Comparison of two self-rating scales to detect depression: HADS and PHQ-9

Maja Hansson, Jayanti Chotai, Annika Nordstöm and Owe Bodlund

ABSTRACT

Background

More than half of patients with depression go undetected. Self-rating scales can be useful in screening for depression, and measuring severity and treatment outcome.

Aim

This study compares the Hospital Anxiety and Depression Scale (HADS) and the Patient Health Questionnaire (PHQ-9) with regard to their psychometric properties, and investigates their agreement at different cut-off scores.

Method

Swedish primary care patients and psychiatric outpatients ($n = 737$) who reported symptoms of depression completed the self-rating scales. Data were collected from 2006 to 2007. Analyses with respect to internal consistency, factor analysis, and agreement (Cohen's κ) at recommended cut-offs were performed.

Results

Both scales had high internal consistency ($\alpha = 0.9$) and stable factor structures. Using severity cut-offs, the PHQ-9 (≥ 5) diagnosed about 30% more patients than the HADS depression subscale (HADS-D; ≥ 8). They recognised the same prevalence of mild and moderate depression, but differed in relation to severe depression. When comparing recommended screening cut-offs, HADS-D ≥ 11 (33.5% of participants) and PHQ-9 ≥ 10 (65.9%) agreement was low ($\kappa = 0.35$). Using the lower recommended cut-off in the HADS-D (≥ 8), agreement with PHQ-9 ≥ 10 was moderate ($\kappa = 0.52$). The highest agreement ($\kappa = 0.56$) was found comparing HADS-D ≥ 8 with PHQ-9 ≥ 12 . This also equalised the prevalence of depression found by the scales.

Conclusion

The HADS and PHQ-9 are both quick and reliable. The HADS has the advantage of evaluating both depression and anxiety, and the PHQ-9 of being strictly based upon the *Diagnostic and Statistical Manual of Mental Disorders*. The agreement between the scales at the best suitable cut-off is moderate, although the identified prevalence was similar. This indicates that the scales do not fully identify the same cases. This difference needs to be further explored.

Keywords

depression; primary health care; psychometrics; self-assessment.

INTRODUCTION

For GPs, identifying patients with depression and monitoring their treatment is an ongoing challenge. To facilitate this process there are several self-rating scales with the ability to screen for depression and anxiety. Depression is one of the most common illnesses, with a point prevalence around 15% in primary care and 5% in the general population.¹⁻⁶ In Sweden, about 25% of primary care patients experience depression and/or anxiety,^{3,7-8} and patients with mild to moderate depression are primarily treated in primary care.

It is well known that there are difficulties in diagnosing patients with depression in primary care, and only about half of patients with depression are identified as such by GPs.^{1,9,10} The time allocated for each patient in Swedish primary care is about 20–30 minutes. Many of the patients with depression also have somatic complaints, which make it difficult to focus on concurrent mental problems and to identify depression within this time limit.¹¹ There is evidence that self-rating scales like the Hospital Anxiety and Depression Scale (HADS),¹²⁻¹⁴ Primary Care Evaluation of Mental Disorders (PRIME-MD),⁹ and the Patient Health Questionnaire (PHQ-9)^{15,16} are valuable to identify depression, assess its severity, and monitor the treatment course in primary care and

M Hansson, MD, PhD student; J Chotai, MD, PhD, associate professor; O Bodlund, MD, PhD, associate professor, Division of Psychiatry, Department of Clinical Science, Umeå University, Umeå; A Nordstöm, PhD, health promotion officer, County Council of Västerbotten, Umeå, Sweden.

Address for correspondence

Maja Hansson, Division of Psychiatry, Department of Clinical Science, Umeå University, S-90185 Umeå, Sweden.
E-mail: maja.hansson@psychiat.umu.se

Submitted: 5 December 2008; Editor's response: 2 February 2009; final acceptance: 10 March 2009.

©British Journal of General Practice

This is the full-length article of an abridged version published in print. Cite this article as: *Br J Gen Pract* 2009; DOI: 10.3399/bjgp09X454070.

in hospital settings. These instruments can be filled in by the patient in the waiting room before seeing the doctor, and they are easy to evaluate.

A few previous studies, from Germany, Australia, and the UK, compared the HADS and PHQ-9.^{16–19} One of these studies showed slightly superior operating characteristics for the PHQ-9 scale,¹⁷ while the majority found the scales equally reliable. The aim of this study is to explore the psychometric properties of the HADS and PHQ-9 in a Swedish clinical population. Also, it aimed to assess the comparability between these two instruments as they are generally employed, and to investigate if modified cut-off scores can make them more similar in performance.

METHOD

Participants

Data were collected from October 2006 to June 2007 at five primary healthcare centres and five psychiatric outpatient clinics in the county of Västerbotten, Sweden. The study population comprised consecutive patients who visited the physician during a 2-week period, reporting symptoms of depression with or without concurrent anxiety. The inclusion criteria were age ≥ 18 years, Swedish-speaking, and the patient considering himself/herself to have depression. After informed consent, all patients filled in a questionnaire about age, sex, reason for visiting the physician, occupation, and the two self-rating scales HADS and PHQ-9, in conjunction with their visit to the physician. Patients with psychiatric or somatic comorbidity were not excluded. To compare age differences, patients were divided into three groups: young adults (18–30 years), middle-aged (31–64 years), and older adults (≥ 65 years).

Response rate was around 70%. The sample comprised 766 participants altogether. Participants with more than two missing values in any one of the two self-rating scales were considered as drop-outs, which left 737 patients for analysis (126 from primary healthcare centres and 611 from psychiatric outpatient clinics).

Hospital Anxiety and Depression Scale (HADS)

The HADS is a self-rating scale first described in 1983 by Zigmond and Snaith.²⁰ It contains two subscales measuring symptoms of depression (HADS-D) and anxiety (HADS-A) during the previous week. It includes seven statements on each disorder, and each response consists of a four-point rating scale (0 to 3); a higher score depicts a worse condition. For each subscale the total score is at most 21. A score of ≥ 11 is considered a clinically significant disorder, whereas a score between 8 and 10 suggests a mild disorder.²⁰

How this fits in

Identifying patients with depression can be a challenge, and self-rating scales such as the Hospital Anxiety and Depression Scale (HADS) and Patient Health Questionnaire (PHQ-9) are helpful complements. There are a few earlier studies comparing these two scales where differences were found in the severity cut-offs, which emphasised further investigation of their psychometric properties. In this study it was found that the scales are both reliable, with stable factor structure and high internal consistency. However, the scales' severity cut-offs showed differences, and the highest agreement was found in the PHQ-9 at a score ≥ 12 compared to HADS depression subscale ≥ 8 .

Several studies suggest a cut-off score of ≥ 8 to be optimal for best sensitivity and specificity.^{12,16,18} The HADS has sensitivity and specificity of about 80%, and a predictive validity for identification of about 70%. It has an excellent ability to detect cases among unselected patients in primary care, with area under the curve ranging from 0.84 to 0.96.¹²

The HADS takes approximately 5 minutes to fill in and it can quickly be evaluated by the physician. The scale has been widely used in many countries for screening of depression and/or anxiety disorder, and it has been validated in both hospitals and primary care settings.^{12,20,21} The HADS was originally designed to screen among patients with physical illness, and therefore it excludes items that measure somatic symptoms.²⁰ Since the HADS does not contain any item regarding suicidal thoughts, it focuses on relatively mild forms of the disorder.

Patient Health Questionnaire (PHQ-9)

The PHQ-9 is a self-rating instrument for depression that was developed in 1999 by Spitzer *et al*, from the PRIME-MD.²² It screens for the occurrence and severity of depression, derived from criteria of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV).⁹ Nine symptoms are described and the patient responds by indicating how much he/she has been bothered by these symptoms over the last 2 weeks. Also, there is a tenth question regarding the patient's functioning. As in the HADS, the response to each item is rated 0–3, which enables the clinician to assess the severity of the disorder. The maximum possible score is 27. Scores of 5–9 indicate mild depression, 10–14 moderate depression, 15–19 moderately severe depression, and ≥ 20 severe depression. Similar to the HADS, the PHQ-9 takes only a few minutes to complete.

Earlier studies in primary care show that PHQ-9 detects depression with a sensitivity around 90% and a specificity ranging from 77% to 88% when using the cut-off ≥ 10 .^{15,17,23,24} However, some studies recommend a higher cut-off such as ≥ 11 or ≥ 12 to improve the balance between sensitivity and

specificity.^{16,24} In the PHQ-9 the cut-offs for depression are not as well established as in the HADS-D, and there is still a 'grey zone' in the range of cut-off scores from 10 to 15.^{15,24,25}

The PHQ-9 has been validated in both primary care settings and hospitals. However, as it is a newer instrument, it has not been used or validated to the same extent as the HADS, and is not yet validated in a Swedish population.

Statistics

All statistical analyses were performed using SPSS (version 15.0). Mean differences were tested using the independent *t*-test. The level of statistical significance was set at $P < 0.05$, and all tests were two-sided. By means of Cohen's κ coefficient, pairwise agreement was tested. Cronbach's α and inter-item correlations were used to calculate the internal consistency of the HADS and its subscales and the PHQ-9. Exploratory factor analysis using principal component analysis (PCA) with varimax rotation was used to explore the factor structure of the two self-rating scales.

RESULTS

Sample characteristics

The age of participants ranged from 18 to 85 years, with a mean of 39 (standard deviation [SD] 15) years, and 72% were female. Among the participants, 47% were working or studying, 38% were on sick leave or had disability pension, 4% were retired, and 10% were unemployed. A few patients had psychotic symptoms ($n = 18$). All analyses were performed both with and without these patients, without changing the results.

Depression and anxiety scores of the population

The mean score on the HADS-D was 8.6 (SD 4.4) without significant differences between patients from primary healthcare centres or psychiatric outpatient clinics, or between males and females. Also the mean PHQ-9 scores were similar in both the primary healthcare centre and psychiatric outpatient clinic populations, with a mean score of 13.4 (SD 7.3). Females had a significantly higher mean score in the PHQ-9 than males: 13.9 versus 12.1 ($P = 0.003$). However, the difference between males and females scoring ≥ 10 on the PHQ-9 was not significant.

The mean HADS-D score and the frequency of patients with HADS-D ≥ 8 did not differ between the different age groups. The mean score in the PHQ-9 decreased with age, and the proportion of patients with PHQ-9 ≥ 10 was 70.5% among the youngest and 51.0% among the oldest group ($P < 0.05$).

According to the HADS, 42.6% of patients

experienced anxiety (HADS-A ≥ 11). Anxiety symptoms were more frequent among the females than the males: 45.0% versus 36.5% ($P = 0.035$).

Reliability

The internal consistency of the HADS-D subscale gave a Cronbach's α of 0.87. The Cronbach's α of the whole HADS was 0.90 and the subscale HADS-A had a Cronbach's α of 0.84. Cronbach's α on the PHQ-9 scale was 0.91.

Factor analysis

PCA with varimax rotation extracted two factors for the HADS, corresponding to an anxiety factor and a depression factor, with eigenvalues of 6.2 and 1.5 respectively. Each item of HADS loaded on only one of the factors, except for item number 7, 'I can sit at ease and feel relaxed', which loaded on both factors. These two factors explained 54.6% of the total variance. A factor analysis of the subscale HADS-D gave one factor with an eigenvalue of 3.9, which explained 56.2% of the total variance.

Factor analysis of the PHQ-9 resulted in a single factor with an eigenvalue of 5.1, explaining 56.5% of the total variance. A factor analysis including all items from both the PHQ-9 and HADS-D was also performed (Table 1). This yielded two factors with eigenvalues of 7.9 and 1.3 respectively, which together explained 57.3% of the total variance. The PHQ-9 items a and b (which include the two key symptoms of depression: anhedonia and depressed mood) loaded on both factors, whereas the HADS-D only loaded on the first factor.

Severity ratings

As seen in Table 2, the PHQ-9 scale diagnosed more than twice as many patients with severe depression in comparison with the HADS-D scale, while the two scales identified equal amounts of mild forms of depression. In total, the PHQ-9 scale identified 30% more patients as depressed compared to the HADS-D scale. However, this is when using PHQ-9 ≥ 5 as the cut-off for any depression.

Agreement at different cut-offs for depression

One-third (33.5%) of the patients had a current clinical depression defined by a HADS-D score ≥ 11 , while two-thirds (65.9%) had PHQ-9 score ≥ 10 ($\kappa = 0.35$). Any depression, defined by HADS-D score ≥ 8 was identified in 57.8% of patients. Compared to the cut-off scores in the PHQ-9, the best agreement ($\kappa = 0.56$) with HADS score ≥ 8 was PHQ-9 ≥ 12 , which also made the prevalence of depression found by the two scales similar (57.8% versus 56.2%). So, with a slightly higher cut-off for PHQ-9, the scales can be fairly equalised.

Table 1. Factor loadings with varimax rotation of the pooled items from HADS-D and PHQ-9.^a

Scale and item	Factor 1	Factor 2
Hospital Anxiety and Depression Scale – depression subscale (HADS-D) items		
2. I still enjoy the things I used to enjoy	0.737	0.275
4. I can laugh and see the funny side of things	0.817	0.210
6. I feel cheerful	0.747	0.293
8. I feel as I am slowed down	0.591	0.396
10. I have lost interest in my appearance	0.536	0.221
12. I look forward with enjoyment to things	0.824	0.218
14. I can enjoy a good book or radio or TV programme	0.500	0.355
Patient Health Questionnaire (PHQ-9) items		
a. Little interest or pleasure in doing things	0.625	0.530
b. Feeling down, depressed or hopeless	0.543	0.643
c. Trouble falling or staying asleep, or sleeping too much	0.161	0.700
d. Feeling tired or having little energy	0.302	0.729
e. Poor appetite or overeating	0.276	0.706
f. Feeling bad about yourself, or that you are a failure or have let yourself or your family down	0.362	0.661
g. Trouble concentrating on things, such as reading the newspaper or watching television	0.269	0.723
h. Moving or speaking so slowly that other people could have noticed. Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual	0.228	0.738
i. Thoughts that you would be better off dead or of hurting yourself in some way	0.376	0.502

^aAll items from both scales are analysed in the same factor analysis. Factor loadings >0.400 are in bold.

DISCUSSION

Summary of main findings

Results of the present study showed that the psychometrics of both the HADS and PHQ-9 were very good, with high internal consistency and a stable factor structure. Factor analysis performed on the pooled items from the two depression scales together gave two factors. Items from each scale loaded on separate factors, except for the first two items in the PHQ-9, which loaded on both factors. This can be explained by the fact that these two main criteria for depression are covered in both scales and that the remaining items of the PHQ-9 are not covered by the HADS. The PHQ-9, which is based on the criteria of the DSM-IV, measures a heterogeneous spectrum of symptoms in major depressive disorder, in contrast to HADS items which emphasise the emotional aspects of depression.

When the recommended severity scores are used, the PHQ-9 seemed to identify around 30% more patients with any type of depressive symptoms in comparison to the HADS-D (HADS-D ≥ 8 and PHQ-9 ≥ 5). A similar difference was seen when comparing the recommended cut-off scores for depression (HADS-D ≥ 11 and PHQ-9 ≥ 10). The cut-off score of ≥ 8 in the HADS-D, which is most frequently recommended,¹² had a higher agreement with PHQ-9 ≥ 10 ($\kappa = 0.52$). However, an even higher κ value of 0.56 was found when comparing HADS-D ≥ 8 with PHQ-9 ≥ 12 . The κ value was still moderate, which indicates that the scales are comparable but do not completely identify the same patients.

In this study, the HADS-D found an equal number of patients with depression both among males and females and in the different age groups. The PHQ-9 on the other hand showed a higher mean score among females, and the youngest patients had a significantly higher score compared to the oldest.

Strengths and limitations of the study

One of the strengths of this study is that this is the first time the Swedish version of the PHQ-9 has been analysed in relation to factor structure and agreement with the HAD-scale. This is of high value as the PHQ-9 has recently been recommended for routine use in primary care and in psychiatric outpatient clinics in (some parts of) Sweden. Comparative studies of diagnostic tools are of value to clinicians when choosing which one to use in clinical practice. This study also contains a fairly high number of participants with both psychiatric

Table 2. Distribution of all patients ($n = 737$) in the severity ratings of mild, moderate, and severe depression according to the HADS-D and PHQ-9.

HADS-D		PHQ-9	
Severity (score range)	n (%)	Severity (score range)	n (%)
Mild (8–10)	179 (24.3)	Mild (5–9)	161 (21.8)
Moderate (11–14)	173 (23.5)	Moderate (10–14)	159 (21.6)
–	–	Moderately severe (15–19)	151 (20.5)
Severe (15–21)	74 (10.0)	Severe (20–27)	176 (23.9)
Any depression (≥ 8)	426 (57.8)	Any depression (≥ 5)	647 (87.8)

HADS-D = Hospital Anxiety and Depression Scale – depression subscale. PHQ-9 = Patient Health Questionnaire.

outpatients and patients from general practice. It might be considered a limitation that the patients investigated were not all from the same kind of setting, but were from both psychiatric outpatient clinics and primary healthcare centres. However, the intention of the study was to evaluate the psychometrics of the self-rating scales, and the results did not differ between the different settings. The patients from primary care had the same mean score in the scales as the patients from psychiatric outpatient clinics.

A limitation of this study is that it does not compare the two self-rating scales in relation to a gold standard diagnostic measure; therefore, it is not possible to calculate sensitivity or specificity of the instruments, which would have been of interest. Although only patients seeking help for depressive symptoms were included, only about 60% of the sample were experiencing current depression according to the scales. This might be explained by the fact that many of the patients, especially those from psychiatric outpatient clinics, were probably already on treatment with antidepressants, and their symptoms might therefore be less pronounced. Unfortunately, this study did not include information on the patients' ongoing treatment. It can also be considered a weakness that only patients with depressive complaints were included. This meant that the probability of the patients being depressed would be higher than among unselected patients. However, this study did not aim to measure the case-finding abilities of the scales.

Comparison with existing literature

Previous studies on the psychometric properties of the HADS and PHQ-9 correspond well with the present results. Previous studies of the PHQ-9 that examined psychometric properties have shown that all items load in one single factor and that it has an internal consistency with Cronbach's α of 0.79 to 0.89.^{16,19,26} In the factor analysis of the HADS, item number 7 ('I can sit at ease and feel relaxed') loaded on both the 'depression' and the 'anxiety' factors. These findings are supported by several earlier studies.^{12,21,27} The HADS has shown good internal consistency in previous studies, with Cronbach's α around 0.82.¹²

In the present study, the PHQ-9 and HADS-D differed in the categorisation of depression severity, where the PHQ-9 found a considerably larger proportion of patients had more severe depression. The difference in severity cut-offs is also noted by Cameron *et al.*¹⁹

The κ value was low when comparing HADS-D ≥ 11 and PHQ-9 ≥ 10 . This might indicate that the cut-off of ≥ 11 in the HADS-D is too high, which is suggested

in other studies;^{12,21} alternatively, the PHQ-9 scale may be over-inclusive, as indicated by, for example, Löwe *et al.*¹⁶

Implications for future research and clinical practice

Future studies comparing the HADS and PHQ-9 to a gold standard are needed to analyse further their differences in cut-offs. It would be valuable to compare the PHQ-9 and HADS in a general population or among all patients attending a clinic or primary healthcare centre, to explore their screening abilities. It would also be interesting to compare the scales in a larger population with more information about the participants to determine if one of the scales is more suitable for a certain group of patients.

Depression can be difficult to diagnose, particularly because many patients with depression present themselves with somatic complaints and the time for the consultation is limited. In these settings, self-rating scales such as the HADS or PHQ-9 can be very helpful to screen for depression. The scales are also suitable for measuring the severity of the disorder, and if used repeatedly the treatment outcome can be monitored.^{21,28}

Advantages with the HADS are that the scale is widely used, validated in different populations, and translated into at least 30 different languages.²¹ The scale's ability to identify depression in patients with concurrent somatic illness might also be an advantage in primary care. Patients with somatic diseases might, for example, have sleeping disturbance, lack of energy, and difficulties in concentrating due to their somatic illness. The HADS also contains separate subscales for depression and anxiety. As there is usually a high comorbidity of depression and anxiety, it is a benefit that the HADS enables measurement of the two disorders separately.

The PHQ-9 takes into account all the demands of the DSM-IV; namely, the number of fulfilled criteria needed for diagnosis, the functional criteria, and a symptom duration of at least 2 weeks. As the scale is derived from the DSM-IV, it might have a greater appeal, at least in research.

Even though both scales are suitable as screening instruments and appropriate for measuring treatment outcome, neither can replace the clinical interview. A clinical assessment should be the natural consequence of a score on a self-rating scale indicating depression.

Funding body

This research was supported by grants from the Department of Psychiatry, University Hospital, Umeå and the Söderström-Königska Foundation (reference number 1938)

Ethics committee

The study only included anonymous data with the patients'

scores on the self-rating scales, age, sex, occupation, and reason for visiting the clinic. The researchers did not have any access to the patients' medical records, or any other personal data. No intervention was conducted in this study. On the basis of this, ethical approval was judged not to be necessary

Competing interests

The authors have stated that there are none

Acknowledgements

We want to thank the participating primary healthcare centres and especially Mariehems Healthcare Centre in Umeå for support and help in recruiting patients from primary care.

Discuss this article

Contribute and read comments about this article on the Discussion Forum: <http://www.rcgp.org.uk/bjgp-discuss>

REFERENCES

- Sartorius N, Ustun TB, Lecrubier Y, Wittchen HU. Depression comorbid with anxiety: results from the WHO study on psychological disorders in primary health care. *Br J Psychiatry* 1996; **168**(Suppl 30): 38–43.
- Wittchen HU, Hofler M, Meister W. Prevalence and recognition of depressive syndromes in German primary care settings: poorly recognized and treated? *Int Clin Psychopharmacol* 2001; **16**(3): 121–135.
- SBU (The Swedish Council on Technology Assessment in Health Care). *Treatment of depression — a systematic literature review* [in Swedish]. Report no 166. Stockholm: SBU; 2004.
- Ayuso-Mateos JL, Vázquez-Barquero JL, Dowrick C, *et al*. ODIN Group. Depressive disorders in Europe: prevalence figures from the ODIN study. *Br J Psychiatry* 2001; **179**: 308–316.
- Andrews G, Henderson S, Hall W. Prevalence, comorbidity, disability and service utilisation. Overview of the Australian National Mental Health Survey. *Br J Psychiatry* 2001; **178**: 145–153.
- Blazer DG, Kessler RC, McGonagle KA, Swartz MS. The prevalence and distribution of major depression in a national community sample: the National Comorbidity Survey. *Am J Psychiatry* 1994; **151**(7): 979–986.
- Allgulander C, Nilsson B. A nationwide study in primary health care: one out of four patients suffers from anxiety and depression [in Swedish]. *Lakartidningen* 2003; **100**(10): 832–838.
- Nordstrom A, Bodlund O. Every third patient in primary care suffers from depression, anxiety or alcohol problems. *Nord J Psychiatry* 2006; **62**: 250–255.
- Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA* 1999; **282**(18): 1737–1744.
- Cameron I, Reid I, Lawton K. Appropriateness of antidepressant prescribing: an observational study in a Scottish primary-care setting. *Br J Gen Pract* 2009; **59**(566): 644–649.
- Tylee A, Gandhi P. The importance of somatic symptoms in depression in primary care. *Prim Care Companion J Clin Psychiatry* 2005; **7**(4): 167–176.
- Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *J Psychosom Res* 2002; **52**(2): 69–77.
- Bodlund O. Anxiety and depression as hidden problem in primary health care. Only one case in four identified [in Swedish]. *Lakartidningen* 1997; **94**(49): 4612–4618.
- Hansson M, Bodlund O, Chotai J. Patient education and group counseling to improve the treatment of depression in primary care: a randomized controlled trial. *J Affect Disord* 2008; **105**(1–3): 235–240.
- Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001; **16**(9): 606–613.
- Löwe B, Spitzer RL, Gräfe K, *et al*. Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. *J Affect Disord* 2004; **78**(2): 131–140.
- Löwe B, Gräfe K, Zipfel S, *et al*. Diagnosing ICD-10 depressive episodes: superior criterion validity of the patient health questionnaire. *Psychother Psychosom* 2004; **73**(6): 386–390.
- Stafford L, Berk M, Jackson HJ. Validity of the Hospital Anxiety and Depression Scale and Patient Health Questionnaire-9 to screen for depression in patients with coronary artery disease. *Gen Hosp Psychiatry* 2007; **29**(5): 417–424.
- Cameron IM, Crawford JR, Lawton K, Reid IC. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. *Br J Gen Pract* 2008; **58**(546): 32–36.
- Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand* 1983; **67**(6): 361–370.
- Herrmann C. International experiences with the Hospital Anxiety and Depression Scale — a review of validation data and clinical results. *J Psychosom Res* 1997; **42**(1): 17–41.
- Spitzer RL, Williams JB, Kroenke K, *et al*. Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. *JAMA* 1994; **272**(22): 1749–1756.
- Wittkamp KA, Naeije L, Schene AH, *et al*. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. *Gen Hosp Psychiatry* 2007; **29**(5): 388–395.
- Gilbody S, Richards D, Barkham M. Diagnosing depression in primary care using self-completed instruments: UK validation of PHQ-9 and CORE-OM. *Br J Gen Pract* 2007; **57**(541): 650–652.
- Williams LS, Brizendine EJ, Plue L, *et al*. Performance of the PHQ-9 as a screening tool for depression after stroke. *Stroke* 2005; **36**(3): 635–638.
- Huang FY, Chung H, Kroenke K, *et al*. Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. *J Gen Intern Med* 2006; **21**(6): 547–52.
- Mykletun A, Stordal E, Dahl AA. Hospital Anxiety and Depression (HAD) scale: factor structure, item analyses and internal consistency in a large population. *Br J Psychiatry* 2001; **179**: 540–544.
- Löwe B, Unützer J, Callahan CM, *et al*. Monitoring depression treatment outcomes with the patient health questionnaire-9. *Med Care* 2004; **42**(12): 1194–1201.