

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

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
Jane Gunn
associate professor

Correspondence to:
B Arroll
b.arroll@
auckland.ac.nz

BMJ 2005;331:884-6

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 (11.1 to 27.7). 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 common in general practice, ranging from 5.5% to 65.0% depending on the definition.¹ The suicide rate in depressed people is at least eight times higher than that of the general population.² 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.^{3,4} Depressed patients often also present with a variety of physical symptoms, leading to excess use of medical services.⁵


Depending on how depression is defined, general practitioners tend to miss between 50% and 75% of cases.⁶ The reasons for this include variations in general practitioners' competencies^{7,8} and the diversity of patients seen in general practice. Often, depressed

patients present with somatic symptoms rather than describing non-somatic criteria for depression. Also, patient factors such as poor insight into emotional illness add to the non-detection of depression.⁹ 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 them.¹⁰

A systematic review by UK authors concluded that screening for depression has little effect on patient outcomes,¹¹ whereas the US Preventive Services Task Force found that it improved both detection and outcomes.⁶

The US group evaluated 41 screening studies and found that the two best tools were the patient health questionnaire¹² and the Beck fast scan for primary care.¹³ The patient health questionnaire has been recommended for screening in general practice.^{14,15} The Beck fast scan for primary care includes a charge for use.

A screening tool for depression using two questions (“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?”)¹⁶ has been developed in written form.¹⁷ These questions have a sensitivity of 96% and a specificity of 57% for depression in patients in whom substance misuse has been excluded.¹⁷ When these questions were asked verbally in an Auckland sample, the sensitivity was 96% and the specificity was 67%.¹⁸ The general practitioner diagnosis after patients had been asked the two questions had a sensitivity of 77% and a specificity of 86%. 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 the help question against the composite international diagnostic interview (mood module only).¹⁹

 Information given on consent form and flow of participants are on bmj.com


 This is the abridged version of an article that was posted on bmj.com on 15 September 2005: <http://bmj.com/cgi/doi/10.1136/bmj.38607.464537.7C>

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.

Methods

Nineteen general practitioners from six practices all agreed to participate in our study. Consecutive patients in the waiting room were invited to participate. After providing written informed consent, the patients completed a 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 patient then completed the mood module of the composite international diagnostic interview.¹⁹ The patient showed the general practitioner his or her written responses. 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 ideal method for detecting depression.

We calculated the sensitivity, specificity, and likelihood ratios for patients who were not currently taking psychoactive drugs.

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 and the general practitioner diagnosis after seeing the patients' written responses to the 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 the help question, 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 two screening questions has good sensitivity and excellent specificity for detecting depression. The sensitivity of 79% for the general practitioner diagnosis of depression is an improvement over the 29-35% often reported.¹⁴ We previously found about five false positive responses for every true positive response when the two screening questions were asked verbally.¹⁸ In our present study this ratio changed from 4.3 to 1.5 when patients responded to either screening question plus the help question. 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.²⁰ In our study this means going from a 5.2% pretest probability of major depression to 48% if patients request help today. Asking a few more questions would confirm or refute the diagnosis of major depression. The validity measures of our screening tool for depression are similar to those of some physical diagnostic tests (see *bmj.com*).

The strength of our study is that it was carried out in a community setting by general practitioners and in consecutive patients, excluding patients taking psychotropic drugs. 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

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.

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

evaluated by the US Preventive Service Task Force.⁶ The best screening tool in that review was the Beck fast scan for primary care. Others have recommended using the patient health questionnaire to detect depression in primary care,¹² 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. 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.

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 Brighthouse for her assistance with gathering data.

Contributors: See bmj.com

Funding: Oakley Mental Health Foundation.

Competing interests: None declared.

Ethical approval: Auckland ethics committee.

- 1 Katon W, Schulberg H. Epidemiology of depression in primary care. *Gen Hosp Psychiatry* 1992;14:237-47.
- 2 Monk M. Epidemiology of suicide. *Epidemiol Rev* 1987;9:51-8.
- 3 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.
- 4 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.
- 5 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.
- 6 Agency for Healthcare Research and Quality. US Preventive Services Task Force <http://www.ahrq.gov/clinic/uspstfix.htm> (accessed 28 Mar 2005).
- 7 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.
- 8 Whewell PJ, Gore VA, Leach C. Training general practitioners to improve their recognitions of emotional disturbance in the consultation. *J Royal Coll Gen Pract* 1988;38:259-62.
- 9 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.
- 10 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.
- 11 Gilbody SM, House A, Shledon TA. Routinely administered questionnaires for depression and anxiety: a systematic review. *BMJ* 2001;322:406-9.
- 12 Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self report version of the prime-MD. *JAMA* 1999;282:1737-44.
- 13 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.
- 14 Nease DE, Malouin JM. Depression screening: a practical strategy. *J Fam Pract* 2003;52:118-26.
- 15 Macarthur Foundation. The Macarthur initiative on depression and primary care. www.depression-primarycare.org (accessed 28 Mar 2005).
- 16 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;271:1749-56.
- 17 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.
- 18 Arroll B, Khin N, Kerse N. Two verbally asked questions are simple and valid. *BMJ* 2003;327:1144-6.
- 19 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.
- 20 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

*Endpiece***We are mad**

We are mad, not only individually, but nationally. We check manslaughter and isolated murders; but what of war and the much-vaunted crimes of slaughtering whole peoples? There are no limits to our greed, none to our cruelty. And as long as such crimes are committed by stealth and by individuals, they are less harmful and less portentous; but cruelties are practised in accordance with acts of senate and popular assembly, and the public is bidden to do that which is forbidden to the individual. Deeds that would be punished by loss of life when committed in secret, are praised by us because uniformed generals have carried them put.

Seneca (5 BC-AD 65), Epistles 95:30-1

Jeremy Hugh Baron, honorary professorial lecturer, Mount Sinai School of Medicine, New York