doi:10.1093/fampra/cml052 Family Practice Advance Access published on 11 October 2006

Screening for depression in primary care

Pille Ööpik[a], Anu Aluoja[b], Ruth Kalda[a] and Heidi-Ingrid Maaroos[a]

Ööpik P, Aluoja A, Kalda R and Maaroos H. Screening for depression in primary care. *Family Practice* 2006; **23:** 693–698.

Introduction

Depression has become one of the most prevalent mental disorders. Depressive disorders or significant depressive symptoms have been found in up to 25% of patients visiting general or family practitioners.[1,2] Thus family doctors frequently meet depressive patients in their everyday work, and the importance of timely evidence-based diagnostics and treatment of the disorder has increased. The diagnosis of depression often presents difficulties, as it is time consuming and requires more clinical investigation.[3] The recognition of depression on the primary care level could be improved by the availability of relevant diagnostic devices. Structured psychiatric interviews, e.g. Composite International Diagnostic Interview (CIDI),[4,5] are reliable diagnostic instruments but these are too time consuming for family doctors and too elaborate for the routine use in primary care. Therefore, short self-rate questionnaires are proposed as screening devices of potential depressive patients in primary care. Self-rate instruments also enable the recognition of persons with minor depressive symptoms who form one of the risk groups for clinical depression. There is good evidence that screening improves the accurate identification of depressed patients in primary care settings.[6] Self-rate instruments vary in the number of symptoms, by duration of the symptoms and by the scale of evaluation. The combination of symptoms in the self-rate instruments is also variable and it is still not clear which combination is best for the differentiation of healthy persons from depressive ones. The most commonly used screening measures for adults in primary care settings include the Beck Depression Inventory,[7] the Zung Self-Depression Scale,[8] the General Health Questionnaire (GHQ)[9] and Patient Health Questionnaire-9 (PHQ-9).[10] Screening by asking two simple questions about the mood and anhedonia was studied in order to facilitate the recognition of depression and it was discovered that these questions were as effective as using longer instruments.[11] The aim of such type of studies was to provide as short and informative screening method of depression as possible which would be easier for patients and faster for GPs. In Estonia a screening scale for depression and anxiety, the Emotional State Questionnaire (EST-Q), which performs well on psychiatric patients and general population, has been developed.[12] Nowadays a later modified version EST-Q2 is used.

One purpose of the present study was to research the suitability of EST-Q2 screening scale depression subscale for screening of depression in general practice. The second purpose was to find out the combination of symptoms allowing the GPs to differentiate patients with depression from patients with other biomedical or psychosocial problems.

Methods

Sample

The recruitment of patients and the design of the research has been carried out according to the PREDICT project.[13] Consecutive attendees aged 18–75 were recruited from April to June 2003 by 23 family doctors (15 from urban and 8 from rural area) who had expressed interest in participating in the study. The principal language of the study was Estonian. The exclusion criteria were non-Estonian speakers, a severe organic mental illness and a terminal illness. After the participants had given their informed consent, a subsequent detailed interview was carried out either at their home or in the general practice within 2 weeks. Then patients filled the EST-Q2 themselves and interviewers administered the CIDI.

Received 25 May 2006; Revised 13 August 2006; Accepted 13 September 2006.
[a]Department of Family Medicine, University of Tartu and [b]Department of Psychiatry, University of Tartu, Tartu, Estonia. Correspondence to Pille Ööpik, Department of Family Medicine, University of Tartu, Puusepa 1a, 50406 Tartu, Estonia. Email: pille.oopik@ut.ee

The family doctors were specially instructed to recruit consecutive patients to avoid the bias for selecting depressive patients. They approached 1370 patients, 1175 patients agreed to take part in the study. Whole 1100 interviews were completed, as 75 patients could not be contacted or had changed their mind about participation. Later 42 interviews were excluded due to incomplete data. The final study sample consisted of 1058 persons, 776 (73%) women (mean age 40.5 ± 15.4) and 282 (37%) men (mean age 42.7 ± 16.2).

The Committee of Ethics of the University of Tartu has approved the study protocol and the informed consent form.

Psychometric instruments

CIDI. CIDI was selected for comparison because the reliability and validity of this instrument has already been established.[14,15] CIDI is a fully structured diagnostic interview providing current (and lifetime) psychiatric diagnoses according to ICD-10 and DSM-IV that was developed by the World Health Organization. A depressive episode was established using the Depression Section of the CIDI. In this study we used 1 month depression determined according to the criteria of the ICD-10.

Screening questionnaires—EST-Q2. A new modification of EST-Q was created in 2002. Items, which did not belong to any subscale, were omitted. EST-Q2 contained subscales of Depression, Anxiety, Agoraphobia-Panic, Fatigue and Insomnia, reflecting symptoms of depressive and anxiety disorders according to ICD-10 and DSM-IV. Each item was rated on a five-point scale ranging from 0 to 4 (0 = not at all; 1 = seldom; 2 = sometimes; 3 = often; 4 = all the time). The participants were asked to report how much the various problems had troubled them during the past 4 weeks, using the scale. The EST-Q2 version consisted of 28 items, the Depression subscale consisted of eight items encompassing cognitive and affective symptoms of depression. The cut-off point for depression was >11 (Appendix 1).

Statistical analysis

Two-by-two tables were constructed, displaying screening instrument (EST-Q2) diagnosis (positive/negative) by CIDI diagnosis (positive/negative). Sensitivity, specificity, false negative and positive rate, positive and negative predictive values were calculated to assess the ability of the screening instruments to render the diagnosis of depression by CIDI.

Furthermore, positive- and negative-likelihood ratios (LR) of the test were assessed.[16] The LR is a way to incorporate sensitivity and specificity of the test into a single measure. LRs tell us how much we should shift our suspicion in a particular test result. The positive LR (sensitivity/1–specificity) indicates how much we have to increase the probability of the disease if the test is positive. The negative LR (1–sensitivity/specificity) reflects how much we have to decrease the probability of the disease if the test is negative. Stepwise logistic regression was used to find the best combination of symptoms for screening depression. We constructed two regression models. In Model 1 we used EST-Q2 Depression scale symptoms as predictors of the CIDI-diagnosed depression. In Model 2 we added the EST-Q2 somatic and behavioural symptoms of depression, and one anxiety symptom *worrying too much* to predictor variables. According to our assessment the scale, in which all the arguments were statistically significant ($P < 0.05$), was the best. Different cut-off points were used to compare the scales. All the analyses were performed using the software package SAS 8.1.

Results

The CIDI interview diagnosed one month depressive episode in 162 (15%) participants, whereas 300 (28%) were screened depressive by the EST-Q2. The EST-Q2 classified 18.8% of the subjects differently when compared with the CIDI. Those 168 (15.9%) persons who did not have depression by the CIDI, but the EST-Q2 screened them as depressive, were classified as 'false positive'. Thirty (2.8%) persons who were diagnosed depressive by the CIDI, but were not depressive according to the EST-Q2, were 'false negative'.

Table 1 presents the results of sensitivity, specificity, false-negative rate, predictive values and LR different cut-off points for 1 month depression.

The EST-Q2 had good specificity, sensitivity, the positive predictive value and the positive LR for

TABLE 1 *Comparison of the test characteristics for the EST-Q2 at cut-off points >11, >10 and >12*

Screening instrument	Sensitivity	Specificity	FN	FP	PPV	NPV	Positive LR	Negative LR
EST-Q2>11	0.81	0.81	0.19	0.19	0.44	0.96	4.3	0.23
EST-Q2>10	0.86	0.77	0.13	0.23	0.4	0.97	3.7	0.18
EST-Q2>12	0.79	0.84	0.2	0.15	0.49	0.96	4.9	0.25

FN: false negative rate; FP: false positive rate; PPV: positive predictive value; NPV: negative predictive value.

the screening of depression at the present cut-off point >11. By decreasing the cut-off by one point, sensitivity and the false-negative rate improved, but the positive predictive value and the positive-LR got worse. By increasing the cut-off point by 1, sensitivity worsened but at the same time specificity, the positive predictive value and the positive-likelihood value improved (Table 1).

Out of the eight most characteristic symptoms, included in the EST-Q2 (Model 1) depression scale, feelings of sadness, loss of interest, self-accusations, loneliness and inability for enjoyment were the best identifiers of depressive patients (Table 2). In this combination worthlessness, suicidal ideation and hopelessness about the future did not discriminate depressive patients from non-depressive ones.

In the Model 2 feeling of sadness, feeling no interest or pleasure in things, feeling of worthlessness, impossibility to enjoy things, excessive worrying about several different things and rest does not restore the strength, were significantly related to having a depressive disorder (Table 2). Self-accusations, recurrent thoughts of death or suicide, feeling lonely, hopelessness about the future, feeling so restless that it is hard to sit still, fatigue or loss of energy, diminished ability to think or concentrate, being easily fatigued, difficulty in falling asleep, restless or disturbed sleep and waking up too early were not significant. We compiled two new scales from symptom combinations, which enabled us to distinguish between the depressive and non-depressive people. The symptoms of the first scale were feeling of sadness, feeling no interest, self-accusations, feeling lonely and no enjoyment (EST-Qnew1). The symptoms of the second scale were feeling of sadness, feeling no interest, worthlessness, no enjoyment, excessive worrying and rest does not restore strength (ESTQnew2). We examined the ability of screening of depression for both scales at different cutoff points in comparison with the original EST-Q2 depression scale (Table 3). At the cut-off point >8 EST-Qnew1 sensitivity was the same as in EST-Q2, but specificity improved from 0.81 to 0.82 and FP, positive predictive value (PPV) improved by order of 0.01 and positive LR improved by order of 0.02. At the cut-off point >7, EST-Qnew1 sensitivity improved by order of 0.06, but all other characteristics became worse. At the cut-off point >11 sensitivity of EST-Qnew2 did not change in comparison with EST-Q2, specificity improved from 0.81 to 0.85, FP, negative LR and PPV of 4.3–5.4, at the same time none of the characteristics became worse. In case of 50% of persons, who were screened depressive, the depressive disorder had also been diagnosed by the CIDI. By decreasing the cut-off point of EST-Qnew2 by one point, sensitivity improved from 0.81 to 0.88 and specificity became worse to 0.78. In comparison with EST-Q2 >10 (Table 1) and EST-Qnew1 >7 (Table 3) specificity became worse the least and sensitivity improved the most.

Discussion

Depressive disorders are common, yet often unrecognized in primary care.[17] Most psychiatric interviews are too elaborate for routine use in general practice.

TABLE 2 *Association between the EST-Q2 symptoms and CIDI-identified depressive episode: logistic regression model 1 and model 2*

Symptoms	Logistic regression model 1		Logistic regression model 2	
	Estimate	OR (95% CI)	Estimate	OR (95% CI)
Feeling of sadness	–0.46**	0.63 (0.46–0.86)	–0.50**	0.60 (0.45–0.82)
Loss of interest	–0.81**	0.45 (0.34–0.59)	–0.64**	0.52 (0.39–0.70)
Feeling of worthlessness	ns	–	–0.29**	0.75 (0.61–0.92)
Self-accusations	–0.33**	0.72 (0.58–0.90)	ns	–
Thoughts of suicide	ns	–	ns	–
Feeling lonely	–0.22*	0.80 (0.65–0.98)	ns	–
Hopelessness	ns	–	ns	–
Impossible to enjoy things	–0.35**	0.70 (0.56–0.89)	–0.26*	0.77 (0.60–0.97)
Excessive worry about several different things	–	–	–0.27*	0.76(0.60–0.97)
Feeling so restless that it is hard to sit still	–	–	ns	–
Fatigue or loss of energy	–	–	ns	–
Diminished ability to think or concentrate	–	–	ns	–
Rest does not restore strength	–	–	–0.35**	0.70 (0.57–0.87)
Being easily fatigued	–	–	ns	–
Difficulty in falling asleep	–	–	ns	–
Restless or disturbed sleep	–	–	ns	–
Waking up too early	–	–	ns	–

*$P < 0.05$; **$P < 0.01$.
ns: no significant.

TABLE 3 *Comparison of the test characteristics for the two new models and the EST-Q2 at fferent cut-off points*

Screening instruments and symptoms	Cutoff point	Sensitivity	Specificity	FN	FP	PPV	NPV	Positive LR	Negative LR
EST-Q2 Depression scale	>11	0.81	0.81	0.19	0.19	0.44	0.96	4.3	0.23
EST-Qnew1: Feeling of sadness	>8	0.81	0.82	0.19	0.18	0.45	0.96	4.5	0.23
Feeling no interest	>7	0.87	0.75	0.13	0.25	0.39	0.97	3.48	0.17
Self-accusations									
Feeling lonely									
Enjoyment									
EST-Qnew2: Feeling of sadness	>11	0.81	0.85	0.19	0.15	0.5	0.96	5.4	0.22
Feeling no interest	>10	0.88	0.78	0.12	0.22	0.43	0.97	4	0.15
Worthlessness									
Enjoyment									
Excessive worry									
Rest does not restore strength									

FN: false negative; FP: false positive rate; PPV: positive predictive value; NPV: negative predictive value.

Throughout the years numerous diagnostic scales and patient self-rating scales have been used. Different measures have considered different depression symptoms and the ideal combination of symptoms has not been found. The number of items in screening measures varies, and it is found that shorter screening measures may be as effective as using longer ones.[4]

The first result of the present study is that the EST-Q2 is applicable to primary care attendees. Sensitivity and specificity of the EST-Q2 depression subscale, using the cut-off point of >11, is good and comparable with other self-rate instruments.[18–20] Sensitivity of screening instruments is considered good when their range is 0.79–0.97 and specificity 0.63–0.86.[4] It was also found in our study. Although our results show that the EST-Q2 has sufficient sensitivity and specificity for using a screening instrument among primary care users testing of different cut-off scores was performed. When we lowered the cut-off point to >10, we received better sensitivity but the rate of false positives increased. Sensitivity can be further improved by lowering the cut-off point, but it yields too many false positives and decreases the positive predictive value. But if we consider the predictive value and the LR more important, a cut-off point of >12 can be used, which yields the highest positive predictive value.[16,20] The major shortcoming of screening instruments is the number of people screened false positive and false negative compared with those really suffering from depression. Some patients with 'false-positive' results on screening may have dysthymia or some anxiety disorder with concomitant depressive symptoms instead of major depression.[4] Screening instruments are not the means for diagnosing but can be the first step in identifying depressive disorders.

Secondly, on the basis of our questionnaire we identified the symptoms, which help the family doctors to discriminate between the patients with and without a depressive disorder, and as a result, formed two new screening scales. The efficacy of screening scales may depend on the symptoms included. The main symptoms of depression are lowered mood, loss of interest and no enjoyment and reduced energy accompanied by other symptoms like reduced concentration and attention, reduced self-esteem, feeling of guilt and worthlessness, pessimism about the future, suicidality, disturbed sleep and appetite. Most self-rate depression screening scales attempt to assess all the symptoms used in the diagnostic criteria. Nevertheless, the value of individual symptoms in screening of depression is not clear. First, we tried to identify which combination of affective and cognitive symptoms of EST-Q2 depression subscales discriminate best the CIDI-identified depression.[21] The typical symptoms of depression, *sad mood* and *loss of interest* occurred to be most significant. The best identifier of depressive episode was the *loss of interest*. That is supported by the study of screening depression with two questions, where the loss of interest yielded the least number of false positives and differentiated the depressive persons from the non-depressive ones best.[8] The other indicator of anhedonia, *impossibility to enjoy things*, was also significantly related to depression. This stresses the importance of anhedonia in recognising depression and supports the idea that while a high negative affect can be general to several negative mood states, lack of a positive affect is specific to depression.[22] From affective-cognitive symptoms of depression also self-accusations and loneliness were significantly related to CIDI-identified depression. When we formed a new self-rate scale with five items from symptoms significantly predicting depression (EST-Qnew1), it appeared to screen depression as well as the existing EST-Q2. This shows that decreasing the number of items in the questionnaire does not necessarily diminish its screening properties but a shorter version can be easier for the patient to fulfill, also, other studies show that shortened screening instruments may give better

results.[23] When we added somatic and behavioural self-rate symptoms to the model, the set of symptoms discriminating depression changed. In this combination in addition to *the sad mood, the loss of interest* and *impossibility of enjoyment*, also *worthlessness, worrying* and *rest does not restore the strength* became significant. Though excessive worrying is a typical symptom of generalized anxiety (GAD), our results suggest that it might be important in identifying primary care patients with depression. Recently the role of repetitive negative cognitions such as rumination and worry in maintaining mood and anxiety disorders has been highlighted.[24] Rumination, which is depression-specific pattern of thought, concerns the past loss or failure, whereas worrisome thoughts are characterized by the anticipated threat in the future. Although the content and the time frame of these cognitive phenomena differ, there might not be a clear distinction between the self-report of symptoms. This is supported by the other studies showing that the score of worrying is equally high in GAD and major depression, and pathological worry is strongly related to depression.[25,26] The result that a symptom of fatigue also differentiates depressive patients is interesting. Although reduced energy belongs to the core symptoms of depression, it is usually omitted from depression screening scales for primary care because it can be a sign of somatic illness.[27,28] Our study shows that even in primary care the patients' fatigue can be a significant identifier of a depressive disorder and should be included in self-report questionnaire as in PHQ-9.[10] Probably with a co-occurring somatic illness some specific aspect of fatigue, like *fatigue not being relieved after rest*, obtains significance as a characteristic of depression. The new scale EST-Qnew2, formed from the second combination of symptoms, yielded a better result in screening depression than EST-Q2 or EST-Qnew1. Specificity, FP, PPV and positive LR improved significantly whereas sensitivity and FN remained the same. The screening properties of EST-Qnew2 are equal or exceed those of common self-administered scales.[18] For instance, the PPV ranged from 0.43 to 0.5 for the EST-Qnew2 which is similar to a well-known primary care screening instrument PHQ-9 (PPV from 0.31 to 0.51 depending on the cut-off).[29]

In conclusion the screening instrument EST-Q2, created on the basis of population and psychiatric patients, is suitable among primary care users. Shortening of the questionnaire does not change its properties but makes its completion easier and quicker. The same approach should be useful for different self-report questionnaires, not only for EST-Q2. Our study approves the symptoms that help family doctors best to identify patients with depression are loss-of-interest, incapability for enjoyment, sad mood, worthlessness, excessive worry and strength not restored by resting. On the basis of these symptoms a new scale EST-Qnew2 was formed which can be used as a new screening instrument.

Acknowledgements

This work was supported by the Estonian Science Foundation grant No 5696. We are also grateful for support from the European Commission's Fifth Framework, grant number Predict-QL4-CT2002-00683 and we thank all PREDICT group members.

References

[1] Satcher D. Global mental health: its time has come. *JAMA* 2001; **285**: 1697.

[2] Meltzer H, Gill B, Petticrew M, Hinds K. OPCS Surveys of Psychiatric Morbidity in Great Britain. Report No 1. *The prevalence of psychiatric morbidity among adults aged* 16–64 living in private households in Great Britain. London: HMSO, 1995.

[3] Lecrubier Y. The Burden of depression and anxiety in general medicine. *J Clin Psychiatry* 2001; **62**: 3–12.

[4] Robins LN, Wing JK, Wittchen H-U *et al*. The composite international diagnostic interview. An epidemiologic instrument for use in conjunction with different diagnostic systems and in different cultures. *Arch Gen Psychiatry* 1988; **45**: 1069–1077.

[5] World Health Organization. *Composite International Diagnostic Instrument (CIDI).Version 2.1*. Geneva: WHO; 1997.

[6] US. Preventive Services Task Force. *Screening for depression: Recommendations and Rationale* [Clinical Guidelines]. *Ann Intern Med* 2002; **136**: 760–764.

[7] Beck AT, Steer RA, Brown GK. *BDI-II, Beck Depression inventory: manual*. 2nd ed. Boston: Harcourt, Brance and Company; 1996.

[8] Zung WWK. A self rating depression rating scale. *Arch Gen Psychiatry* 1965; **12**: 63–70.

[9] Goldberg DP, Williams P. *The user's guide to the general health questionnare*. Windsor: NFER-Nelson, 1988.

[10] Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatr Annals* 2002; **32**: 509–521.

[11] Arroll B, Khin N, Kerse N. Screening for depression in primary care with two verbally asked questions: cross sectional study. *BMJ* 2003; **327**: 1144–1146.

[12] Aluoja A, Shlik J, Vasar V, Luuk K, Leinsalu M. Development and psychometric properties of the Emotional State Questionnaire, a self-report questionnaire for depression and anxiety. *Nord J Psychiatry* 1999; **53**: 443–449.

[13] King M, Weich S, Torres F *et al*. Prediction of depression in European general practice attendees: The PREDICT study. *BMC Public Health* 2006; **6(1)**: 6.

[14] Witchen HU. Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): a critical review. *J Psychiatry Res* 1994; **28**: 57–84.

[15] Andrews G, Peters L. The psychometric properties of the Composite International Diagnostic Interview. *Soc Psychiatry Psychiatr Epidemiol* 1998; **33**: 80–88.

[16] Peters L, Andrews G. A procedural validity study of the computerized version of the Composite International Diagnostic Interview. *Psychol Med* 1995; **25**: 1269–1280.

[17] Katon W, Schulberg H. Epidemiology of depression in primary care. *Gen Hosp Psychiatry* 1992; **14**: 237–247.

[18] Henkel V, Mergel R, Kohnen R, Allgaier A-K, Möller H-J, Hegerl U. Use of brief depression screening tool in primary care: consideration of heterogeneity in different patient groups. *Gen Hosp Psychiatry* 2004; **26**: 190–198.

[19] Dutton GR, Grothe KB, Jones GN, Whitehead D, Kendra K, Brantley PJ. Use of the Beck Depression Inventory-II with

African American primary care patients. *Gen Hosp Psychiatry* 2004; **26**: 437–442.

[20] Arroll B, Goodyear-Smith F, Kerse N, Fishman T, Gunn J. Effect of the addition of a 'help' question to two screening questions on specificity for diagnosis of depression in general practice: diagnostic validity study. *BMJ* 2005; **331**: 884.

[21] World Health Organization. *The ICD-10 Classification of Mental and Behavioral Disorders: Clinical descriptions and diagnostic guidelines*. Geneva: WHO; 1992.

[22] Clark LA, Watson D. Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *J Abnorm Psychol* 1991; **100**: 316–336.

[23] Whooley MA, Avins AL, Miranda J, Browner WS. Case-finding instruments for depression. Two questions are as good as many. *J Gen Intern Med* 1997; **12**: 439–445.

[24] Papageorgiou C, Wells A. Nature, functions, and beliefs about depressive rumination. In Papageorgiou C, Wells A. (eds). *Depressive rumination: Nature, theory and treatment*. New York: Wiley; 2003: 3–20.

[25] Starcevic V. Pathological worry in major depression: a preliminary report. *Behav Res Ther* 1995; **33**: 55–56.

[26] Muris P, Roelofs J, Rassin E, Franken I, Mayer B. Mediating effects of rumination and worry on the links between neuroticism, anxiety and depression. *Pers Individ Dif* 2005; **39**: 1105–1111.

[27] Beck AT, Guth D, Steer RA, Ball R. Screening for major depression disorders in medical inpatients with the Beck Depression Inventory for Primary Care. *Behav Res Ther* 1997; **35**: 785–791.

[28] Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983; **67**: 361–370.

[29] Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depressionseverity measure. *J Gen Intern Med* 2001; **16**: 606–613.