Cochrane
Library

Cochrane Database of Systematic Reviews

Screening and case finding instruments for depression (Review)

Gilbody S, House A, Sheldon T

Gilbody S, House A, Sheldon T.
Screening and case finding instruments for depression.
Cochrane Database of Systematic Reviews 2005, Issue 4. Art. No.: CD002792.
DOI: [10.1002/14651858.CD002792.pub2](https://doi.org/10.1002/14651858.CD002792.pub2).

www.cochranelibrary.com

TABLE OF CONTENTS

HEADER	1
ABSTRACT	1
PLAIN LANGUAGE SUMMARY	2
BACKGROUND	3
OBJECTIVES	3
METHODS	3
RESULTS	6
DISCUSSION	8
AUTHORS' CONCLUSIONS	11
ACKNOWLEDGEMENTS	11
REFERENCES	12
CHARACTERISTICS OF STUDIES	15
DATA AND ANALYSES	22
Analysis 1.1. Comparison 1 Recognition of depression following feedback, Outcome 1 Recognition of depression following feedback [all studies].	22
Analysis 1.2. Comparison 1 Recognition of depression following feedback, Outcome 2 Recognition of depression following feedback [unselected patients].	23
Analysis 1.3. Comparison 1 Recognition of depression following feedback, Outcome 3 Recognition of depression following feedback [high risk patients].	23
Analysis 2.1. Comparison 2 Management of depression following feedback, Outcome 1 Any intervention for depression.	24
Analysis 2.2. Comparison 2 Management of depression following feedback, Outcome 2 Prescription of anti-depressants.	24
Analysis 3.1. Comparison 3 Outcome of depression following feedback, Outcome 1 Short term outcome of depression (0-6 months) - Dichotomous outcomes from depression rating scales.	25
Analysis 3.2. Comparison 3 Outcome of depression following feedback, Outcome 2 Short term outcome of depression (0-6 months) following feedback [depression rating scale endpoint scores].	25
WHAT'S NEW	25
HISTORY	25
CONTRIBUTIONS OF AUTHORS	25
DECLARATIONS OF INTEREST	26
SOURCES OF SUPPORT	26
NOTES	26
INDEX TERMS	26

[Intervention Review]

Screening and case finding instruments for depression

Simon Gilbody¹, Allan House², Trevor Sheldon¹

¹Department of Health Sciences, University of York, York, UK. ²Leeds Institute of Health Sciences, University of Leeds, Leeds, UK

Contact address: Simon Gilbody, Department of Health Sciences, University of York, Seebohm Rowntree Building, York, YO10 5DD, UK.
sg519@york.ac.uk.

Editorial group: Cochrane Common Mental Disorders Group

Publication status and date: Edited (no change to conclusions), published in Issue 1, 2009.

Citation: Gilbody S, House A, Sheldon T. Screening and case finding instruments for depression. *Cochrane Database of Systematic Reviews* 2005, Issue 4. Art. No.: CD002792. DOI: [10.1002/14651858.CD002792.pub2](https://doi.org/10.1002/14651858.CD002792.pub2).

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

Screening or case finding instruments have been advocated as a simple, quick and inexpensive method to improve detection and management of depression in non-specialist settings, such as primary care and the general hospital. However, screening/case finding is just one of a number of strategies that have been advocated to improve the quality of care for depression. The adoption of this seemingly simple and effective strategy should be underpinned by evidence of clinical and cost effectiveness.

Objectives

To determine the clinical and cost effectiveness of screening and case finding instruments in: (1) improving the recognition of depression; (2) improving the management of depression, and (3) improving the outcome of depression.

Search methods

The researchers undertook electronic searches of The Cochrane Library (Issue 4, 2004); The Cochrane Depression, Anxiety and Neurosis Group's Register [2004]; EMBASE (1980-2004); MEDLINE (1966-2004); CINAHL (to 2004) and PsycLIT (1974-2004). References of all identified studies were searched for further trials, and the researchers contacted authors of trials.

Selection criteria

Randomised controlled trials of the administration of case finding/screening instruments for depression and the feedback of the results of these instruments to clinicians, compared with no clinician feedback. Trials had to be conducted in non-mental health settings, such as primary care or the general hospital. Studies that used screening strategies in addition to enhanced care, such as case management and structured follow up, were specifically excluded.

Data collection and analysis

Citations and, where possible, abstracts were independently inspected by researchers, papers ordered, re-inspected and quality assessed. Data were also independently extracted. Data relating to: (1) the recognition of depression; (2) the management of depression and (3) the outcome of depression over time were sought. For dichotomous data the Relative Risk (RR), 95% confidence interval (CI) were calculated on an intention-to-treat basis. For continuous data, weighted and standardised mean difference were calculated. A series of a priori sensitivity analyses relating to the method of administration of questionnaires and population under study were used to examine plausible causes of heterogeneity.

Main results

Twelve studies (including 5693 patients) met our inclusion criteria. Synthesis of these data gave the following results:

(1) the recognition of depression: according to case note entries of depression, screening/case finding instruments had borderline impact on the overall recognition of depression by clinicians (relative risk 1.38; 95% confidence interval 1.04 to 1.83). However, substantial

heterogeneity was found for this outcome. Screening and feedback, irrespective of baseline score of depression has no impact on the detection of depression (relative risk 1.00; 95% confidence interval 0.89 to 1.13). In contrast, three small positive studies using a two stage selective procedure, whereby patients were screened and only patients scoring above a certain threshold were entered into the trial, did suggest that this approach might be effective (relative risk 2.66; 95% confidence interval 1.78 to 3.96). Separate pooling according to this variable reduced the overall level of heterogeneity. Publication bias was also found for this outcome.

(2) the management of depression: according to case note entries for active interventions and prescription data, a selected subsample of all studies reported this outcome and found that there was an overall trend to showing a borderline higher intervention rate amongst those who received feedback of screening/case finding instruments (relative risk 1.35; 95% confidence interval 0.98 to 1.85), although substantial heterogeneity between studies existed for this outcome. This result was dependant upon the presence of one highly positive study.

(3) the outcome of depression: few studies reported the impact of case finding/screening instruments on the actual outcome of depression, and no statistical pooling was possible. However, three out of four studies reported no clinical effect ($p < 0.05$) at either six months or twelve months.

No studies examined the cost effectiveness of screening/case finding as a strategy.

Authors' conclusions

There is substantial evidence that routinely administered case finding/screening questionnaires for depression have minimal impact on the detection, management or outcome of depression by clinicians. Practice guidelines and recommendations to adopt this strategy, in isolation, in order to improve the quality of healthcare should be resisted. The longer term benefits and costs of routine screening/case finding for depression have not been evaluated. A two stage procedure for screening/case finding may be effective, but this needs to be evaluated in a large scale cluster randomised trial, with a prospective economic evaluation.

PLAIN LANGUAGE SUMMARY

Screening and case finding instruments for depression

The use of depression screening or case finding instruments has little or no impact on the recognition, management or outcome of depression in primary care or the general hospital.

BACKGROUND

Depression affects between 5% and 10% of individuals and is the third most common reason for consultation in general practice/primary care (Singleton 2001). Eighty percent of patients with depression consult with non-specific physical complaints (Kessler 1999), without spontaneously divulging the psychological nature of their problems (Goldberg 1986). It has been reported that depressive symptoms are not recognized in around 50% of attending patients with depressive disorders (ascertained by research diagnostic interview) in general practice/primary care (Dowrick 1995). Similar problems are also found in general hospital settings, where there is substantial under identification and unmet need (Katon 1996).

Depression is associated with marked reduction in functional capacity and quality of life (Wells 1989). Use of general medical services by depressed patients is 50% to 100% higher than utilisation by similar patients without depressive illness (Simon 1997). The increased economic burden of depression arises due the loss of functioning and productivity and the increased utilisation of medical services, and exceeds the resources devoted to treatment (Greenberg 2003).

A number of approaches have been advocated to improve the quality of care for depression in non specialist settings. One approach involves enhancements in the process of care, such as case management (Von Korff 2001). Another approach involves the use of educational strategies, aimed at improving the skills and practice of non-specialist health professionals (Thompson 2000). One solution that has been advocated is the use of brief self report measures of depression (Wright 1994), which can be used to aid identification - hitherto referred to as case finding/screening instruments. Some interventions have included combinations of all three of these approaches - enhanced care, clinician education and screening (e.g. Wells 1989). Approaches using screening alone may seem superficially appealing in that instruments are available which are brief; easy to administer and score; psychometrically robust (valid and reliable) and acceptable to patients (Williams 2002).

The use of screening instruments has been supported in important policy pronouncements, including guidelines issued by the US Agency for Healthcare Research and Quality (AHRQ) (Pignone 2002) and the UK National Institute of Clinical Excellence (NICE 2004). The potential for these instruments to improve the behaviour of non-specialists in their ability to recognise and manage depression is substantial, but cannot be assumed. The administration of screening and case finding instruments is also not a cost-neutral exercise. The administration of questionnaires for each and every patient in primary care and general hospital settings will require the investment of substantial resources and time.

Even simplistic quality improvement strategies must be supported by evidence of clinical benefit and cost effectiveness. We therefore set out to undertake a systematic review of randomised evaluations of the impact of routine screening and case finding for depression. Our review specifically addresses the use of screening instruments alone, rather than interventions which also offer substantial enhancement in the process of care (such as active case management).

OBJECTIVES

To determine the clinical and cost effectiveness of screening and case finding instruments in: (1) improving the recognition of depression; (2) improving the management of depression, and (3) improving the outcome of depression.

METHODS

Criteria for considering studies for this review

Types of studies

Randomised controlled trials

Types of participants

Studies must include:

1. Patients in non-psychiatric settings, such as general hospital patients and non-selected general practice attenders.

Studies relating to the following patient groups were excluded:

1. Patients being managed by specialist psychiatric services.
2. Child and adolescent populations.
3. Those with learning disabilities or dementia.

Types of interventions

Included studies were those comparing the introduction of a routine form of screening or case finding assessment with a normal routine pattern of care.

Routine care (the control/comparator condition) involved usual patient-doctor interaction, with non-standardised history taking, investigation, referral, intervention and follow up. This did not usually involve the use of screening/case finding instruments, but relied on the traditional channels of patient doctor communication and informal assessment of outcome using clinical history taking, psychiatric/physical examination and recording of progress in clinical notes.

The active intervention involved the addition of a standardised screening/outcome assessment instrument to routine care, with information from the screening/case finding assessment being fed back to the clinician. Hence, depression could be assessed by researchers in both intervention and control conditions, but the active component in an intervention involved the *feeding back* of this information to the clinician.

Depression screening or case finding instruments (i.e. those actively incorporated into routine care) could be either self-completed or interviewer-completed paper and pencil questionnaires. Broader measures of quality of life, which often contain a mental health component, in addition to questions on physical functioning, social functioning and cognition (Bowling 1997) have also been subject to systematic review (Gilbody 2002), and are excluded from the present review

Studies that include substantial enhancements in the process of care - such as the use of case managers or a 'collaborative care' approach were not included in this reviews, since these involve a complex package of care, of which screening might only form a component (Von Korff 2001). The effectiveness of screening within these studies will be difficult to disentangle from other active components. Case management and collaborative care have

already been subject to systematic review (e.g. [Gilbody 2003a](#)) and a related Cochrane review of this topic is currently underway.

Types of outcome measures

Outcome was studied as defined by the authors of each study, with particular attention given to the impact of screening/case finding tools on the following:

1. Recognition of depression - as indicated by recording a diagnosis of depression or reporting depression as an active problem in case notes.

2. Management of depression - as indicated by an active intervention for depression (other than non-specific clinician counseling) by the physician. Such interventions might include outside referral to a specialist mental healthcare professional or the initiation of an empirically supported treatment such as antidepressants or a specific form of psychotherapy.

3. Outcome of depression - as indicated by recovery from depression (however defined) or scores on a depression-specific instrument at follow up.

4. Cost (direct and indirect).

Outcomes were classified as short term (0-6 months); medium term (6-12 months) and; longer term (12 months and over).

Search methods for identification of studies

We used the Collaborative Review Group search strategy on the following databases:

DATABASES We searched: MEDLINE (2004); PsycLIT (2004); EMBASE (2004); CINAHL (2004); BNI/RCN (2004); CDSR (2004); and the Trials Register of the Cochrane

Depression, Anxiety and Neurosis Group (2004).

ELECTRONIC SEARCH STRATEGY

A. Cochrane MEDLINE (2004) optimal RCT search strategy and depression search terms were combined with the following search strategy:

- 1 "Health-Status-Indicators"
- 2 "Outcome-and-Process-Assessment-(Health-Care)"/ all subheadings
- 3 "Outcome-Assessment-(Health-Care)"/ all subheadings
- 4 "Quality-of-Life"/ all subheadings
- 5 (outcome measure*) in ti,ab
- 6 (health outcome*) in ti,ab
- 7 (quality of life) in ti,ab
- 8 measure* in ti,ab
- 9 assess* in ti,ab
- 10 (score* or scoring) in ti,ab
- 11 index in ti,ab
- 12 indices in ti,ab
- 13 scale* in ti,ab
- 14 monitor* in ti,ab
- 15 #8 or #9 or #10 or #12 or #11 or #13 or #14
- 16 outcome* in ti,ab
- 17 #16 near3 #15
- 18 #1 or #2 or #3 or #4 or #5 or #6 or #7
- 19 #17 or #18

B. Cochrane EMBASE (2004) optimal RCT search strategy and depression search terms were combined with the following:

- 1 "health-survey"/ all subheadings
- 2 explode "quality-of-life"/ all subheadings
- 3 "outcomes-research"/ all subheadings

- 4 health outcome* in ti,ab
- 5 quality of life in ti,ab
- 6 outcome measure* in ti,ab
- 7 measure* in ti,ab
- 8 (score* or scoring) in ti,ab
- 9 index in ti,ab
- 10 indices in ti,ab
- 11 scale* in ti,ab
- 12 monitor* in ti,ab
- 13 assess* in ti,ab
- 14 #7 or #8 or #9 or #10 or #11 or #12 or #13
- 15 outcome* in ti,ab
- 16 #15 near3 #14
- 17 #1 or #2 or #3 or #4 or #5 or #6
- 18 #16 or #17

C. We searched PsycLIT (2004) with depression search terms and the following strategy:

- 1 explode "Treatment-Outcomes"
- 2 explode "Psychological-Assessment"
- 3 explode "Quality-of-Life"
- 4 (outcome* or process*) near3 assessment*
- 5 health status indicator*
- 6 health status
- 7 health outcome* in ti,ab
- 8 quality of life in ti,ab
- 9 outcome measure* in ti,ab
- 10 measure* in ti,ab
- 11 assess* in ti,ab
- 12 (score* or scoring) in ti,ab
- 13 index in ti,ab
- 14 indices in ti,ab
- 15 scale* in ti,ab
- 16 monitor* in ti,ab
- 17 #10 or #11 or #12 or #13 or #14 or #15 or #16
- 18 outcome* in ti,ab
- 19 #18 near3 #17
- 20 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9
- 21 #19 or #20

D. We searched CINAHL (2004) with depression search terms and the following strategy:

- 1 explode "Health-Status"/ all topical subheadings / all age subheadings
- 2 explode "Health-Status-Indicators"/ all topical subheadings / all age subheadings
- 3 explode "Outcome-Assessment"/ all topical subheadings / all age subheadings
- 4 "Outcomes-(Health-Care)"/ all topical subheadings / all age subheadings
- 5 explode "Quality-of-Life"/ all topical subheadings / all age subheadings
- 6 health outcome* in ti,ab
- 7 quality of life in ti,ab
- 8 outcome measure* in ti,ab
- 9 measure* in ti,ab
- 10 assess* in ti,ab
- 11 (score* or scoring) in ti,ab
- 12 index in ti,ab
- 13 indices in ti,ab
- 14 scale* in ti,ab

Screening and case finding instruments for depression (Review)

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

15 monitor* in ti,ab
16 #9 or #10 or #11 or #12 or #13 or #14 or #15
17 outcome* in ti,ab
18 #17 near3 #16
19 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8
20 #18 or #19

E. We searched the British Nursing Index/RCN (2004) with depression search terms and the following strategy:

1 health status
2 status indicator*
3 (outcome* or process*) near3 assessment*
4 health outcome*
5 quality of life
6 outcome* measure*
7 assess*
8 score* or scoring
9 index
10 indices
11 scale*
12 monitor*
13 #7 or #8 or #9 or #10 or #11 or #12
14 outcome*
15 #14 near3 #13
16 #1 or #2 or #3 or #4 or #5 or #6
17 #16 or #15

E. We searched the CCTR cd-ROM (2004, issue 3) with depression search terms and the following strategy:

1. HEALTH-STATUS-INDICATORS:ME
2. OUTCOME-AND-PROCESS-ASSESSMENT-HEALTH-CARE:ME
3. OUTCOME-ASSESSMENT-HEALTH-CARE:ME
4. QUALITY-OF-LIFE:ME
5. OUTCOME:TI AND MEASURE*:TI
6. OUTCOME:AB AND MEASURE*:AB
7. HEALTH:TI AND OUTCOME*:TI
8. HEALTH:AB AND OUTCOME*:AB
9. QUALITY:TI NEAR LIFE:TI
10. QUALITY:AB NEAR LIFE:AB
11. MEASURE:TI OR MEASURE:AB
12. ASSESS*:TI OR ASSESS*:AB
13. SCORE*:TI OR SCORING:TI OR SCORE*:AB OR SCORING:AB
14. INDEX:TI OR INDEX:AB
15. INDICES:TI OR INDICES:AB
16. SCALE*:TI OR SCALE*AB
17. MONITOR*:TI OR MONITOR*:AB
18. #11 OR #13 OR #14 OR #15 OR #16 OR #17
19. OUTCOME*:TI OR OUTCOME*:AB
20. #19 AND #18
21. #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10
22. #21 OR #20

F. Citation searches on located trials

G. CONTACTS

We attempted to contact first authors of all identified studies and made contacts with a number of content experts who might have known of research (published and unpublished) in this area.

Data collection and analysis

All potential trials were read by two researchers individually. Each determined if the trial fulfilled inclusion criteria.

Quality Assessment:

Randomised trials of behavioural interventions, designed to influence clinical practice have to cope with the problem of 'cross contamination' between subjects (Gail 1996). For example, a key aim of research summarised in this review is to encourage clinicians to become more aware of, and responsive to, psychological problems through the active feedback standardised questionnaires. However, if clinicians see some patients with and some patients without questionnaire scores, there is a danger that clinical management will change for both intervention and control groups. In other words, receiving feedback for some patients may sensitise the clinician to psychological problems amongst all their patients. The consequence of this 'cross contamination' is that any real advantage of the intervention may be diluted, and no difference may be found between control and intervention arms. One solution to this problem is to use individual clinicians as the unit of randomisation, so that individual clinicians are randomised to receive feedback for all their patients if the clinician is allocated to the intervention arm, and to not receive feedback for any of their patients if the clinician is allocated to a control/ usual care arm (Elbourne 1997; Bland 1997; Ukoumunne 1999). Hence, those studies which use 'cluster randomisation', where clinicians or clinical teams are the unit of randomisation will be judged to be more robust than those which randomise individual patients. Differences between studies where the patient is the unit of randomisation and those where clinicians are randomised were explored in a sensitivity analysis (see below).

In addition we assessed methodological quality in accordance with the Cochrane Handbook, and with the validated scale of Jadad (Jadad 1996) which considers method of randomisation, allocation concealment and intention to treat.

Data Management:

We independently extracted data on the following for each included trial:

- Setting
- Patient population
- Unit of randomisation (patient or clinician)
- Intervention
- Outcomes studied
- Follow up
- Results
- Method of analysis (was there an appropriate 'unit of analysis'?)
- Attrition

Data were extracted by one reviewer (SG) and independently checked by a second reviewer.

Data analysis

For those studies which were sufficiently similar in terms of patients, populations, interventions and outcomes, we undertook a quantitative data synthesis using a fixed effects model for non-heterogenous data. Between-study heterogeneity was assessed using the I² statistic (Higgins 2003) which describes the percentage of total variation across studies that is due to heterogeneity rather

than chance. The I^2 statistic has several advantages over other measures of heterogeneity such as χ^2 , including greater statistical power to detect clinical heterogeneity when fewer studies are available. As a guide, I^2 values of 25% may be considered 'low', 50% 'moderate' and 75%, 'high'. Where significant heterogeneity was found ($I^2 > 50\%$), we undertook a random effects meta-analysis, and sought to explore sources of heterogeneity according to specified a priori causes of heterogeneity outlined below. We anticipated in advance that studies may be sufficiently similar in terms of the populations studied and routine instruments used to allow pooling for the following outcomes to be undertaken:

- Clinicians' recognition of depression
- Intervention for depression
- Severity of depression at follow up

The analysis maintained the study groups according to the original randomisation procedure, and we synthesised data according to the following procedure:

1. For categorical (binary) outcomes, such as recognition of depression or intervention for depression, we calculated risk ratios (Relative Risk - RR).
2. For continuous outcomes, such as patient scores on the GHQ and BDI, we calculated weighted mean differences (WMD) where there was a common metric between studies, and standardised mean differences (SMD) if different scales were used to measure the same underlying construct. Endpoint scores were pooled as a first preference, since assumptions about normality of data can be checked by ensuring that the standard deviation of endpoint scores, when multiplied by 1.65 is less than the mean endpoint score. Pooling was not attempted for data that were non-normally distributed.

For studies which employed a 'cluster randomisation' (such as randomisation by clinician or primary care practice), we sought evidence that clustering had been accounted for by the authors of studies in their analysis. Failure to account for intra class correlation in clustered studies is commonly encountered in primary research, and leads to a 'unit of analysis' error (Divine 1992) - whereby p values are spuriously low, confidence intervals unduly narrow and statistical significance overestimated - causing type I errors (Bland 1997; Gulliford 1999). Where clustering is not accounted for, we present an (*) symbol next to data. Subsequent versions of this review will seek to contact first authors of studies, to seek intra class correlation co-efficients of their clustered data and to adjust for this (Gulliford 1999). In the interim, we also adjusted for clustering using a plausible Intra-cluster correlation co-efficient (ICC) of 0.05 (Thompson 2000) to check the robustness of our initial analyses to unit of analysis error (Donner 2002).

Economic data

There are major difficulties encountered when conducting systematic reviews of cost effectiveness data (Gilbody 1999). First, studies often adopt different perspectives; account for different types of cost data; use different methods of discounting future healthcare costs/benefits and are conducted at different points in time. Second, they are conducted in different countries and healthcare settings with different funding and reimbursement systems, making international comparisons difficult. Third, economic evaluations often make fundamental errors in the analysis of data, by for example, applying parametric analyses to

highly skewed cost data. Where cost data are presented and a formal cost effectiveness analysis is undertaken, their methods and results were simply be described. No formal statistical pooling was attempted for cost or cost-effectiveness data.

Sensitivity analyses

We specified a number of a priori sensitivity analyses, in order to explore the effect of the following on the results obtained in this review:

- Whether the setting of the study or population in receipt of routine screening/case finding influenced the success of the strategy
- Whether the mode of administration of routine screening and outcome measures (e.g. self completed vs interviewer completed) influenced the success of the strategy.
- Whether the mode of feedback of results to the clinician influenced the success of the strategy (e.g. presentation of raw scores, with no interpretation/normative values versus feedback of only high scores with a clear written indication that this indicates a likely emotional disorder).
- Whether cluster randomised studies produced a different result from non-clustered studies.

Where a formal data synthesis had been performed, we visually inspected forest plots to check for obvious between study variation in results and conducted formal tests for heterogeneity. Where significant heterogeneity was found, we sought causes, including the a priori ones outlined above.

Publication bias

Psychiatric research is especially prone to publication bias (Gilbody 2003b), and we sought evidence of any inherent publication bias in our results by plotting 'funnel plots' where this was feasible (Egger 1997). Funnel plot tests for asymmetry were separately conducted using STATA version 8.0, using the *metabias* command.

RESULTS

Description of studies

Twelve studies met our inclusion criteria (Callahan 1994, Dowrick 1995, German 1987, Hoeper 1984, Lewis 1996, Linn 1980, Magruder-Habib 1990, Moore 1978, Weatherall 2000; Weatherall 2000, Whooley* 2000; Williams 1999; Zung 1983)

Setting

Nine studies were conducted in primary care settings, two in general medical outpatients (German 1987; Linn 1980), and one in an elderly inpatient setting (Weatherall 2000). Details of these studies are provided in 'Characteristics of included studies' section.

Patients

There were broadly two types of patient populations who underwent randomisation: (1) 'unselected feedback', where patients were included, irrespective of their baseline score on the instrument under evaluation or pre-existing probability of having some deficit or disorder as measured on the instrument under evaluation (Dowrick 1995; Hoeper 1984; German 1987; Linn 1980; Weatherall 2000; Whooley* 2000; Williams 1999), and (2) 'high risk feedback', whereby patients were only randomised if they scored above a certain level on the instrument under evaluation, or were known to have a pre-existing but unrecognised deficit or

disorder such as depression (Callahan 1994; Lewis 1996; Magruder-Habib 1990; Moore 1978; Zung 1983). For example, an unselected population was recruited by Hoeper 1984, who administered the General Health Questionnaire (see below) to all primary care attenders, and randomised these patients to either have their GHQ score fed back to the clinician or to be withheld - irrespective of their score on the GHQ. Conversely, Magruder-Habib 1990 recruited only patients with a likely pre-existing diagnosis of depression using a two stage procedure. All outpatient attenders were first given the Zung SDI (Zung 1965), and those with high scores were then screened using a standardised diagnostic interview schedule. Only those with a confirmed diagnosis of hitherto unrecognised depression were then randomised to have their Zung SDI score fed back to clinicians in the course of the interview.

Some studies, by the nature of the services under evaluation (e.g. US veterans administration hospitals (Magruder-Habib 1990)) included a greater proportion of elderly patients, or were specifically targeted at elderly patients (Whooley* 2000).

Screening/case finding instrument

The most commonly used instruments were self-completed scales designed to detect depression e.g. Beck Depression Inventory - BDI (Beck 1961); General Health Questionnaire - GHQ (Goldberg 1972); Zung SDI (Zung 1965). Instruments were generally administered in the waiting room by research assistants prior to consultation.

Active intervention and control

The active intervention broadly involved the feedback of instrument test results to the clinician - generally in the form of a sheet containing summary scores and an explanation of the importance of high scores in terms of the likely presence of a psychological disorder. For example, German 1987 provided summary sheets with GHQ scores together with the following statement:

'it has been shown that above a critical symptom level, a psychiatrist is likely to make a psychiatric diagnosis of a non-psychotic emotional disorder. Higher levels of GHQ scores indicate increasing probability of current emotional distress. A score higher than four is regarded as a 'positive' or abnormal result'

The control condition was generally the administration of the case finding instrument to the patient, without the score on this scale being fed back to the clinician.

Trial endpoints & follow up

The most commonly collected trial endpoints were:

- The detection of depression by the clinician during the course of the clinical interview, and;
- The initiation of treatment or intervention for depression.

The detection of and treatment for depression was established by the use of clinician questionnaires or interviews following a patient consultation, or by case note review, whereby written evidence was sought to determine whether the clinician had noted a depressive disorder as being present, or if they had initiated any interventions for an emotional problem (e.g. Magruder-Habib 1990).

Risk of bias in included studies

All studies described themselves as 'randomised', with few giving specific details of method of randomisation and concealment of

allocation. Failure to specify these facets of design are important since they have been shown to be sources of bias in randomised studies (Schulz 1995). One study that did give details of method of randomisation, used a suboptimal alternate allocation (Weatherall 2000).

In the majority of studies, the unit of randomisation was the individual patient, with individual clinicians seeing both intervention and control patients (i.e. receiving feedback from case finding/screening questionnaires for some patients and not for others). This raises problems of 'cross contamination' between subjects and controls and a Hawthorne effect, whereby practice is changed for both subjects and controls by virtue of participation in a study (Roethlisberger 1939). The implications of this facet of study design are explored in more detail in the discussion section. One study used individual clinicians or practices as the unit of randomisation (Whooley* 2000), so that cross contamination was avoided by single clinicians receiving either the control or experimental condition, but not both for their individual patients. This study failed to account for clustering in the analysis of results, making it prone to a 'unit of analysis error' (Divine 1992).

Sample size varied between 52 and 2209, and three studies (Dowrick 1995; Weatherall 2000; Williams 1999) included a power calculation or discussion of the sample size required to detect a specified difference in outcomes between treatment and control groups.

Effects of interventions

1. Effect of screening/case finding on the recognition of depression/anxiety

Nine studies (4194 patients) presented data on the effect of case finding/screening on the recognition of depression (Callahan 1994; Dowrick 1995; German 1987; Hoeper 1984; Linn 1980; Magruder-Habib 1990; Moore 1978; Williams 1999; Whooley* 2000). Visual inspection of forest plots and statistical testing for these data demonstrated substantial heterogeneity between studies (chi square=28.44, p=0.00002; I²=75.4%) and a pooling showed that overall, within this heterogeneous group of studies, screening and case finding had a borderline positive impact on the rate of clinician recognition of depression (eight studies; relative risk 1.38; 95% confidence interval 1.04 to 1.83). An exploration of possible a priori sources of heterogeneity showed that the most plausible explanation was the method of scoring and patient randomisation. Selection and randomisation of patients according to their pre-existing score being above a certain 'cut off' (high risk feedback) seemed to produce a positive and significant effect size (relative risk 2.66; 95% confidence interval 1.78 to 3.96), whereas 'unselected feedback' produced no improvement in recognition rates by clinicians (relative risk 1.00; 95% confidence interval 0.89 to 1.13). Subgroup analyses according to this plausible source of heterogeneity sought to reduce the overall levels of statistical heterogeneity (high risk feedback studies: Chi²=0.2, p=0.99, I²=0%; unselected feedback studies: Chi²=5.73, p=0.23, I²=30.2%)

When a test for publication bias was applied to all studies reporting this outcome, an asymmetrical plot was obtained and Eggers's test was significant (Intercept (0 if unbiased) = 2.59; 95% CI = 0.81 to 4.35 p=0.0108). This test became non-significant when only 'unselected feedback' studies were selected.

2. Effect of screening/case finding on the management of depression

Eight studies (2272 patients) presented data on the impact of case finding/screening on management of depression (Callahan 1994; Dowrick 1995; German 1987; Lewis 1996; Linn 1980; Magruder-Habib 1990; Weatherall 2000; Whooley* 2000). Interventions for depression were variously defined, and included the prescription of anti-depressants, referral for psychotherapy or counseling, psychiatric referral or referral to a mental health specialist, and stopping drugs that were known to cause depression (especially in the elderly). Sufficient data were provided to allow two comparisons to be made. Firstly, where any of the above interventions were recorded in both intervention and control groups ('any intervention for depression'); and secondly, where the specific prescription or initiation of anti-depressants (an empirically supported treatment) was recorded in both intervention and control groups ('prescription of anti-depressants'). For the outcome 'any intervention for depression', there was a borderline higher intervention rate amongst those who received feedback from screening/case finding instruments (relative risk 1.35; 95% confidence interval 0.98 to 1.85), although substantial heterogeneity between studies existed for this outcome ($\text{Chi}^2=35.6$, $p=0.0001$, $I^2=80.3\%$). This outcome was highly dependant upon the presence of two of the eight studies (Callahan 1994; Magruder-Habib 1990) which showed a much stronger effect size than other studies, and which were also studies that employed 'high risk feedback'. For the outcome 'prescription of anti-depressants', a borderline significant effect size was noted (relative risk 1.15; 95% confidence interval 1.01 to 1.32). Substantial heterogeneity existed for this outcome ($\text{Chi}^2=27.4$, $p=0.00001$, $I^2=81.8\%$) and the effect size was highly dependant upon the presence of a single 'high risk feedback' study (Callahan 1994), the removal of which substantially reduced the level of heterogeneity amongst the remaining seven studies.

One well conducted large scale negative study (Hoeper 1984) did not report outcomes relating to the management of depression. When a test for publication bias was applied to all studies reporting this outcome, a symmetrical plot was obtained and Eggers's test was non-significant ($p=0.36$).

3. Effect of screening/case finding on the outcome of depression

Four studies (Callahan 1994; Dowrick 1995; Lewis 1996; Whooley* 2000) presented data on the impact of screening/case finding on the short term (0-6 month) outcome of depression. Unfortunately, for three studies (Dowrick 1995; Lewis 1996; Whooley* 2000), depression outcomes were not presented in a form that allowed statistical pooling (no standard deviation or change data rather than endpoint data presented). However, each of these three studies recorded non-significant results ($p<0.05$). One study (Callahan 1994) presented dichotomous outcomes for depression (a fall below 'caseness' on the Hamilton Depression rating scale at six month follow-up), which was non-significant (relative risk 1.08; 95% confidence interval 0.49 to 2.40). One study (Dowrick 1995) presented 12 month follow up data, and found no significant difference (between group difference on BDI = zero, 95% confidence interval -3 to +4, $p=0.924$).

Insufficient data were available to test for publication bias with respect to this outcome.

4. Other outcomes

No study examined the costs and resource use associated with routine case finding/screening. No study examined patients' views about the usefulness or acceptability of standardised instruments for detecting psychiatric disorders.

DISCUSSION

Main findings

The main findings of this review are that screening/case finding instruments, when routinely administered and fed back, have little impact on the overall recognition rates of depression. Three relatively small size studies did find a positive effect for depression case finding/screening, but these involved a logistical more complex two stage procedure of screening and the selective feedback of positive results for those who score above a certain threshold. The actual management of depression by clinicians was much more variable and there was no clear association between routine case finding/screening and improved management. Lastly, there were very few data on the actual outcomes of depression, but there seemed to be little impact on actual outcomes for depression at six and twelve months. Each of these findings will now be explored in more detail

1. Effect of screening/case finding on the recognition of depression

The finding that routinely administered case finding/screening instruments for depression have little impact on the actual recognition of depression is a robust finding, based upon several large scale and well conducted studies. Where substantial heterogeneity was found to exist between studies, a plausible and significant cause for this heterogeneity was established. Positive studies were almost consistently those which had employed a logistical more complex two stage screening and feedback method - 'high risk feedback'. This positive finding must be tempered by the fact that the positive studies were all of relatively small size - compared to those which employed 'unselected feedback'. A consideration of why 'routine feedback' is ineffective, whilst 'high risk feedback' is possibly effective is discussed below (see - 'clinical implications').

2. Effect of screening/case finding on management of depression

Broadly defined management outcomes ('any intervention for depression') and a specific empirically supported treatment ('prescription of anti-depressants') were not clearly influenced by screening. Studies reporting these outcomes were fewer and of smaller size than those reporting recognition rates. Substantial heterogeneity existed between studies and a ready source of this heterogeneity was not apparent. However, it is worthy of note that several large scale studies (e.g. Hoeper 1984) that had found no effect on recognition rates for depression had not studied or reported on the actual management of depression. The studies that showed a positive effect for the outcome of depression recognition were also those that studied and found a positive effect for management, in terms of outside referral and anti-depressant prescription (e.g. Callahan 1994). Although non-significant, the results of this outcome were also substantially altered the presence of one strongly positive study (Callahan 1994). The impact of screening on the actual management of depression therefore seems to be minimal.

3. Effect of screening/case finding on the outcome of depression

In comparison with the previous two endpoints - clinician recognition and management - the impact of case finding/screening on the actual outcome of depression was rarely reported,

and was mostly in a form that precluded meta-analytic pooling. Of those studies that did report this outcome, it is worth of note that all had found a positive effect on clinician management, but each failed to show any significant impact on depression outcome (when *p* values alone were considered). This is perhaps the most important outcome and no benefit for screening/case finding is shown within this review.

Methods of the review

Comparisons with other reviews in this area

Traditional (non-systematic) review articles in this area (Meakin 1992; Wright 1994) have produced contradictory recommendations without any clear indication as to how their authors have arrived at their conclusions. The present review, in contrast, produces a series of conclusions with a clear and explicit outline of the methods by which those conclusions were arrived at. This demonstrates the major advantage of systematic reviews over traditional review articles. Other researchers have also sought to apply systematic review methods to establish the effectiveness of screening for depression (e.g. Pignone 2002 - on behalf of the US Preventative Services Task Force). The results presented in the present Cochrane review are generally less supportive of routine screening/case finding than those of Pignone 2002. Comparisons of the methods and results of both reviews shows the following differences:

- The present review is a more up to date summary of the research evidence, with more comprehensive searches and databases;
- Several studies (based on 2349 patients) were included in the present review, which were not identified or included by Pignone (German 1987; Hoeper 1984; Weatherall 2000)
- Several studies were excluded from the present review which were included by Pignone 2002, as they fell without of our tight inclusion criteria. For example, Pignone 2002 included several large positive studies, which were clearly not randomised and are excluded from the present review (Johnstone 1976; Reilfer 1996).

Perhaps most importantly, a more heterogenous sample of studies were included by Pignone 2002, including those where screening was embedded within a complex package of patient care and clinician support (quality improvement strategies) (Katzelnick 2000; Rost 2002; Wells 2000). Of particular note was the inclusion of a large positive US study - the Partners in Care study (Wells 2000). The components of this strategy included: intensive face to face clinician education; computerised decision support and individualised treatment algorithms; the active support of a nurse case manager; regular consultation-liaison with a specialist mental health clinician (psychologist or psychiatrist). Screening was only one component of this package and the effectiveness of screening/case finding as an active ingredient cannot be assumed or established.

Once again, the use of systematic research methods allows the critical reader to disentangle the reasons behind discordant results from two reviews of apparently similar areas of clinical practice and policy.

Identification of studies

It is possible that further research exists which was not identified within the literature search strategies. Some of this research emerged following the publication of an earlier version of this review in a paper journal (Gilbody 2001), and highlights the

advantages of publication in peer reviewed journals in enhancing the quality and comprehensiveness of the review. Similarly, data were left incompletely reported in several studies, but which potentially could have been included in this review. The present review is therefore still likely to be incomplete, but will be published and updated in line with existing and emerging data as reviewed within the Cochrane library.

Examination of heterogeneity

The research included in the present review was subject to a large degree of heterogeneity. This became apparent when the methods and results of individual studies were described in a systematic way. Important sources of heterogeneity anticipated in advance were those relating to the mode of administration and feedback of outcomes measure (the 'unselected' versus 'high risk' approach). The present review illustrates the complementary nature of quantitative and more qualitative approaches to the examination and exploration of sources of heterogeneity. The use of separate statistical pooling for divergent approaches to feedback can be defended on both a statistical and an intuitive basis. The clinical implications of 'unselected' versus 'high risk' feedback are explored in more detail in the following sections.

Publication bias

The present review has highlighted two problems in the examination of the influence of publication bias: the difficulty in applying tests for publication bias, and the difficulty in interpreting the tests that are used.

All published forms of research are potentially subject to publication bias, and there are reasons why psychiatric research is likely to be just as susceptible as research in other areas and specialities (Gilbody 2003b). Conventional tests for publication bias, such as the funnel plot, rely upon two criteria being satisfied: First, studies must be sufficiently similar in terms of participants and interventions to justify a formal statistical pooling in the form of a meta-analysis. Secondly, the published literature must include a sufficient number of studies with a wide range of sample sizes, providing a mix of smaller studies and one or more larger studies with which to construct a funnel plot.

When applying this method of analysis to the group of studies that included the detection of mood disorders as an outcome following feedback, then the second criterion was fulfilled, with a range of study sizes between 80 and 2203. However, for reasons outlined previously, there was felt to be substantial heterogeneity between studies. When a funnel plot was applied, the asymmetrical plot that was obtained was likely to be a reflection of underlying heterogeneity, where this was also a function of sample size. Egger 1997 urges caution in making the assumption that asymmetrical funnel plots are only indicative of publication bias, and the present review provides an interesting example of this. Petticrew 1999 have also demonstrated the potential for heterogeneity to produce asymmetrical funnel plots, where differences in effect size were related to the underlying quality of observational research in the area of heart disease.

Implications for clinical practice

It is perhaps surprising that the uniform administration of well validated case finding instruments, such as the GHQ, with sensitivities and specificities of over 70 and 90% respectively in their ability to detect psychiatric disorders (Goldberg 1972; Goldberg 1988), has not been found to substantially influence

actual clinician behaviour. Routine case finding only becomes effective in increasing the rate of recognition of emotional disorders when there is some form of screening procedure, whereby an instrument is administered, scored by someone other than the clinician, and only those with high scores have their results fed back to the clinician (e.g. [Callahan 1994](#); [Magruder-Habib 1990](#)). Routine administration combined with selective feedback is, however, unlikely to form a model for routine practice, nor does it reflect current UK practice, since this strategy is likely to require that an additional person be employed in order to administer score and feedback outcome measures to the clinician.

There are a number of possible explanations for the observed result. The most likely explanation is the relationship between psychometrics and clinical decision making. It is predictive value (rather than sensitivity and specificity) which is of most interest to clinicians in the context of routine care - i.e. the proportion of those predicted by the test as having the disease who turn out to have the disease ([Sackett 1991](#)). Crucially, positive predictive value increases according to the prevalence of a disorder in the population tested. Whilst unrecognised emotional disorders form a significant portion of the clinical caseload in non-psychiatric services, this is rarely going to exceed 15%. The consequence is that of those patients with a positive screening result, only 50% will turn out to actually have an emotional disorder (i.e. be 'true positives') ([Hoepfer 1984](#)). Equally, the workload and outside referral rate is likely to rise dramatically if all positive test results are acted upon when positive predictive value is much lower than quoted sensitivities and specificities. Clinicians may intuitively recognise this fact and will be unwilling to act on positive test results ([Goldberg 1986](#)).

A major limitation of the research presented in this review is the fact that case definition of depression is generally based upon a questionnaire score above a certain cut off point, rather than some 'gold standard', such as a standardised research interview. Thus, the principle trial endpoint - rates of recognition of emotional disorders - uses this imperfect form of case definition. Research shows that questionnaires consistently overestimate the true prevalence of clinically important emotional disorders (i.e. those worthy of intervention) - e.g. [Feldman 1987](#). It should perhaps therefore be less surprising that clinicians in this review uniformly ascribed far fewer patients as having emotional problems than did questionnaires. However, the negative result for feedback suggests that questionnaire results, in effect, add little to the clinical encounter. Calls for the routine application of such questionnaires in non-psychiatric settings ([Wright 1994](#)) are therefore not supported.

A second explanation is that non-psychiatrists do not feel best equipped to deal with emotional disorders, even when these are uncovered using screening questionnaires. Screening is therefore a necessary, but not sufficient, condition in facilitating the appropriate management of these psychological problems. Supporting this conclusion is the observation from other systematic reviews (e.g. [Gilbody 2003a](#)) that feedback is most effective when it is accompanied by an educational programme and the provision of a dedicated outside referral agency who will readily assume responsibility for management (e.g. [Wells 1989](#)). The results of the present review 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 ([Thompson 2000](#)).

Implications for health policy

Screening tests can only be justified if the instrument is: (i) accurate; (ii) results in a more effective treatment than would otherwise be the case and; (iii) does so with a favourable ratio of costs to benefits ([Cochrane 1971](#); [Mant 1990](#)). The accuracy of an instrument is traditionally determined by the examination of sensitivity, specificity and predictive value ([Knottnerus 2002](#)). Several of the authors of studies in this review justified their choice of instrument with reference to sensitivity and specificity as determined in prior validation studies. Only one examined or published these key psychometric properties within the populations that were recruited or randomised ([Williams 1999](#)). However, it is predictive value which is of most interest to clinicians in the context of routine care - i.e. the proportion of those predicted by the test as having the disease who turn out to have the disease ([Sackett 1991](#)). Predictive value increases as the incidence of disease in the population under investigation increases and this is essentially what is happening when the instrument is administered to all patients and only those with positive score have their results 'fed back'. This is a likely explanation of the improved recognition by clinicians when feedback occurs with only 'high risk' patients as opposed to feedback with all patients. Further research might seek to evaluate the routine use of outcome measures using basic psychometric criteria such as sensitivity, specificity and predictive value.

The second criterion which must be fulfilled for a screening instrument is that its use should result in effective treatment. The evidence outlined in the present review shows that this is under researched, and the research that has been published reflects a positive subsamples of the existing studies. Even when routine feedback does change clinical management, a general finding was that there was no subsequent improvement in clinical outcome (e.g. [Dowrick 1995](#)). The last criterion to be satisfied is that the benefits of screening should outweigh cost. Cost can include the costs (monetary, time and forgone opportunity) incurred through the introduction of routine outcome measurement, and no studies in this review measured this. Additionally, cost involves the harm which might be done through routine outcome measurement in terms of the initiation of treatment for those wrongly identified as having some psychological disorder ('false positives'), or the initiation of resource intensive referral or intervention for those who might be identified as having some emotional problem, but which might be self limiting. Further research is needed in all these respects and in the absence of such research, then it would be imprudent to recommend the introduction of routine outcome measurement in routine care settings.

One substantial policy initiative in recent years has been the development, evaluation and implementation of enhanced models of care for depression ([Von Korff 2001](#)). Enhanced models of care, such as 'collaborative care', are generally effective ([Gilbody 2003a](#)). Several trials of enhanced care include screening strategies (e.g. [Wells 1989](#), [Katzelnick 2000](#)). The degree to which screening is a necessary or active ingredient in enhancing the impact of these strategies is not addressed by the current review. It remains possible, as has been advanced by the US Preventative Services Task Force, that 'larger benefits have been observed in studies in which the communication of screening results is coordinated

with effective follow up and treatment' (Pignone 2002). To our knowledge, no studies have compared enhanced packages of care (including screening) with enhanced packages of care (without screening), and we cannot use the results of this review to comment on this assertion. A related Cochrane review addressing this question is currently in preparation. Unfortunately, influential policy guidelines such as those produced in the UK (NICE 2004) have taken this assertion to mean that screening - by its self - can be an effective strategy to improve the quality and outcome of care. The results of our review can clearly be used to refute the superficial appeal of screening strategies - alone - in improving the outcomes of depression.

Implications for research

Further studies into the impact of routine case finding/screening instruments for depression may be justified; including the potentially effective 'high risk strategy' versus 'routine feedback'. Such studies should examine the impact on clinician management and longer term outcome of depression. There remains a need to understand the role of routinely administered instruments in clinical care and decision making. Qualitative research, conducted within the context a randomised trial will help us understand how and why this approach is either used or ignored by clinicians. The cost effectiveness of routine case finding/screening should also be established using a concurrent economic evaluation within a randomised trial. Cluster randomisation rather than individual randomisation should be the gold standard method of research for this and other strategies designed to influence clinician behaviour, but this has rarely been employed.

AUTHORS' CONCLUSIONS

Implications for practice

There is substantial evidence that routinely administered case finding/screening questionnaires for depression have little or no impact on the detection and management of depression by clinicians. Practice guidelines and recommendations to adopt this strategy in isolation in order to improve the quality of healthcare should be resisted.

Implications for research

The longer term benefits and costs of routine screening/case finding for depression have not been evaluated. A two stage procedure for screening/case finding may be effective, but this needs to be evaluated in a large scale cluster randomised trial, with a prospective economic evaluation. Further research is need to assess whether screening enhances the effectiveness of approaches such as 'case management'.

ACKNOWLEDGEMENTS

We are grateful to Kate Misso and Hugh McGuire for conducting literature searches. We are indebted to the UK Medical Research Council for funding a Research Training Fellowship, within which this review was initially conducted.

REFERENCES

References to studies included in this review

Callahan 1994 {published data only}

Callahan CM, Dittus RS. Primary care physicians' medical decision making for late life depression. *Journal of General Internal Medicine* 1996;**11**:218-9.

* Callahan CM, Hendrie HC, Dittus RS, Brater DC. Improving treatment of late life depression in primary care: a randomized clinical trial. *Journal of the American Geriatrics Society* 1994;**42**:839-46.

Dowrick 1995 {published data only}

Dorwick C. Does testing for depression influence diagnosis or management by general practitioners?. *Family Practice* 1995;**12**:461-5.

* Dowrick C, Buchan I. Twelve month outcome of depression in general practice: does detection or disclosure make a difference?. *BMJ* 1995;**311**:1274-6.

German 1987 {published data only}

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.

Shapiro S, German PS, Skinner EA, VonKorf M, Turner RW, Klein LE, et al. An experiment to change the detection and management of mental morbidity in primary care. *Medical Care* 1987;**25**:327-39.

Hoeper 1984 {published data only}

* Hoeper EW, Nycz GR, Kessler JD, Pierce WE. The usefulness of screening for mental illness. *Lancet* 1984;**1**:33-5.

Lewis 1996 {published data only}

* Lewis G, Sharp D, Bartholomew J, Pelosi AJ. Computerized assessment of common mental disorders in primary care: effect on clinical outcome. *Family Practice* 1996;**13**:120-6.

Linn 1980 {published data only}

Linn LS, Yager J. Screening for depression in relationship to subsequent patient and physician behaviour. *Medical Care* 1980;**20**:1233-45.

Linn LS, Yager J. The effect of screening, sensitisation and feedback on notation of depression. *Journal of Medical Education* 1980;**20**:942-53.

Magruder-Habib 1990 {published data only}

* Magruder Habib K, Zung WW, Feussner JR. Improving physicians' recognition and treatment of depression in general medical care. Results from a randomized clinical trial. *Medical Care* 1990;**28**:239-50.

Moore 1978 {published data only}

* Moore JT, Silimperi DR, Bobula JA. Recognition of depression by family medicine residents: the impact of screening. *Journal of Family Practice* 1978;**7**:509-13.

Weatherall 2000 {published data only}

* Weatherall M. A randomized controlled trial of the Geriatric Depression Scale in an inpatient ward for older adults. *Clinical Rehabilitation* 2000;**14**:186-91.

Whooley 2000 {published data only}

* Whooley MA, Stone B, Soghikian K. Randomized trial of case-finding for depression in elderly primary care patients. *Journal of General Internal Medicine* 2000;**15**:293-300.

Williams 1999 {published data only}

* Williams JW, Mulrow CD, Kroenke K. Case-finding for depression in primary care: a randomized trial. *American Journal of Medicine* 1999;**106**:36-43.

Zung 1983 {published data only}

* Zung WW, Magill M, Moore JT, George DT. Recognition and treatment of depression in a family medicine practice. *Journal of Clinical Psychiatry* 1983;**44**:3-6.

References to studies excluded from this review

Calkins 1994 {published data only}

* Calkins DR, Rubenstein LV, Cleary PD. Functional disability screening of ambulatory patients: a randomised controlled trial in a hospital based group practice. *Journal of General Internal Medicine* 1994;**9**:590-2.

Gold 1989 {published data only}

* Gold I, Baraff LJ. Psychiatric screening in the emergency department: its effect on physician behaviour. *Annals of Emergency Medicine* 1989;**18**:875-80.

Goldsmith 1989 {published data only}

* Goldsmith G, Brodwick M. Assessing the functional status of older patients with chronic illness. *Family Medicine* 1989;**21**:38-41.

Johnstone 1976 {published data only}

* Johnstone A, Goldberg D. Psychiatric screening in General Practice. *Lancet* 1976;**1**:605-12.

Katzelnick 2000 {published data only}

Katzelnick DJ, Simon GE, Pearson SD, Manning WG, Helstad CP, Henk HJ. Randomized trial of a depression management program in high utilizers of medical care. *Archives of Family Medicine* 2000;**9**:345-51.

Kazis 1990 {published data only}

* Kazis LE, Callahan LF, Meenan RF, Pincus TS. Health status reports in the care of patients with rheumatoid arthritis. *Journal of Clinical Epidemiology* 1990;**43**:1243-53.

Reilfer 1996 {published data only}

* Reilfer DR, Kessler HS, Bernhard EJ, Leon AC, Martin G. Impact of screening for mental health concerns on health service utilisation and functional status in primary care patients. *Archives of Internal Medicine* 1996;**156**:2593-9.

Rost 2001 {published data only}

Rost K, Nutting PA, Smith J, Werner J, Duan N. Improving depression outcomes in community primary care practice: a randomised trial of the QuEST intervention. *Journal of General Internal Medicine* 2001;**16**:143-9.

Rubenstein 1989 {published data only}

* Rubenstein LV, Calkins DR, Young RT. Improving patient functioning: a randomised trial of functional disability screening. *Annals of Internal Medicine* 1989;**111**:836-42.

Rubenstein 1995 {published data only}

* Rubenstein LV, McCoy JM, Cope DW, Barrett PA, Hirsch SH, Messer KS. Improving patient quality of life with feedback to physicians about functional status. *Journal of General Internal Medicine* 1995;**10**:607-14.

Street 1994 {published data only}

* Street RL Jr, Gold WR, McDowell T. Using health status surveys in medical consultations. *Medical Care* 1994;**32**:732-44.

Wagner 1997 {published data only}

* Wagner AK, Ehrenberg BL, Tran TA, Bungay KM, Cynn DJ, Rodgers WH. Patient based health status measurement in clinical practice: a study of its impact in epilepsy patients. *Quality of Life Research* 1997;**6**:329-41.

Wasson 1992 {published data only}

* Wasson J, Hays R, Rubenstein L, Nelson E, Leaning J, Johnson D, et al. The short-term effect of patient health status assessment in a health maintenance organization. *Quality of Life Research* 1992;**1**:99-106.

Wells 2000 {published data only}

Wells KA, Sherbourne C, Schoenbaum M, Duan N, Meridith L, Unutzer, J. Impact of disseminating quality improvement programmes for depression in managed primary care: a randomized controlled trial. *JAMA* 2000;**283**:212-20.

Additional references
Beck 1961

Beck AT, Ward CH. An inventory for measuring depression. *Archives of General Psychiatry* 1961;**4**:561-71.

Bland 1997

Bland JM, Kerry SM. Statistics notes. Trials randomised in clusters. *BMJ* 1997;**710**:600.

Bowling 1997

Bowling, A. Measuring Health: A review of quality of life measurement scales. 2nd Edition. Milton Keynes: Open University Press, 1997.

Cochrane 1971

Cochrane AL, Holland WW. Validation of screening procedures. *British Medical Bulletin* 1971;**27**:3-8.

Divine 1992

Divine GW, Brown JT, Frazer LM. The unit of analysis error in studies about physicians' patient care behavior. *Journal of General Internal Medicine* 1992;**7**:623-29.

Donner 2002

Donner A, Klar N. Issues in the meta-analysis of cluster randomized trials. *Statistics in Medicine* 2002;**21**:2971-80.

Egger 1997

Egger M, Davey-Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple graphical test. *BMJ* 1997;**315**:629-35.

Elbourne 1997

Elbourne D. Guidelines are needed for evaluations that use a cluster approach. *BMJ* 1997;**315**:1620-1.

Feldman 1987

Feldman E, Mayou R, Hawton K, Arden M, Smith EB. Psychiatric disorders in medical in-patients. *Quarterly Journal of Medicine* 1987;**63**:405-12.

Gail 1996

Gail MH, Mark SD, Carroll RJ, Green SB, Pee D. On design considerations and randomization-based inference for community intervention trials. *Statistics In Medicine* 1996;**15**:1069-92.

Gilbody 1999

Gilbody SM, Petticrew M. Rational decision making in mental health: the role of systematic reviews in clinical and economic evaluation. *Journal of Mental Health Policy and Economics* 1999;**2**:99-107.

Gilbody 2001

Gilbody SM, House AO, Sheldon TA. Routinely administered questionnaires for depression and anxiety: a systematic review. *BMJ* 2001;**322**:406-9.

Gilbody 2002

Gilbody SM, House AO, Sheldon TA. Routine administration of health related quality of life (HRQoL) and needs assessment tools - a systematic review. *Psychological Medicine* 2002;**32**:1345-56.

Gilbody 2003a

Gilbody S, Whitty P, Grimshaw J, Thomas R. Educational and organizational interventions to improve the management of depression in primary care: a systematic review. *JAMA* 2003;**289**:3145-51.

Gilbody 2003b

Gilbody SM, House AO, Sheldon TA. Routine outcome and needs assessment for those with schizophrenia. *Cochrane Library* 2003, Issue 1.

Goldberg 1972

Goldberg D. The Detection of Psychiatric Illness by Questionnaire. Oxford: Oxford University Press, 1972.

Goldberg 1986

Goldberg, D. The use of the general health questionnaire in clinical work. *British Medical Journal* 1986;**293**:1188-9.

Goldberg 1988

Goldberg DP, Williams P. The user's guide to the General Health Questionnaire. Windsor: NFER - Nelson, 1988.

Greenberg 2003

Greenberg PE, Kessler RC. The economic burden of depression in the United States: how did it change between 1990 and 2000?. *Journal of Clinical Psychiatry* 2003;**64**:1465-75.

Gulliford 1999

Gulliford MC, Ukoumunne OC, Chinn S. Components of variance and intraclass correlations for the design of community-based surveys and intervention studies: data from the Health Survey for England 1994. *American Journal Of Epidemiology* 1999;**149**:876-83.

Higgins 2003

Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**:557-60.

Hoeper 1984

Hoeper EW, Nycz GR, Kessler JD, Pierce WE. The usefulness of screening for mental illness. *Lancet* 1984;**1**:33-5.

Jadad 1996

Jadad AR, Moore RA, Carroll D. Assessing the quality of reports of randomized clinical trials: Is blinding necessary?. *Controlled Clinical Trials* 1996;**17**:1-12.

Katon 1996

Katon W. The impact of major depression on chronic medical illness. *General Hospital Psychiatry* 1996;**18**:215-9.

Kessler 1999

Kessler D, Lloyd K, Lewis G, Gray DP. Cross sectional study of symptom attribution and recognition of depression and anxiety in primary care. *BMJ* 1999;**318**:436-40.

Knottnerus 2002

Knottnerus JA, van Weel C. Evaluation of diagnostic procedures. *BMJ* 2002;**324**:477-80.

Mant 1990

Mant D, Fowler G. Mass screening: theory and ethics. *BMJ* 1990;**300**:916-8.

Meakin 1992

Meakin CJ. Screening for depression in the medically ill. *British Journal of Psychiatry* 1992;**160**:212-6.

NICE 2004

National Institute for Clinical Excellence. Depression: core interventions in the management of depression in primary and secondary care. London: HMSO, 2004.

Petticrew 1999

Petticrew M, Gilbody SM, Sheldon TA. Relation between hostility and coronary heart disease. *BMJ* 1999;**319**:917-8.

Pignone 2002

Pignone MP, Gaynes BN, Rushton JL, Burchell CM, Orleans CT, Mulrow CD. Screening for depression in adults: a summary of the evidence for the U.S. Preventive Services Task Force. *Annals of Internal Medicine* 2002;**136**:765-76.

Roethlisberger 1939

Roethlisberger FJ, Dickinson WJ. Management and the Worker. Cambridge, MA: Harvard University Press, 1939.

Sackett 1991

Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical Epidemiology: A basic science for clinical medicine. Boston, MA.: Little, Brown and Company, 1991.

Schulz 1995

Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;**273**:408-12.

Simon 1997

Simon GE, Katzelnick DJ. Depression, use of medical services and cost-offset effects. *Journal of Psychosomatic Research* 1997;**42**:333-4.

Singleton 2001

Singleton N, Bumpstead R, O'Brien M, Lee A, Meltzer HY. Office of National Statistics: Psychiatric Morbidity Among Adults Living in Private Households. London: HMSO, 2001.

Thompson 2000

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.

Ukoumunne 1999

Ukoumunne OC, Gulliford MC, Chinn S, Sterne JA, Burney PG, Donner A. Methods in health service research. Evaluation of health interventions at area and organisation level. *BMJ* 1999;**319**:376-9.

Von Korff 2001

Von Korff M, Goldberg D. Improving outcomes of depression: the whole process of care needs to be enhanced. *BMJ* 2001;**323**:948-9.

Ware 1993

Ware JE, Snoww KK, Kosinski M, Gandek B. SF-36 Health Survey: Manual and Interpretation Guide. Boston, MA.: The Health Institute, New England Medical Centre., 1993.

Wells 1989

Wells KB, Stewart A, Hays RD, Burnam MA, Rogers W, Daniels M, et al. The functioning and well-being of depressed

patients. Results from the Medical Outcomes Study. *JAMA* 1989;**262**:914-9.

Williams 2002

Williams JW, Pignone M, Ramirez G, Stellato CP. Identifying depression in primary care: a literature synthesis of case-finding instruments. *General Hospital Psychiatry* 2002;**24**:225-37.

Wright 1994

Wright A. Should general practitioners be testing for depression?. *British Journal of General Practice* 1994;**44**:132-5.

Young 1987

Young JB, Chamberlain MA. The contribution of the Stanford Health Assessment questionnaire in rheumatology clinics. *Clinical Rehabilitation* 1987;**1**:97-100.

Zung 1965

Zung WW. A self rating depression rating scale. *Archives of General Psychiatry* 1965;**12**:63-70.

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

Callahan 1994

Methods	RCT: Individual patients randomised
Participants	Elderly US primary care patients with a score above 15 on the Hamilton Depression Rating Scale (HDRS) (n=175)
Interventions	<p>Instrument: HDRS</p> <p>Int: Three additional appointments made over a three-month period with the primary care physician. Clinicians provided with written patient specific materials, including HDRS scores; an interpretation of their meaning; a list of all medications and a specific instruction that drugs causing depression should be reviewed; and a written instruction that the presence of depression should be examined and managed appropriately - clinical algorithm provided. (n=100)</p> <p>Cont: No written feedback and no extra visits scheduled (n=75)</p>
Outcomes	<p>Diagnoses of depression.</p> <p>Discontinuation of drugs causing depression.</p> <p>Initiation of antidepressants.</p> <p>Psychiatric referrals.</p> <p>Depression scores</p> <p>Functional status scores. (Symptom Impact Profile - SIP) - not used</p> <p>Follow up at six months</p>

Notes

Risk of bias

Bias	Authors' judgement	Support for judgement
Allocation concealment?	Unclear risk	B - Unclear

Dowrick 1995

Methods	RCT: Individual patients randomised
---------	-------------------------------------

Dowrick 1995 (Continued)

Participants	Consecutive GP attenders (n=116) in Liverpool, UK, with depression score above 14 on the BDI.
Interventions	Instrument: BDI Int: BDI administered pre consultation and depression scores disclosed to GP (n=52). Cont. 1: BDI administered, but not fed back to GP (n=64).
Outcomes	Diagnoses of depression BDI scores at 6 & 12 months
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Allocation concealment?	Low risk	A - Adequate

German 1987

Methods	RCT: Individual patients randomised
Participants	US adult & elderly general medical outpatient attenders (n=1242). Separate interventions for high (n=488) and low (n=754) GHQ scorers.
Interventions	Instrument: GHQ Int: GHQ administered pre consultation and results fed back to clinician, together with an indication that score was high and suggested 'psychiatric diagnosis'. (n=165) Cont: GHQ administered, but not fed back. (n=323)
Outcomes	Detection of depression by clinicians. Presence of depression according to diagnostic interview (DIS). Treatment initiated for depression.
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Allocation concealment?	Unclear risk	B - Unclear

Hoeper 1984

Methods	RCT: Individual patients randomised
Participants	Adult US primary care patients (n=2309)

Hooper 1984 (Continued)

Interventions	Instrument: GHQ Int: GHQ administered by researcher and scores fed back to clinician, with information that a score >5 indicated mental illness. Cont:GHQ administered, but not fed back to clinicians.	
Outcomes	Physician diagnoses of depression at reference visit.	
Notes		
<i>Risk of bias</i>		
Bias	Authors' judgement	Support for judgement
Allocation concealment?	Unclear risk	B - Unclear

Lewis 1996

Methods	RCT: Individual patients randomised	
Participants	UK General practice attenders at a single practice with GHQ-12 score >2. (n=454)	
Interventions	<p>Instrument: GHQ-12 & computerised assessment of psychiatric symptomatology.</p> <p>Int1: GHQ administered & placed in notes with no interpretation or instruction on the presence of mental disorder (n=227 patients).</p> <p>Int 2: Patient asked to complete a computerised assessment and the results of this assessment fed back to the clinician (n=227 patients).</p> <p>Cont: No feedback given (n=227 patients).</p>	
Outcomes	<p>Consultation rates & clinician attribution of encounters as due to psychological or physical problems.</p> <p>Prescription of a psychotropic drug.</p> <p>Rates of outside mental health referrals to outside agencies.</p> <p>GHQ scores at 6 weeks, 3 & 6 months.</p>	
Notes	GHQ assessment used in analysis. computerised group (int 2) ignored.	
<i>Risk of bias</i>		
Bias	Authors' judgement	Support for judgement
Allocation concealment?	Low risk	A - Adequate

Linn 1980

Methods	RCT: Individual patients randomised.	
Participants	New referrals to US medical outpatients (n=150) - mean age 56.	

Linn 1980 (Continued)

Interventions	Instrument: Zung self rating depression scale (SDS). Int 1: SDS administered prior to consultation and results placed at front of notes, together normative values. Physician also asked about depression post consultation. Int 2: SDS fed back to clinician following consultation. Int 3: SDS provided pre consultation, but clinician's impression of depression not elicited. Int 4.: SDS given to clinician following consultation, no impression of depression sought. Int 5.: no screening by SDS, but impression of depression sought.Cont.: no screening by SDS, no physician opinion sought.	
Outcomes	Depression noted in charts. Initiation of treatment for depression.	
Notes		
Risk of bias		
Bias	Authors' judgement	Support for judgement
Allocation concealment?	Unclear risk	B - Unclear

Magruder-Habib 1990

Methods	RCT: Individual patients randomised.	
Participants	Male adult US veterans (mean age 60) attending a US general internal medicine OP clinic with Zung SDS score >50.	
Interventions	Instrument: Zung self rating for depression scale (SDS). Int: SDS administered and fed back to physicians at first clinic assessment visit - placed at front of clinic notes (n=48). Cont: SDS administered but not fed back to clinicians (n=52).	
Outcomes	Recognition of depression. Initiation of management of depression. Scores on SDS at 3, 6, 9 & 12 months.	
Notes		
Risk of bias		
Bias	Authors' judgement	Support for judgement
Allocation concealment?	Unclear risk	B - Unclear

Moore 1978

Methods	RCT: Individual patients randomised.
Participants	US primary care patients 20-60 years with Zung SDS score >50 (n=96).
Interventions	Instrument: Zung SDS score. Int: SDS administered and score fed back, together with written indication of level of depression (n=50). Cont: screened, but no feedback (n=46).
Outcomes	Recognition of depression as indicated in case notes.
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Allocation concealment?	Unclear risk	B - Unclear

Weatherall 2000

Methods	RCT: Individual patients.
Participants	Elderly inpatients, in New Zealand (n=100).
Interventions	Instrument: Geriatric depression rating scale (GDS) Int: GDS administered, together with the Mini Mental State Examination. Scores written in the notes (by hand) and an interpretation of the significance of scores given (n=50). Cont: An Activity of Daily living questionnaire administered in place of the GDS (n=50)
Outcomes	Rate of prescription of antidepressants. Follow up at discharge and three months.
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Allocation concealment?	High risk	C - Inadequate

Whooley 2000

Methods	RCT: Primary care clinics randomised.
Participants	Sequential US family practice attenders over 65 years (n=2,346), of whom 331 had depression.
Interventions	Instrument: Geriatric Depression Scale (GDS) administered by a research assistant.

Whooley 2000 (Continued)

Int: GDS administered, and scored by research assistant. Scores fed back to physicians, with an indication that the score suggested moderate (score 6-10) or severe (11+) depression. In addition, clinic attenders screened positive were offered a series of organised educational sessions. (n=162)

Cont: GDS administered, but scores not fed back. Educational sessions not offered (usual care). (n=169)

Outcomes	<p>Physician diagnosis of depression (case note review, by blinded researcher).</p> <p>Prescription of antidepressants.</p> <p>Healthcare utilisation (number of clinical visits and hospitalisations).</p> <p>Depression scores of the GDS.</p> <p>Outcomes all measured at two years. NB only those with screen positive depression followed up (n=331)</p>
Notes	Cluster randomised study; unit of analysis error

Risk of bias

Bias	Authors' judgement	Support for judgement
Allocation concealment?	Unclear risk	B - Unclear

Williams 1999

Methods	RCT: Individual patients randomised
Participants	Sequential attenders at a US family medicine clinic (n=969)
Interventions	<p>Instrument: CES-D Questionnaire or Single item question 'Have you felt depressed or sad much of the time in the past year?'</p> <p>Int 1: CES-D self administered, scored by researcher and results fed back to clinicians as either 'positive' or negative'. N=323</p> <p>Int 2: Single item question asked and answer yes or no fed back to clinician. N=330</p> <p>Cont: Usual care. N=316.</p> <p>NB all clinicians were given a copy of the 'Quick reference guide for clinicians on the management of depression (Depression Guideline Panel, 1993)</p>
Outcomes	<p>Sensitivity and specificity of the instruments.</p> <p>Recognition of depression from case note review - corroborated by DSM-III-R interview schedule.</p> <p>Severity of depression from DSM-III-R symptom counts.</p> <p>Treatment for depression (referral, antidepressants. Patient and physician satisfaction with care and use of questionnaires).</p> <p>Functional status from the SF36</p>
Notes	<p>Results for Int 1 and Int 2 combined by the authors.</p> <p>Planned follow up only available for 230 patients (positive for depression).</p>

Williams 1999 (Continued)

Risk of bias

Bias	Authors' judgement	Support for judgement
Allocation concealment?	Low risk	A - Adequate

Zung 1983

Methods	RCT: Individuals patients randomised.
Participants	US patients with undetected depression attending a family medicine centre (n=143)
Interventions	Instrument: Zung self rating for depression scale (SDS) Int: Patients' (n=102) SDS results attached to the front of the medical record and the clinician verbally informed of the positive result and asked to evaluate the patient carefully for the presence of depressive disorder. Cont: Patients' (n=41) SDS results not fed back to the clinician.
Outcomes	Notation of depression in the medical notes. SDS scores at 4 weeks & clinical improvement (operationally defined as a decrease of at least 12 points from baseline).
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Allocation concealment?	Unclear risk	B - Unclear

Characteristics of excluded studies [ordered by study ID]

Study	Reason for exclusion
Calkins 1994	Not a specific depression/anxiety measure; Generic Health Related Quality of Life measure.
Gold 1989	Not a specific depression/anxiety measure; Generic Health Related Quality of Life measure.
Goldsmith 1989	Not a specific depression/anxiety measure; Generic Health Related Quality of Life measure.
Johnstone 1976	Non randomised study.
Katzelnick 2000	Complex quality improvement programme.
Kazis 1990	Not a specific depression/anxiety measure; Generic Health Related Quality of Life measure.
Reilfer 1996	Non randomised study.
Rost 2001	Complex quality improvement programme.

Study	Reason for exclusion
Rubenstein 1989	Not a specific depression/anxiety measure; Generic Health Related Quality of Life measure.
Rubenstein 1995	Not a specific depression/anxiety measure; Generic Health Related Quality of Life measure.
Street 1994	Non randomised study.
Wagner 1997	Not a specific depression/anxiety measure; Generic Health Related Quality of Life measure.
Wasson 1992	Not a specific depression/anxiety measure; Generic Health Related Quality of Life measure.
Wells 2000	Complex quality improvement programme.

DATA AND ANALYSES

Comparison 1. Recognition of depression following feedback

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Recognition of depression following feedback [all studies]	9	4194	Risk Ratio (M-H, Random, 95% CI)	1.38 [1.04, 1.83]
2 Recognition of depression following feedback [unselected patients]	6	3825	Risk Ratio (M-H, Fixed, 95% CI)	1.00 [0.89, 1.13]
3 Recognition of depression following feedback [high risk patients]	3	369	Risk Ratio (M-H, Fixed, 95% CI)	2.66 [1.78, 3.96]

Analysis 1.1. Comparison 1 Recognition of depression following feedback, Outcome 1 Recognition of depression following feedback [all studies].

Analysis 1.2. Comparison 1 Recognition of depression following feedback, Outcome 2 Recognition of depression following feedback [unselected patients].

Analysis 1.3. Comparison 1 Recognition of depression following feedback, Outcome 3 Recognition of depression following feedback [high risk patients].

Comparison 2. Management of depression following feedback

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Any intervention for depression	8	2272	Risk Ratio (M-H, Random, 95% CI)	1.35 [0.98, 1.85]
2 Prescription of anti-depressants	6	2017	Risk Ratio (M-H, Fixed, 95% CI)	1.15 [1.01, 1.32]

Analysis 2.1. Comparison 2 Management of depression following feedback, Outcome 1 Any intervention for depression.

Analysis 2.2. Comparison 2 Management of depression following feedback, Outcome 2 Prescription of anti-depressants.

Comparison 3. Outcome of depression following feedback

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Short term outcome of depression (0-6 months) - Dichotomous outcomes from depression rating scales	1	175	Risk Ratio (M-H, Fixed, 95% CI)	1.08 [0.49, 2.40]
2 Short term outcome of depression (0-6 months) following feedback [depression rating scale endpoint scores]	1	454	Std. Mean Difference (IV, Fixed, 95% CI)	1.60 [1.39, 1.82]

Analysis 3.1. Comparison 3 Outcome of depression following feedback, Outcome 1 Short term outcome of depression (0-6 months) - Dichotomous outcomes from depression rating scales.

Analysis 3.2. Comparison 3 Outcome of depression following feedback, Outcome 2 Short term outcome of depression (0-6 months) following feedback [depression rating scale endpoint scores].

WHAT'S NEW

Date	Event	Description
5 November 2008	Amended	Converted to new review format.

HISTORY

Protocol first published: Issue 4, 2000

Review first published: Issue 4, 2005

Date	Event	Description
18 August 2005	New citation required and conclusions have changed	Substantive amendment

CONTRIBUTIONS OF AUTHORS

SG: initial idea for the review. Responsible for protocol development, peer review comment syntheses; scrutiny of literature searches; data extraction and analysis; writing final paper. TS and AH. protocol development and intellectual input; data verification and comments on each draft of the review. SG is the guarantor of the review.

Screening and case finding instruments for depression (Review)

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

DECLARATIONS OF INTEREST

None.

SOURCES OF SUPPORT

Internal sources

- Department of Health Sciences, University of York, UK.
- Academic Unit of Psychiatry, University of Leeds, UK.

External sources

- Medical Research Council, UK.

NOTES

23/8/05:

This review focuses on depression only. On the advice of colleagues, and with the agreement of CCDAN, the title of this review has been altered from

"Routine outcome assessment for depression and anxiety"

to

"Screening and case finding instruments for depression"

This was done between the publication of the protocol and submission of the review, as it was felt to better reflect the nature of the intervention and avoid confusion about the contents of the review.

The authors are considering undertaking a separate review of screening and case finding instruments for anxiety.

INDEX TERMS

Medical Subject Headings (MeSH)

Depression [*diagnosis]; Hospitals, General; Mass Screening [*methods]; Primary Health Care; Randomized Controlled Trials as Topic

MeSH check words

Humans