

Primary care

Effect of the addition of a “help” question to two screening questions on specificity for diagnosis of depression in general practice: diagnostic validity study

B Arroll, F Goodyear-Smith, N Kerse, T Fishman, Jane Gunn

Abstract

Objective To determine the validity of two written screening questions for depression with the addition of a question inquiring if help is needed.

Design Cross sectional validation study.

Setting 19 general practitioners in six clinics in New Zealand.

Participants 1025 consecutive patients receiving no psychotropic drugs.

Main outcome measures Sensitivity, specificity, and likelihood ratios of the two screening questions, the help question, combinations of the screening and help questions, and diagnosis by general practitioners.

Results The help question alone had a sensitivity of 75% (95% confidence interval 60% to 85%) and a specificity of 94% (93% to 96%). The positive likelihood ratio for the help question was 13.0 (9.5 to 17.8) and the negative likelihood ratio was 0.27 (0.17 to 0.44). The likelihood ratio for patients wanting help today was 17.5 (11.8 to 31.9). The general practitioner diagnosis had a sensitivity of 79% (65% to 88%) and a specificity of 94% (92% to 95%).

Conclusion Adding a question inquiring if help is needed to the two screening questions for depression improves the specificity of a general practitioner diagnosis of depression.

Introduction

Depression is an important public health problem. Researchers estimate that by 2020 unipolar depression will be second only to ischaemic heart disease as the leading cause of disability adjusted life years.¹ Depression is common in general practice, with estimates ranging from 5.5% to 65.0% depending on the definition.² The suicide rate in depressed people is at least eight times higher than that of the general population.³ Most people who complete suicide have a mental disorder, and in 50% of cases depression is associated with the suicide.³ On a population basis the most important effect of major depression may be decreased quality of life and productivity rather than suicide. This effect is widespread and has been shown to be comparable to levels associated with major physical illnesses.^{4,5} Depressed patients often also present with a variety of physical symptoms, leading to excess use of medical services.⁶

Depending on how depression is defined, general practitioners tend to miss between 50% and 75% of cases.⁷ The reasons for this vary. General practitioners vary in competencies, skills, communication skills, knowledge base, duration of consultation, and attitudes about their patients, and about symptoms.^{8,9} Patients

who attend general practice also differ. Often, depressed patients present with somatic symptoms, including gastrointestinal, skeletal muscle, and cardiovascular symptoms, rather than describing non-somatic criteria for depression. In addition, patient factors such as poor insight into emotional illness add to the non-detection of depression.¹⁰ Many of the studies that assess detection rates by general practitioners use screening or detection tools that do not agree with each other, and therefore general practitioners may not agree with some or all of those tools.¹¹

A systematic review by UK authors concluded that screening for depression has little effect on patient outcomes.¹² The authors did not, however, pool their data, unlike the US Preventive Services Task Force.⁷ This group found that screening for depression can improve both detection and outcomes and therefore recommended its use in primary care.

The US group evaluated 41 screening studies and found that the two best tools (highest combination of sensitivity and specificity) were the patient health questionnaire¹³ and the Beck fast scan for primary care.¹⁴ The patient health questionnaire consists of nine questions and has been recommended for screening in general practice.^{15,16} The Beck fast scan for primary care consists of seven questions and includes a charge for use. The length of these two questionnaires and the costs incurred by the Beck tool makes a shorter questionnaire with no charges an attractive alternative.

A screening tool for depression using two questions (from the original prime-MD questionnaire)¹⁷ has been developed in written form.¹⁸ These two questions are “during the past month have you often been bothered by feeling down, depressed or hopeless?” and “during the past month have you often been bothered by little interest or pleasure in doing things?” These questions have a sensitivity of 96% and a specificity of 57% for depression in patients in whom substance misuse has been excluded.¹⁸ When these questions were asked verbally in an Auckland sample, the sensitivity was 96% and the specificity was 67%.¹⁹ The general practitioner diagnosis after patients had been asked the two questions had a sensitivity of 77%, a specificity of 86%, a positive likelihood ratio of 5.4, and a negative likelihood ratio of 0.27 (the positive predictive value was 27% and the negative predictive value 98.2%). We have since extended these two questions by adding a question that asks “is this something with which you would like help?” with three possible responses: “no,” “yes, but not today,” or “yes.” We validated the two questions plus



Information given on consent form and flow of participants are on [bmj.com](#)

Table 1 Sensitivity, specificity, and likelihood ratios of screening questions for depression in primary care, help question, combination of screening and help questions, and general practitioner diagnosis

Variable	% sensitivity (95% CI)*	True positive responses*	% specificity (95% CI)†	True negative responses†	Positive likelihood ratio (95%CI)	Negative likelihood ratio (95%CI)
Help question alone	75 (60 to 85)	35	94 (93 to 96)	838	13.0 (9.5 to 17.8)	0.27 (0.17 to 0.44)
Two screening questions alone	96 (86 to 99)	45	78 (76 to 81)	697	4.4 (3.9 to 5.1)	0.05 (0.01 to 0.21)
Either screening question plus help question	96 (86 to 99)	45	89 (87 to 91)	795	9.1 (7.4 to 11.1)	0.05 (0.012 to 0.19)
General practitioner diagnosis	79 (65 to 88)	37	94 (92 to 95)	835	13 (9.6 to 17.4)	0.23 (0.13 to 0.39)

Ideal method for all variables is composite international diagnostic interview.
 *Compared with 47 true positive responses in composite international diagnostic interview.
 †Compared with 889 true negative responses in composite international diagnostic interview.

the help question against the composite international diagnostic interview (mood module only).²⁰

Methods

We approached 19 general practitioners from six practices, all of whom agreed to participate in our study. Consecutive patients in the waiting room were invited to participate. Written informed consent was sought (see bmj.com). After consenting, the patients completed a written document, which included the two screening questions with a help question and a list of psychoactive drugs. We considered a response to either of the screening questions as a positive answer. Response to the help question was considered positive if patients responded by wanting help but not today or wanting help today. We also considered a response to be positive if the patient responded to either screening question plus the help question or to both screening questions plus the help question. The drug list included all available antidepressants, anti-anxiety agents, antipsychotics, and anticonvulsants. The patient then completed the mood module of the composite international diagnostic interview.²⁰ The research assistant did not look at the responses to the screening questions until the patient had completed the module. The patient showed the general practitioner his or her written responses to the screening and help questions. The general practitioners could ask any questions. They then completed a form with their opinion on whether the patient was depressed. Patients were not able to start treatment before completing the composite international diagnostic interview, which is considered the reference standard for detecting depression. This instrument takes the participants' answer—arrived at without any interpretation, probe, or explanation by the interviewer—as valid data for arriving at diagnoses. It has been shown to have excellent test characteristics in primary care with moderate to excellent ($\kappa = 0.58-0.97$) concordance with diagnoses in the international classification of disease, 10th revision.²⁰ It has the added advantage of being able to be administered by a non-clinical interviewer.

We calculated the sensitivity, specificity, and likelihood ratios according to the calculator on the University of Toronto website (www.cebm.utoronto.ca) for patients who were not currently taking psychoactive drugs. Our study was designed and analysed according to the STARD statement.

Results

We approached 1094 consecutive patients attending general practice. Overall, 1025 agreed to participate (94% response rate; see bmj.com).

Table 1 reports the measures of validity (sensitivity, specificity, likelihood ratios) for the questions answered. It also reports the

general practitioner diagnosis after seeing the patients' written response to the screening and help questions. The number of false positive responses to true positive responses for the two screening questions alone compared with either screening question plus the help question was 4.3 (192/45) versus 1.5 (54/37). Table 2 reports the likelihood ratios for a positive response to wanting help today, wanting help but not today, and not wanting help, all without the screening questions. When compared with the composite international diagnostic interview, the general practitioners had a sensitivity of 79% and a specificity of 94% for detecting major depression when using the two screening questions with the help question, giving a positive predictive value of 41% and a negative predictive value of 98.8%.

Discussion

The addition of a help question to the two screening questions from the Prime-MD questionnaire has a good sensitivity and an excellent specificity for a screening questionnaire for depression. The sensitivity of 79% for the general practitioner diagnosis of depression is an improvement over the 29-35% often reported.¹⁵ We previously found about five false positive responses for every true positive response when the two screening questions were asked verbally.¹⁹ In our present study this ratio changed from 4.3 to 1.5 when patients responded to either screening question plus the help question. This is much improved and provides a way around the traditional issue of large numbers of false positives in screening studies. Another way of looking at these results is that the likelihood ratio for asking for help today is 17.5, which is high and as such will significantly raise the post-test probabilities above the pretest value.²¹ In our study this means going from a 5.2% pretest probability of major depression to 48% if patients request help today in response to the help question. Asking a few more questions would confirm or refute the diagnosis of major depression. This likelihood ratio is better than that associated with the elevation of the ST segment in the diagnosis of myocardial infarction (likelihood ratio 11.20) and D-dimer levels above 1092 ng/ml for diagnosing deep vein thrombosis (3.1) although

Table 2 Likelihood ratio for answering help question with "yes, help today," "yes, but not today," and "no help," without consideration of two screening questions

Answer to help question	Diagnosis on composite international diagnostic interview*		Likelihood ratio (95% CI)
	Positive	Negative	
Help today	25	27	17.5 (11.1 to 27.7)
Help, but not today	10	24	7.9 (4.0 to 15.5)
No help	12	838	0.27 (0.17 to 0.44)

*Diagnosis on reference standard composite international diagnostic interview computer program.

not as good as venography for diagnosing deep vein thrombosis in patients with symptoms (47.5; see www.cebm.utoronto.ca). The validity measures of our screening tool for depression are therefore similar to those of physical diagnostic tests.

The strength of our study is that it was carried out in a community setting by general practitioners and in consecutive patients, excluding patients who were receiving psychotropic drugs. The patients were not attending general practice for any specific predetermined clinical reason. The response rate was high at 94% and it is the first validity assessment of the two questions administered with the help question. A weakness of our study is that we had no non-screened comparison group.

For studies of screening for depression in general practice the prevalence is usually reasonably low (5% for major depression in our study). The likelihood ratio for a negative test result does not therefore need to be low to rule out depression when the test result is negative; in our study a patient with a negative response to the help question would have a 1% chance of being depressed. Also, the two verbally asked questions had a similar likelihood ratio for a positive result when compared with the 41 screening studies for depression evaluated by the US Preventive Service Task Force.⁷ The best screening tool in that review was the Beck fast scan for primary care, with a positive likelihood ratio of 97 and a negative likelihood ratio of 0.03. Comparable values in our previous study were 4 and 0.17 for the Beck fast scan for primary care and 19.7 and 0.4 for the patient health questionnaire.¹⁹ Others have recommended using the patient health questionnaire to detect depression in primary care,¹³ but our two screening questions are shorter than the questionnaire, have similar likelihood ratios, and enable clinicians to pursue the issue of depression with the help question.

We suggest that these questions be presented to all new patients attending general practice and to patients who have not been to see their general practitioner for about two years. The intensity of administration would need to be decided by clinicians themselves. In our study, only one patient who had major depression did not respond positively to either of the two questions and the help question. Patients who responded to the help question with either help needed today or help needed, but not today had a 48% and 29% chance of having major depression, respectively. A positive response to either screening question plus the help question (table 1) signals a 32% chance of having major depression and a negative response signals a 99.7% chance of not having depression. Any of these three options therefore yields a high return. In practice any patient who answers yes to one or both of the screening questions or answers yes to the help question should be asked three or four more questions about depression, as the screening questions are almost identical to the first two questions of the *Diagnostic and Statistical Manual of Mental Disorder*, fourth edition, revised, for major depression (five symptoms are needed for a diagnosis of major depression).

Our explanation for the improvement in validity with the patient answering either screening question plus the help question is that it circumvents the many patients who respond to just one of the two screening questions and do not request help. Most of these responses are false positives and the help question seems to sort out those with major depression from those without. Patients who respond to both screening questions with or without the help question are another high risk group, therefore two out of three responses has a high validity.

We thank S Brighouse for her assistance with gathering data.

What is already known on this topic

High false positive responses are related to poor specificity in screening and diagnostic tests

Two screening questions have good sensitivity but poor specificity for major depression

General practitioner diagnosis with the two verbally asked questions has reasonable sensitivity and specificity for major depression

What this study adds

Response to two screening questions plus a question on whether help is wanted today or sometime have good sensitivity and specificity for major depression

General practitioner diagnosis with the two written screening questions plus the help question had similar sensitivity but improved specificity for major depression than without the help question

Contributors: BA, FG-S, NK, TF, and JG were involved in the design, interpretation of data, and drafting of the paper. BA analysed the data. He is guarantor.

Funding: Oakley Mental Health Foundation.

Competing interests: None declared.

Ethical approval: Auckland ethics committee.

- Murray CJ, Lopez AD. Alternative projections of mortality and disability by cause 1990-2020. *Lancet* 1997;349:498-504.
- Katon W, Schulberg H. Epidemiology of depression in primary care. *Gen Hosp Psychiatry* 1992;14:237-47.
- Monk M. Epidemiology of suicide. *Epidemiol Rev* 1987;9:51-8.
- Broadhead WE, Blazer DG, George LK, Tse CK. Depression, disability days and days lost from work in a prospective epidemiologic survey. *JAMA* 1990;264:2524-8.
- The Counselling Versus Antidepressants In Primary Care Study Group. How disabling is depression? Evidence from a primary care sample. *Br J Gen Pract* 1999;49:95-8.
- Waxman HM, McCreary C, Weinrit RM, Carner EA. A comparison of somatic complaints among depressed and non-depressed older persons. *Gerontologist* 1985;25:501-7.
- Agency for Healthcare Research and Quality. US Preventive Services Task Force <http://www.ahrq.gov/clinic/uspstfix.htm> (accessed 28 Mar 2005).
- Millar T, Goldberg DP. Link between the ability to detect and manage emotional disorders: a study of general practitioner trainees. *Br J Gen Pract* 1991;41:357-9.
- Whewell PJ, Gore VA, Leach C. Training general practitioners to improve their recognition of emotional disturbance in the consultation. *J Royal Coll Gen Pract* 1988;38:259-62.
- Good M, Good B, Cleary P. Do patients attitudes influence physician recognition of psychological problems in primary care. *J Fam Pract* 1987;25:53-9.
- The Mental Health General Practice Investigation Research Group. General practitioner recognition of mental illness in the absence of a 'gold standard.' *Aust NZ J Psychiatry* 2004;38:789-94.
- Gilbody SM, House A, Shledon TA. Routinely administered questionnaires for depression and anxiety: a systematic review. *BMJ* 2001;322:406-9.
- Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self report version of the prime-MD. *JAMA* 1999;282:1737-44.
- Steer RA, Cavalieri TA, Leonard DM, Beck AT. Use of the Beck depression inventory for primary care to screen for major depression disorders. *Gen Hosp Psychiatry* 1999;21:106-11.
- Nease DE, Malouin JM. Depression screening: a practical strategy. *J Fam Pract* 2003;52:118-26.
- Macarthur Foundation. The Macarthur initiative on depression and primary care. www.depression-primarycare.org (accessed 28 Mar 2005).
- Spitzer RL, Williams JB, Kroenke K, Linzer M, deGruy III FV, Hahn SR, et al. Utility of a new procedure for diagnosing mental disorders in primary care. The prime-MD1000 study. *JAMA* 1994;14:1749-56.
- 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;12:439-45.
- Arroll B, Khin N, Kerse N. Two verbally asked questions are simple and valid. *BMJ* 2003;327:1144-6.
- Jordanova V, Wickramasinghe C, Gerada C, Prince M. Validation of two survey diagnostic interviews among primary care attendees: a comparison of CIS-R and CIDI with SCAN ICD-10 diagnostic categories. *Psychol Med* 2004;34:1013-24.
- Guyatt G, Rennie DRE. *Users guide to the medical literature*. Chicago: AMA Press, 2002. (Accepted 23 August 2005)

doi 10.1136/bmj.38607.464537.7C

Primary care

Department of General Practice and Primary Health Care, School of Population
Health, University of Auckland, Private Bag 92019, Auckland, New Zealand
B Arroll *professor*
F Goodyear-Smith *senior lecturer*
N Kerse *associate professor*

T Fishman *senior lecturer*
Department of General Practice University of Melbourne, Australia
J Gunn *associate professor*
Correspondence to: B Arroll b.arroll@auckland.ac.nz