# Routinely administered questionnaires for depression and anxiety: systematic review

Simon M Gilbody, Allan O House, Trevor A Sheldon

NHS Centre for Reviews and Dissemination, University of York YO10 5DD
Simon M Gilbody *MRC fellow in health services research*
Academic Unit of Psychiatry and Behavioural Sciences, University of Leeds LS2 9LT
Allan O House *professor liaison psychiatry*
Department of Health Studies, University of York
Trevor A Sheldon *head*
Correspondence to: S M Gilbody Academic Unit of Psychiatry and Behavioural Sciences, University of Leeds LS2 9LT s.m.gilbody@leeds.ac.uk

*BMJ* 2001;322:406–9

## Abstract

**Objectives** To examine the effect of routinely administered psychiatric questionnaires on the recognition, management, and outcome of psychiatric disorders in non-psychiatric settings.
**Data sources** Embase, Medline, PsycLIT, Cinahl, Cochrane Controlled Trials Register, and hand searches of key journals.
**Methods** A systematic review of randomised controlled trials of the administration and routine feedback of psychiatric screening and outcome questionnaires to clinicians in non-psychiatric settings. Narrative overview of key design features and end points, together with a random effects quantitative synthesis of comparable studies.
**Main outcome measures** Recognition of psychiatric disorders after feedback of questionnaire results; interventions for psychiatric disorders; and outcome of psychiatric disorders.
**Results** Nine randomised studies were identified that examined the use of common psychiatric instruments in primary care and general hospital settings. Studies compared the effect of the administration of these instruments followed by the feedback of the results to clinicians, with administration with no feedback. Meta-analytic pooling was possible for four of these studies (2457 participants), which measured the effect of feedback on the recognition of depressive disorders. Routine administration and feedback of scores for all patients (irrespective of score) did not increase the overall rate of recognition of mental disorders such as anxiety and depression (relative risk of detection of depression by clinician after feedback 0.95, 95% confidence interval 0.83 to 1.09). Two studies showed that routine administration followed by selective feedback for only high scorers increased the rate of recognition of depression (relative risk of detection of depression after feedback 2.64, 1.62 to 4.31). This increased recognition, however, did not translate into an increased rate of intervention. Overall, studies of routine administration of psychiatric measures did not show an effect on patient outcome.
**Conclusions** The routine measurement of outcome is a costly exercise. Little evidence shows that it is of benefit in improving psychosocial outcomes of those with psychiatric disorder managed in non-psychiatric settings.

## Introduction

Disorders such as anxiety and depression are especially prevalent in primary care and general hospital settings and yet often go unrecognised.[1 2] Psychiatric screening and outcome questionnaires have been advocated as an aid to the detection of cases and clinical decision making.[3] Self completed instruments such as the general health questionnaire are acceptable to patients, have adequate sensitivity and specificity in their ability to identify disorders such as anxiety and depression, and are sensitive to change.[4] The routine use of these instruments might, therefore, be a simple and cost effective means of improving the recognition, management, and outcome of psychiatric disorders in non-psychiatric settings.

If psychiatric questionnaires are to be of value, however, clinicians must routinely use them and act on their results. In short, questionnaires must change professional behaviour such that psychiatric disorders are more readily recognised and better managed, thereby having an improved outcome. Otherwise their implementation is a cumbersome, costly, and bureaucratic exercise. We systematically reviewed the evidence on the use of routinely administered psychiatric questionnaires in non-psychiatric settings.

## Methods

### Search strategy

We searched Medline (1966-2000), Embase (1981-2000), Cinahl (1982-2000), PsycLIT (to 2000), and the Cochrane Controlled Trials Register (to 2000). We also hand searched several key journals and scrutinised reference lists for additional studies (see website).

### Inclusion criteria

Participants were those being treated in non-psychiatric settings. Included studies were those in which the intervention involved the use of any standardised measure of psychiatric symptoms as a screening and outcome assessment instrument in routine care, with results being fed back to clinicians. The control intervention involved routine care, with results not being fed back to clinicians.

### Outcomes

We sought outcome data on rates of detection of psychiatric disorders, initiation of treatment or referral for psychiatric disorders, the outcome of psychiatric disorders, consulting behaviour and service use, patient satisfaction with care and patient-doctor communication, and cost (direct and indirect)

### Data extraction and validity assessment

Study inclusion, quality assessment, and data extraction were conducted by two reviewers, and differences were resolved by discussion. Study quality, particularly the method of randomisation, was judged with accepted criteria.[5] Additionally, we established the unit of randomisation—whether by individual patient or by cluster.[6]

### Data synthesis

Individual studies are reported separately. Where appropriate, results from different studies were pooled using a random effects model.[7] Where incomplete data were reported, we attempted to contact the first author.

bmj.com

*Details of the search terms and studies appear on the BMJ's website*



BMJ: first published as 10.1136/bmj.322.7283.406 on 17 February 2001. Downloaded from http://www.bmj.com/ on 18 March 2024 by guest. Protected by copyright.

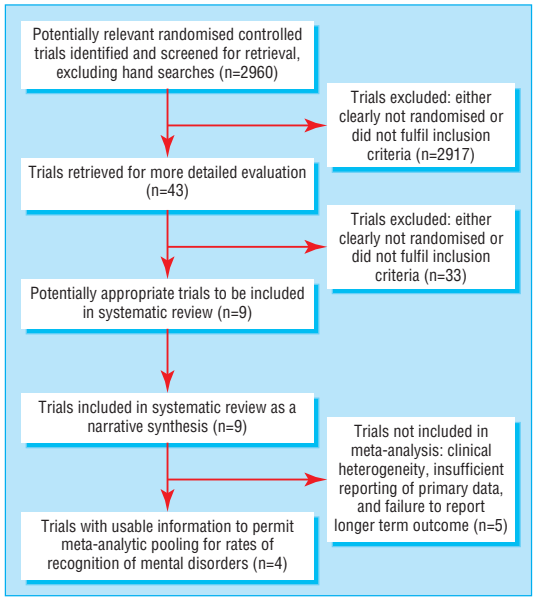



**Fig 1** Flow of participants of studies through review process

Relative and absolute risks are reported for dichotomous outcomes.

## Results

We identified nine studies conducted in non-psychiatric settings (fig 1). Six studies were in the setting of primary care[8–13] and three in general medical outpatients (see table on website).[14–16]

### Study design and quality

The method of randomisation was rarely described. One study used a pseudorandomised design.[8] In most studies the unit of randomisation was the patient, with individual clinicians receiving questionnaire results for some patients and not for controls—raising the problem of cross contamination and sensitisation between participants.

Questionnaires used included the Beck depression inventory,[17] the general health questionnaire (versions 12 and 28),[4] and the Zung self rated depression scale.[18] One study[11] combined an anxiety questionnaire (anxiety scores from symptom check list 90)[19] with a health status questionnaire (the short form 36).[20] Instruments were administered by research assistants before consultation.

The interventions involved the feedback of the test results to the clinician—generally as a sheet containing summary scores and an explanation of the importance of high scores indicating a possible psychological disorder. In most studies instruments were administered once and were used as instruments for "case finding" for the purposes of identifying problems at an assessment interview. In only one study was the outcome battery administered and results fed back sequentially during the course of care.[11]

Broadly, two types of participants were randomised: all patients, irrespective of their score on the instrument or likelihood of having pre-existing psychiatric disorder ("unselected patients"), and those with a probable psychiatric disorder, with a score above some cut-off point or a positive diagnostic interview ("high risk patients").

### Effects of routine screening and outcome measurement

*Recognition of emotional problems and minor psychiatric disorders*

The earliest study showed a large effect for the detection of depression through feedback of results from the general health questionnaire, increasing detection of depression in unselected patients seen by a single doctor (the study author) by 11%.[8] More methodologically robust studies, however, showed no overall effect of feedback for unselected patients.[10 15] Statistical pooling of studies that used feedback for all patients did not show an effect (DerSimonian-Laird pooled relative risk of detection of depression 0.95, 95% confidence interval 0.83 to 1.09; fig 2).[7] Insufficient data were presented in the earlier (and most positive) study to confirm the size of the result reported by the authors.[8] One study, comprising six arms, presented several different variations of time and mode of feedback, and pooling of the separate arms was not justified.[14] The inclusion of this trial did not, however, materially alter our results.

Three studies used a "high risk" approach, targeting feedback at a selected population of patients with a probable diagnosis of depression (Zung score greater than 50, Beck depression inventory score less than 14 or positive diagnostic interview schedule).[9 12 16] Pooling two studies that reported the detection of depression at the key index consultation showed that feedback increased the rate of recognition of depression by 27% (95% confidence interval 14% to 40%, DerSimonian-Laird pooled relative risk of detection of depression 2.64, 1.62 to 4.31; see fig 2).[9 7 16] Dowrick and Buchan reported diagnoses of depression from case notes at six and 12 months after feedback and found no overall effect (relative risk of detection of depression: six months, 0.82, 0.32 to 2.07; 12 months 1.71, 0.93 to 3.14).[12] Similarly, Magruder-Habib et al showed that the benefit of screening had diminished by 12 months and was of borderline significance (relative risk 1.63, 1.00 to 2.58).[16]

One study specifically employed the routine measurement of outcome in addition to the active education of clinicians into the nature and management of untreated anxiety.[11] This combined approach increased the rate of recognition of anxiety disorders (defined as "chart notations") from 19% to 32% in the intervention arm (relative risk of recognition 1.72, 1.25 to 2.37).

*Initiation of treatment for emotional problems*

Six studies investigated the effect of the feedback of questionnaire results on the rate of intervention for emotional problems.[11–16] All but one found no effect.[16] Heterogeneity of methods and definition of an active intervention meant that overall pooling was not justified.

The study that specifically targeted the recognition and intervention for anxiety showed increased mental health referrals (10% *v* 3%, relative risk of outside referral 2.94, 1.33 to 6.51).[11]

*Subsequent outcome of emotional disorders*

Surprisingly few studies examined the effect of routine measurement on the actual outcome of the patient over time. The earliest study, using retrospective patient recall, reported that patients with unrecognised depression, on whom feedback was given, had a



BMJ: first published as 10.1136/bmj.322.7283.406 on 17 February 2001. Downloaded from http://www.bmj.com/ on 18 March 2024 by guest. Protected by copyright.

shorter illness (2.8 months *v* 5.3 months).[8] However, general health questionnaire scores at 12 months were similar for depressed patients on whom feedback was given compared with controls.

No overall effect of feedback on longer term outcome was detected in two other studies.[12 13] These showed that unrecognised depressive symptoms resolved over a six to 12 month period, irrespective of detection and feedback.

The combination of an intensive educational and feedback intervention targeted at anxiety problems did not improve anxiety scores either on the symptom check list 90 or the mental health component of the short form 36.[11]

*Consulting behaviour*

Two studies examined the effect of feedback of outcome data on subsequent number of consultations with a doctor over six to 12 months and found no increase.[8 13] Feedback did, however, increase the proportion of consultations labelled as "psychiatric" by the doctor.

*Other outcomes*

No study reported the costs of routine measurement of outcome or clinicians' and patients' perceptions of their usefulness or acceptability.

## Discussion

Routine administration of validated outcome measures has not been shown to influence clinicians' behaviour. The recognition of emotional disorders seems to be increased only when there is some form of screening procedure, whereby an instrument is administered, scored by someone other than the clinician, and the results of those with high scores only fed back to the clinician.[16] Increased recognition does not, however, necessarily translate into improved management of depression or improved outcome.

There are several explanations for the lack of effect in unselected patients. The first relates to the psychometric properties of questionnaires and clinicians' perception of their value. What is of most interest to clinicians in the context of routine care is predictive value—that is, the proportion of those predicted by the test as having the disease who turn out to have the disease—not sensitivity and specificity.[21] Crucially, positive predictive value increases according to the prevalence of a disorder. Whereas unrecognised emotional disorders form a major portion of the clinical caseload in non-psychiatric services, their prevalence rarely exceeds 15%. Consequently only 50% of those patients with a positive screening result has a clinically important emotional disorder ("true positives").[10] Clinicians may intuitively recognise this and be unwilling to act on positive test results.[22] This review shows that unselected questionnaire results add little to the clinical encounter. Calls for the routine application of such questionnaires in non-psychiatric settings[3] seem therefore not to be supported.

A second explanation is that clinicians who have not been trained in psychiatry are not confident in dealing with emotional disorders. Supporting this conclusion is the observation that feedback is most effective when it is accompanied by an educational programme and provision of a dedicated outside referral agency that assumes responsibility for management.[11] Our results also complement recent research, which shows that simple educational interventions such as the provision of guidelines on the detection and management of depression in primary care have little impact.[23] However, more complex strategies for quality improvement, in which the feedback of individualised positive test results is accompanied by increased resources and local educational interventions delivered by opinion leaders, can result in improved outcome for depression.[24]

A third explanation relates to the methods employed in most of the included studies. In all but one, patients were randomised to have questionnaire results fed back to the clinician or not.[11] It is possible that receiving feedback on some patients influences how other patients are managed. This cross contamination could dilute estimates of benefit. A more appropriate design would be a cluster randomisation study, whereby individual clinicians rather than individual patients are randomised.[6] The largest and most striking result came from a study with several additional methodological problems, including inadequate randomisation, differences in the way in which cases were established between control and intervention arms, and difficulties generalising beyond the practice style of a single motivated doctor.[8]

Our results also show that more patients with emotional disorders would be recognised if every patient had a questionnaire administered, scored by someone other than the clinician, and only the positive test results fed back to the clinician, then. Clinicians therefore ignore raw scores on psychometric questionnaires when they have to add them up and interpret them themselves. This has implications for how screening tests should be implemented and evaluated in routine care settings. More user friendly formats for administration, such as computer based self completed questionnaires[13] and the administration of these questionnaires by other staff are possibilities. However the resources used in administering, scoring, and feeding back results for all patients are substantial and may not be justified by the likely benefits.

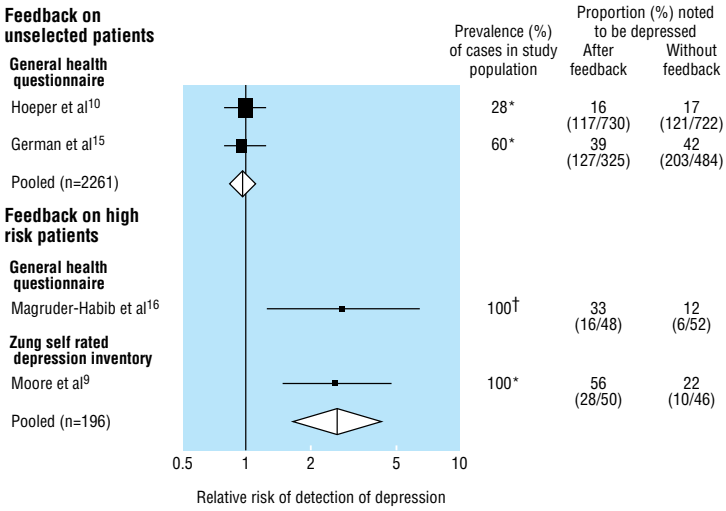



**Fig 2** Meta-analysis of effect of clinician feedback on detection of depression. *According to scores above specified cut-off point on psychiatric questionnaire. †According to structured DSM III interview schedule. Size of black squares represents relative sizes of individual studies

BMJ: first published as 10.1136/bmj.322.7283.406 on 17 February 2001. Downloaded from http://www.bmj.com/ on 18 March 2024 by guest. Protected by copyright.

## What is already known on this topic

Much psychiatric morbidity goes undetected in general practice and general hospital settings

Self completed psychiatric questionnaires have acceptable validity and reliability and might be used as outcomes measures to guide clinical practice, yet research on the impact of results fed back to clinicians is contradictory

## What this study adds

The routine administration of psychiatric questionnaires with feedback to clinicians does not improve the detection of emotional disorders or patient outcome, although those with high scores may benefit

The widely advocated use of simple questionnaires as outcomes measures in routine practice is not supported; more research is needed before this strategy is adopted

This review will also be published and updated in line with emerging published and unpublished evidence on the *Cochrane Library*. We thank Kate Misso for performing the literature searches. The NHS Centre for Reviews and Dissemination and the Department of Health Studies are part of the Medical Research Council Health Services Research Collaboration.

Contributors: SMG had the original idea for the review, initiated and managed it, designed the protocol, scrutinised the literature searches, extracted and tabulated the data, conducted the analyses, and drafted all versions of the paper; he will act as guarantor. TAS and AOH helped initiate the review, extracted the data, provided ongoing academic supervision of the review, and commented on protocols and all drafts of the paper.

Funding: SMG is supported by the UK Medical Research Council Health Services research training fellowship programme.

Competing interests: None declared.

1 Goldberg D, Huxley P. *Mental illness in the community*. London: Tavistock, 1980.
2 Van Hemert AM, Hengeveld MW, Bolk JH, Rooijmans HG, Vandenbroucke JPSO. Psychiatric disorders in relation to medical illness among patients of a general medical out-patient clinic. *Psychol Med* 1993;23:167-73.
3 Wright A. Should general practitioners be testing for depression? *Br J Gen Pract* 1994;44:132-5.
4 Goldberg DP, Williams P. *The user's guide to the general health questionnaire*. Windsor: NFER-Nelson, 1988.
5 Jadad AR, Moore RA, Carroll D. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled Clin Trials* 1996;17:1-12.
6 Campbell MK, Grimshaw JM. Cluster randomised trials: time for improvement. *BMJ* 1998;317:1171.
7 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clin Trials* 1986;7:177-88.
8 Johnstone A, Goldberg D. Psychiatric screening in general practice. *Lancet* 1976;1:605-12.
9 Moore JT, Silimperi DR, Bobula JA. Recognition of depression by family medicine residents: the impact of screening. *J Fam Pract* 1978;7:509-13.
10 Hoeper EW, Nycz GR, Kessler JD, Pierce WE. The usefulness of screening for mental illness. *Lancet* 1984;1:33-5.
11 Mazonson PD, Mathias SD, Fifer SK, Beusching DP, Patrick DL. The mental health patient profile: does it change primary care physicians practice patterns? *J Am Board Fam Pract* 1994;9:336-45.
12 Dowrick C, Buchan I. Twelve month outcome of depression in general practice: does detection or disclosure make a difference? *BMJ* 1995;311:1274-6.
13 Lewis G, Sharp D, Bartholomew J, Pelosi AJ. Computerized assessment of common mental disorders in primary care: effect on clinical outcome. *Fam Pract* 1996;13:120-6.
14 Linn LS, Yager J. The effect of screening, sensitisation and feedback on notation of depression. *J Med Educ* 1980;20:942-53.
15 German PS, Shapiro S, Skinner EA. Detection and management of mental health problems of older patients by primary care providers. *JAMA* 1987;257:489-96.
16 Magruder Habib K, Zung WW, Feussner JR. Improving physicians' recognition and treatment of depression in general medical care. Results from a randomized clinical trial. *Med Care* 1990;28:239-50.
17 Beck AT, Steer RA. *The Beck depression inventory manual*. San Antonio, TX: Psychological Corporation, Harcourt Brace Jovanovich, 1987.
18 Zung WWK. A self rating depression rating scale. *Arch Gen Psychiatry* 1965;12:63-70.
19 Derogatis LR. *Symptom checklist-90-R (SCL-90-R) administration, scoring and procedures manual*. 3rd ed. Minneapolis, MN: National Computer Systems, 1994.
20 Ware JE, Snow KK, Kosinski M, Gandek B. *SF-36 health survey: manual and interpretation guide*. Boston, MA: Health Institute, New England Medical Centre, 1993.
21 Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical epidemiology: a basic science for clinical medicine*. Boston, MA: Little, Brown, 1991.
22 Goldberg D. The use of the general health questionnaire in clinical work. *BMJ* 1986;293:1188-9.
23 Thompson C, Kinmonth J, Stevens L, Peveler RC, Stevens A, Ostler KJ, et al. Effects of a clinical-practice guideline and practice-based education on detection and outcome of depression in primary care: Hampshire Depression Project randomised controlled trial. *Lancet* 2000;355:50-7.
24 Wells KA, Sherbourne C, Schoenbaum M, Duan N, Meridith L, Unutzer J, et al. Impact of disseminating quality improvement programmes for depression in managed primary care: a randomized controlled trial. *JAMA* 2000;283:212-20.

*(Accepted 1 December 2000)*

*One hundred years ago*

## Looking backwards in 2001

The Editor of the *Scarborough Post* has had the courage to celebrate the first year of the new century by issuing a number dated January 1st 2001. The general news is a little scanty, for there is only a telegram announcing fighting in Ecuador, which sounds rather like the nineteenth century, but there is a series of retrospective articles which are decidedly ingenious. The men at the end of the twentieth century, living in flats, using electricity for all sorts of domestic purposes, getting their meals delivered ready cooked from central stations by pneumatic tubes, and themselves travelling 150 miles an hour in "monorail" trains or amusing themselves in air ships, look back with pitying astonishment at the discomforts which their fore-fathers at the end of the nineteenth century endured. Whether they live any longer or are much healthier we are not told, and as no mention is made of any sport except racing and hunting, and nobody need walk anywhere, the generality must be in danger of dying prematurely of gout. However they have their hair cut in antiseptic toilet saloons, and the dead are cremated. Medicine still flourishes, but some of our correspondents will be glad to know that those who follow it wear a distinctive garb, which is however nothing more than the silk top hat and formal frock coat of the nineteenth century. Other classes and professions, except the solicitors, are more picturesquely clad, but we may observe that even in the last quarter of the nineteenth century not every medical man wore the garb which is to be his in the future, and we well remember the excitement caused in a cathedral town some years ago, when, during a visit of the British Medical Association, one of the leading surgeons of the city appeared so arrayed. People are still to have the old-fashioned diseases, and we are given the curious piece of information that "law and medicine have become more specialised, and in addition to the general practitioner we have our travelling specialists in pneumonia, scarlet fever, diabetes, and other disorders, who visit Scarborough twice a week." Our contemporary's prophetic powers seem to have broken down a little here, and, we would fain hope that there will not be quite so many advertisements of quack secret remedies as the print before us would appear to assume.

(*BMJ* 1901;i:40)



BMJ: first published as 10.1136/bmj.322.7283.406 on 17 February 2001. Downloaded from http://www.bmj.com/ on 18 March 2024 by guest. Protected by copyright.