Identifying depression in primary care: a comparison of different methods in a prospective cohort study

Verena Henkel, Roland Mergl, Ralf Kohnen, Wolfgang Maier, Hans-Jürgen Möller, Ulrich Hegerl

Department of Psychiatry, Ludwig-Maximilians-University Munich, Nußbaumstr 7, D-80336 Munich, Germany
Verena Henkel *psychiatrist*
Roland Mergl *psychologist*
Hans-Jürgen Möller *professor*
Ulrich Hegerl *professor*

Institute for Medical Research Management and Biometrics (IMEREM), Scheurlstr 21, D-90478 Nuremberg, Germany
Ralf Kohnen *professor*

Department of Psychiatry, University of Bonn, Sigmund-Freud-Str 25, D-53105 Bonn, Germany
Wolfgang Maier *professor*

Correspondence to: V Henkel verena.henkel@psy.med.uni-muenchen.de

BMJ 2003;326:200–1

Depressive disorders are a major health problem in primary care, and at least half of these disorders remain undetected.[1] There are two recommended approaches to diagnosing depression in primary care: one is to perform routine screening, and the other is to evaluate patients only when the clinical presentation triggers the suspicion of depression. Our aim was to compare these two approaches, and to compare three different screening tools in order to evaluate which would be most appropriate for use in primary care. From among the many available screening tools, we selected three brief, self rating instruments: one disorder-specific (the depression module of the brief patient health questionnaire (B-PHQ, 9 items)),[2] one broad based (the general health questionnaire (GHQ-12, 12 items)),[3] and one that is less restricted to both issues (WHO-5 wellbeing index (WHO-5, 5 items)).[4]

Methods and results

Eighteen primary care facilities participated in our prospective cohort study. The study protocol was approved by our local ethics committee. On one given day, all patients who presented in one of the practices were asked to complete the three screening questionnaires before seeing a doctor. The doctors who treated the patients remained blind to the questionnaire results until they had completed a brief "physician's encounter form" to indicate their clinical assessment of their patient's current diagnoses.

Within a period not exceeding six days after they had completed the questionnaires, the patients were contacted by telephone for a fully structured, standardised psychiatric interview (composite international diagnostic interview (CIDI)) conducted by a trained psychologist blind to the screening results. We chose the composite international diagnostic interview as the reference standard because its reliability and validity have been established.[5] The interviewing psychologists met a high standard of inter-rater reliability.

The main outcome measures were, firstly, the family doctors' performance in detecting depression without any tool to help guide diagnosis decisions and, secondly, the test accuracy of the screening questionnaires. We calculated sensitivity, specificity, and predictive values using two-by-two tables. We used two statistical tests to compare differences of characteristics of test accuracy (table).

For 431 patients, all screening questionnaires, the composite international diagnostic interview, and the physician's encounter form were completed. Of these patients, 17% suffered from any depressive disorder and 83% did not.

Comment

The sensitivity of the family doctors' unaided clinical diagnoses was 65%. With standard cut-off points, the briefest screening questionnaire (and therefore the most practical to use), the WHO-5, produced significantly greater sensitivity (93%) and a better negative predictive value (98%) than the other questionnaires (see table). However, the brief patient health questionnaire and unaided clinical diagnosis produced better specificity. The brief patient health questionnaire also produced the best positive predictive value. However, since screening tools are designed to identify all patients at risk for a disorder, sensitivity and negative predictive value are the most important operating characteristics.

Comparison of test accuracy of screening questionnaires for depression and family doctors' unaided clinical diagnosis. Values are means (95% confidence intervals) unless stated otherwise

Measures of test accuracy	Screening questionnaires			Unaided clinical diagnosis (UCD)	Significant differences ($P \leq 0.05$, one sided tests)§
	WHO-5*	GHQ-12†	B-PHQ‡		
Sensitivity (%)	93 (85 to 98)	85 (74 to 92)	78 (66 to 87)	65 (53 to 76)	WHO-5>GHQ-12, B-PHQ>UCD
Negative predictive value (%)	98 (95 to 99)	95 (92 to 98)	95 (92 to 97)	91 (88 to 94)	WHO-5>B-PHQ>UCD, GHQ-12>UCD
Specificity (%)	64 (59 to 69)	62 (57 to 67)	85 (81 to 89)	74 (69 to 79)	B-PHQ>UCD>WHO-5, UCD>GHQ-12
Positive predictive value (%)	34 (28 to 41)	31 (25 to 38)	51 (42 to 61)	34 (26 to 42)	B-PHQ>WHO-5>GHQ-12, B-PHQ>UCD

*WHO-5 wellbeing index (scoring procedure as indicated in *World Health Organization info package*[4]).
†General health questionnaire (scoring procedure as indicated in Goldberg 1978[3]).
‡Brief patient health questionnaire (scoring procedure as indicated in Spitzer et al 1999[2]).
§McNemar's test to compare sensitivities and specificities, analogue of McNemar's test to compare predictive values.

Our results suggest that the use of WHO-5 could improve family doctors' ability to detect depression, supporting the World Health Organization's recommendation that every patient in primary care should participate in a screening process with the completion of WHO-5 as a standard first step, done in the waiting room.[4] The questionnaire can easily be scored by hand. Patients who score positively for depression should be examined by their doctor in order to confirm a diagnosis of depression or to rule out normal distress or physical causes of depression. At this stage, doctors could use the brief patient health questionnaire as a checklist.

We hope that our results favouring such a simple, two stage screening process for depression in primary care, starting with the questionnaire WHO-5, will encourage further research in other countries.

We thank Simone Braun, Kathrin Allgaier, Petra Ohlendorf, Isabelle Seidscheck, and Evelyn Poth for data collection. We thank Jan Stefanek and Simone Braun for conducting the ROC-analyses presented in an earlier draft of this paper.

Contributors: VH had the idea for this paper and drafted the paper. RM analysed the data. RK, WM, H-JM, and UH commented on the study protocol and the text of the paper. UH is the speaker of the "German Research Network on Depression." VH and UH are guarantors for the study.

Funding: The study was funded by grants from the German Federal Research Ministry within the programme "German Research Network on Depression" and by additional funds from Pfizer and Novartis.

Competing interests: None declared.

1 Paykel ES, Tylee A, Wright A, Priest RG, Rix S, Hart D. The defeat depression campaign: psychiatry in the public arena. *Am J Psychiatry* 1997;154(6 suppl):59-65.
2 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;282:1737-44.
3 Goldberg DG. *Manual of the general health questionnaire.* Windsor: NFER Publishing, 1978.
4 *World Health Organization info package: Mastering depression in primary care.* Frederiksborg: World Health Organization, Regional Office for Europe, Psychiatric Research Unit, 1998.
5 Andrews G, Peters L. The psychometric properties of the composite international diagnostic interview. *Soc Psychiatry Psychiatr Epidemiol* 1998;33:80-8.

(Accepted 15 August 2002)