlock, E. P., Beil, T. L., & Gaynes, B. N. (2009). Screening for depression in adult patients in primary care settings: A systematic evidence review. *Annals of Internal Medicine, 151*, 793–803.

Olsen, L. R., Jensen, D. V., Noerholm, V., Martiny, K., & Bech, P. (2003). The internal and external validity of the Major Depression Inventory in measuring severity of depressive states. *Psychological Medicine, 33*, 351–356.

Osman, A., Kopper, B. A., Barrios, F., Gutierrez, P. M., & Bagge, C. L. (2004). Reliability and validity of the Beck Depression Inventory-II with adolescent psychiatric inpatients. *Psychological Assessment, 16*, 120–132.

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin, 128*, 72–73.

Pibernik-Okanovic, M., Peros, K., Szabo, S., Begic, D., & Metelko, Z. (2005). Depression in Croatian Type 2 diabetic patients: prevalence and risk factors. A Croatian survey from the European Depression in Diabetes (EDID) Research Consortium. *Diabetic Medicine, 22*, 942–945.

Pignone, M. P., Gaynes, B. N., Rushton, J. L., Burchell, C. M., Orleans, C. T., Mulrow, C. D., et al. (2002). Screening for depression in adults: A summary of the evidence for the U.S. Preventive Services Task Force. *Annals of Internal Medicine, 136*, 765–776.

Pintea, S., & Moldovan, R. (2009). The receiver-operating characteristic (ROC) analysis: Fundamentals and applications in clinical psychology. *Journal of Cognitive and Behavioral Psychotherapies, 9*, 49–66.

Priebe, S., Bogic, M., Ajdukovic, D., Franciskovic, T., Galeazzi, G. M., Kucukalic, A., et al. (2010). Mental disorders following war in the Balkans. *Archives of General Psychiatry, 67*, 518–528.

Reeves, G. M., Rohan, K. J., Langenberg, P., Snitker, S., & Postolache, T. T. (2012). Calibration of response and remission cut-points on the Beck Depression Inventory-Second Edition for monitoring seasonal affective disorder treatment outcomes. *Journal of Affective Disorders, 138*, 123–127.

Richards, D. (2011). Prevalence and clinical course of depression: A review. *Clinical Psychology Review, 31*, 1117–1125.

Segal, D. L., Coolidge, F. L., Cahill, B. S., & O'Riley, A. A. (2008). Psychometric properties of the Beck Depression Inventory—II (BDI-II) among community-dwelling older adults. *Behavior Modification, 32*, 3–20.

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*, 1001–1009.

Sikorski, C., Luppa, M., König, H. H., van den Bussche, H., & Riedel-Heller, S. G. (2012). Does GP training in depression care affect patient outcome? A systematic review and meta-analysis. *BMC Health Services Research*. doi:10.1186/1472-6963-12-10.

Spitzer, R. L., Williams, J. B., Gibbon, M., & First, M. B. (1994). *Structured clinical interview for DSM-IV axis I disorders—patient edition*. New York: New York State Psychiatric Institute.

Steer, R. A., Kumar, G., Ranieri, W. F., & Beck, A. T. (1998). Use of the Beck Depression Inventory-II with adolescent psychiatric outpatients. *Journal of Psychopathology and Behavioral Assessment, 20*, 127–137.

Steer, R. A., Rissmiller, D. J., & Beck, A. T. (2000). Use of the Beck Depression Inventory-II with depressed geriatric inpatients. *Behavior Research and Therapy, 38*, 311–318.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173–180.

Stojanovic-Spehar, S., Blazekovic-Milakovic, S., Amerl-Sakic, V., Kolic, N., & Supe, S. (2009). Depression prevalence and estimation of psychosocial parameters within adult population in City of Zagreb. *Psychiatria Danubina, 21*, 497–507.

Streiner, D. L., & Cairney, J. (2007). What's under the ROC? An introduction to Receiver Operating Characteristics Curves. *The Canadian Journal of Psychiatry, 52*, 121–128.

Tarricone, I., Stivanello, E., Poggi, F., Castorini, V., Marseglia, M. V., Fantini, M. P., et al. (2012). Ethnic variation in the prevalence of depression and anxiety in primary care: A systematic review and meta-analysis. *Psychiatry Research, 195*, 91–106.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analyses. *Psychometrika, 18*, 1–10.

Uslu, R. I., Kapci, E. G., Oncu, B., Ugurlu, M., & Turkcapar, H. (2008). Psychometric properties and cut-off scores of the Beck Depression Inventory-II in Turkish Adolescents. *Journal of Clinical Psychology in Medical Settings, 15*, 225–233.

Vanheule, S., Desmet, M., Groenvynck, H., Rosseel, Y., & Fontaine, J. (2008). The factor structure of the Beck Depression Inventory II: An evaluation. *Assessment, 15*, 177–187.

Vanheule, S., Desmet, M., Verhaeghe, P., & Bogaerts, S. (2007). Alexithymic depression: Evidence for a depression subtype? *Psychotherapy and Psychosomatics, 76*, 135–136.

VanVoorhis, C. R. W., & Blumentritt, T. L. (2007). Psychometric properties of the Beck Depression Inventory-II in a clinically identified sample of Mexican American adolescents. *Journal of Child and Family Studies, 16*, 789–798.

Viljoen, J. L., Iverson, G. L., Griffiths, S., & Woodward, T. S. (2003). Factor structure of the Beck Depression Inventory-II in a medical outpatient sample. *Journal of Clinical Psychology in Medical Settings, 10*, 289–291.

Wancata, J., & Friedrich, F. (2011). Depression: a diagnosis aptly used? *Psychiatria Danubina, 23*, 406–411.

Weissman, M. M., Neria, Y., Gameroff, M. J., Pilowsky, D. J., Wickramaratne, P., Lantigua, R., et al. (2010). Positive screens

for psychiatric disorders in primary care: a long-term follow-up of patients who were not in treatment. *Psychiatric Services, 61*, 151–159.

Whisman, M. A., Perez, J. E., & Ramel, W. (2000). Factor structure of the Beck Depression Inventory-Second Edition (BDI-II) in a student sample. *Journal of Clinical Psychology, 56*, 545–551.

Wiebe, J. S., & Penley, J. A. (2005). A psychometric comparison of the Beck Depression Inventory-II in English and Spanish. *Psychological Assessment, 17*, 481–485.

World Health Organization (WHO). (1998). *Info package: Mastering depression in primary care*. Frederiksborg: WHO, Regional Office for Europe, Psychiatric Research Unit.

World Health Organization (WHO). (2001). What is depression? In *Depression*. Retrieved from http://www.who.int/mental_health/management/depression/definition/en/.

World Health Organization (WHO). (2006). *Mastering depression in primary care* (version 22). Retrieved from http://www.gp-training.net/protocol/psychiatry/who/whodep.htm#A.