

## Screening and case-finding instruments for depression: a meta-analysis

Simon Gilbody DPhil, Trevor Sheldon DSc, Allan House DM

∞ See related article page 1023

### ABSTRACT

**Background:** Screening and case-finding has been proposed as a simple, quick and cheap method to improve the quality of care for depression. We sought to establish the effectiveness of screening in improving the recognition of depression, the management of depression and the outcomes of patients with depression.

**Methods:** We performed a Cochrane systematic review of randomized controlled trials conducted in nonmental health settings that included case-finding or screening instruments for depression. We conducted a meta-analysis and explored heterogeneity using meta-regression techniques.

**Results:** Sixteen studies with 7576 patients met our inclusion criteria. We found that the use of screening or case-finding instruments were associated with a modest increase in the recognition of depression by clinicians (relative risk [RR] 1.27, 95% confidence interval [CI] 1.02 to 1.59). Questionnaires, when administered to all patients and the results given to clinicians irrespective of baseline score, had no impact on recognition (RR 1.03, 95% CI 0.85 to 1.24). Screening or case finding increased the use of any intervention by a relative risk of 1.30 (95% CI 0.97 to 1.76). There was no evidence of influence on the prescription of antidepressant medications (RR 1.20, 95% CI 0.87 to 1.66). Seven studies provided data on outcomes of depression, and no evidence of an effect was found (standardized mean difference  $-0.02$ , 95% CI  $-0.25$  to  $0.20$ ).

**Interpretation:** If used alone, case-finding or screening questionnaires for depression appear to have little or no impact on the detection and management of depression by clinicians. Recommendations to adopt screening strategies using standardized questionnaires without organizational enhancements are not justified.

Une version française de ce résumé est disponible à l'adresse [www.cmaj.ca/cgi/content/full/178/8/997/DC1](http://www.cmaj.ca/cgi/content/full/178/8/997/DC1)

CMAJ 2008;178(8):997-1003

DOI:10.1503/cmaj.070281

Depression affects 5%–10% of people,<sup>1</sup> but about half of these cases are missed in primary care<sup>2</sup> and in general hospital settings.<sup>3</sup> The use of screening and case-finding instruments to improve the quality of care

for depression has been supported by recommendations from the Canadian Task Force on Preventive Health Care,<sup>4</sup> the US Preventive Services Task Force<sup>5</sup> and the UK National Institute of Clinical Excellence.<sup>6</sup> The potential for these instruments to improve the ability of nonspecialists to recognize and manage depression is substantial but cannot be assumed. Even simple quality-improvement strategies must be supported by evidence of clinical benefit.<sup>7</sup>

Previous systematic reviews in this area have produced seemingly conflicting results. One review of the use of screening without further enhancement of care reported no overall benefit;<sup>8</sup> however, there were substantial differences between the included studies and some suggestion that isolated screening may be effective in certain circumstances. A review by the US Preventive Services Task Force supported screening<sup>9</sup> and recommended its use in conjunction with additional enhancement of care. Recommendations by the UK National Institute of Clinical Excellence<sup>6</sup> stated that screening should be offered for populations at increased risk of depression, but the research evidence used to support this recommendation was not clear. A more recent review showed definitively that enhanced care for depression (collaborative care) is effective in improving the outcome of depression<sup>10</sup> but that the use of screening as a component of multifaceted quality enhancement was not a necessary condition for improved outcomes.

In many health care systems, the use of screening questionnaires in primary care without additional enhancement of care has become the most commonly used quality-improvement strategy for depression care.<sup>7</sup> We conducted a systematic review to determine the specific clinical effectiveness of screening and case-finding instruments without additional enhancement of care in improving the recognition, management and outcome of depression. In particular, we sought to distinguish between studies that evaluated screening alone from those that included other elements to improve the organization and delivery of care. We also sought to examine whether the effectiveness of screening alone might vary according to circumstance and patient population.

From the Department of Health Sciences and Hull-York Medical School (Gilbody), University of York; Department of Health Sciences (Sheldon), University of York; Academic Unit of Psychiatry (House), University of Leeds, United Kingdom.

## Methods

We conducted this review according to the methods recommended by the Cochrane Collaboration.<sup>11</sup>

### Search

We searched the following databases without language restrictions from inception to December 2007: MEDLINE; EMBASE; CINAHL; PsycLIT; EconLIT; BNI/RCN; Cochrane Database of Systematic Reviews; the Trials Register of the Cochrane Depression, Anxiety and Neurosis Group; Cochrane Library; NHS Economic Evaluations Database; and the Database of Reviews of Effectiveness. A detailed example of the strategy used to search MEDLINE is available online (Appendix 1, [www.cmaj.ca/cgi/content/full/178/8/997/DC2](http://www.cmaj.ca/cgi/content/full/178/8/997/DC2)).

### Study selection

We included randomized controlled trials that investigated screening and case-finding instruments among patients in nonpsychiatric settings (e.g., general hospital or primary care) and that compared the introduction of a routine form of screening or case-finding instrument with usual or routine care. The active intervention involved the addition of standardized depression screening or an outcome-assessment instrument to routine care with information from the assessment given to the clinician. We excluded studies with substantial enhancements in the process of care<sup>12,13</sup> (e.g., case managers, nursing interventions, collaborative care) because these interventions consist of enhanced care of which screening is only 1 element.<sup>12</sup>

### Outcomes

The outcomes of interest were the rates of detection or recognition of depression by the clinician indicated by a clear entry in the medical record and the rates of intervention for depression, such as initiation of pharmacological or psychosocial intervention or active referral to specialist care for depression. We also included outcomes of depression, which were classified as short term (< 6 months), medium term (6–12 months) and long term (> 12 months).

### Quality assessment

All included studies were scrutinized independently by 2 researchers (S.G. with either A.H. or T.S.). We assessed the methodologic quality of the included studies with reference to the method of randomization and allocation concealment.<sup>11</sup> In addition, we examined clustering and “unit-of-analysis error” for studies that randomized clinicians or practices rather than individual patients.<sup>14</sup> Study inclusion or exclusion, quality assessment and data extraction were first performed by 1 reviewer (S.G.) and independently checked by a second reviewer (A.H. or T.S.). Differences of opinion were resolved by a third reviewer.

### Data synthesis

We performed a meta-analysis<sup>15</sup> of risk ratios for dichotomous variables (recognition and treatment of depression) using a random-effects analysis. To analyze outcomes of de-

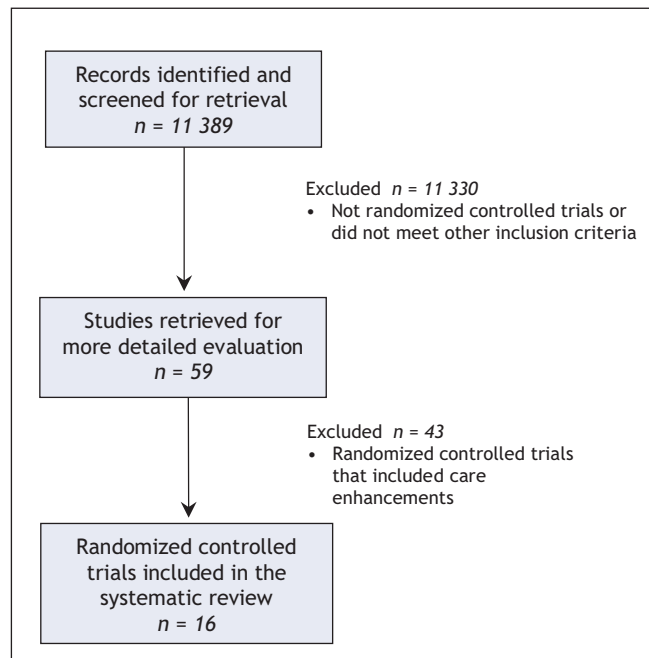
pression, we calculated standardized and weighted mean differences for continuous variables, and we transformed dichotomous variables to a standardized mean difference.<sup>16</sup> We assessed between-study heterogeneity using the  $I^2$  statistic.<sup>17</sup> As a guide, an  $I^2$  value of 25% is considered low, 50% is considered moderate and 75% is considered high for the likelihood of heterogeneity.<sup>17</sup>

If substantial between-study heterogeneity was found ( $I^2 > 50\%$ ), we examined the following likely sources of clinical heterogeneity that had been defined a priori: clinical setting (general hospital v. primary care); patient population (studies that randomized all patients irrespective of their score on a depression questionnaire [unselected patients] v. those that randomized only preselected patients who exceeded a prestated cut-off point on a questionnaire [high-risk patients]); type of instrument (depression-specific measures [e.g., Beck Depression Inventory]<sup>18</sup> v. mixed anxiety and depression measures [e.g., General Health Questionnaire]<sup>19</sup> or depression-specific measures embedded within a range of psychiatric subscales [e.g., Primary Care Evaluation of Mental Disorders]).<sup>20</sup>

For studies with substantial heterogeneity, we explored the causes using meta-regression.<sup>21</sup> The amount of heterogeneity explained by these a priori causes was examined by reductions in the  $I^2$  inconsistency statistic.

Studies that randomized by cluster (clinician or practice) but failed to incorporate this clustering into their analysis (unit-of-analysis error) were re-analyzed by use of an appropriate method<sup>14,22</sup> (incorporating an intra-class correlation coefficient of 0.02).<sup>23</sup>

We performed all analyses using STATA (version 8) with the “metan” and “metareg” series of commands.



**Figure 1:** Quality of Reporting of Meta-analysis (QUOROM) flow chart showing the number of studies screened and included in the meta-analysis.

## Results

### Search results

Of the 11 389 records identified, we selected 59 for further scrutiny. Of these, 16 met our full inclusion criteria (Figure 1; Appendix 2, available online at [www.cmaj.ca/cgi/content/full/178/8/997/DC2](http://www.cmaj.ca/cgi/content/full/178/8/997/DC2)).<sup>24-39</sup> The majority of studies were excluded because they involved a substantial enhancement of care (over and above screening) or were nonrandomized. Appendix 3 presents the excluded studies and the reasons for exclusion (available online at [www.cmaj.ca/cgi/content/full/178/8/997/DC2](http://www.cmaj.ca/cgi/content/full/178/8/997/DC2)).

### Study characteristics

Of the included studies, 12 were conducted in primary care settings,<sup>24-27,29-31,33,34,37-39</sup> 2 in general hospital outpatient settings,<sup>28,32</sup> 1 in an emergency department<sup>35</sup> and 1 in an inpatient setting with elderly patients.<sup>36</sup> The sample sizes were between 51 and 2209, and 4 studies<sup>26,27,36,38</sup> included a power calculation. Twelve of the studies were performed in the United States.

We identified 2 distinct populations of randomized patients: 9 studies included unselected patients,<sup>26-30,32,35,36,38</sup> and 7 studies included high-risk patients.<sup>24,25,31,33,34,37,39</sup> Two studies included a greater proportion of elderly patients or were specifically targeted to elderly patients.<sup>33,37</sup>

The interventions evaluated in these studies involved the feedback of 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. The control condition was generally the same case-finding instrument administered to the patient but without the results being given to the clinician. Nine studies used depression-specific instruments<sup>24,25,27,32-34,37-39</sup> and 6 used less specific measures in which depression was one of a number of psychiatric disorders covered.<sup>26,28-31,35,36</sup>

### Assessment of methodologic quality

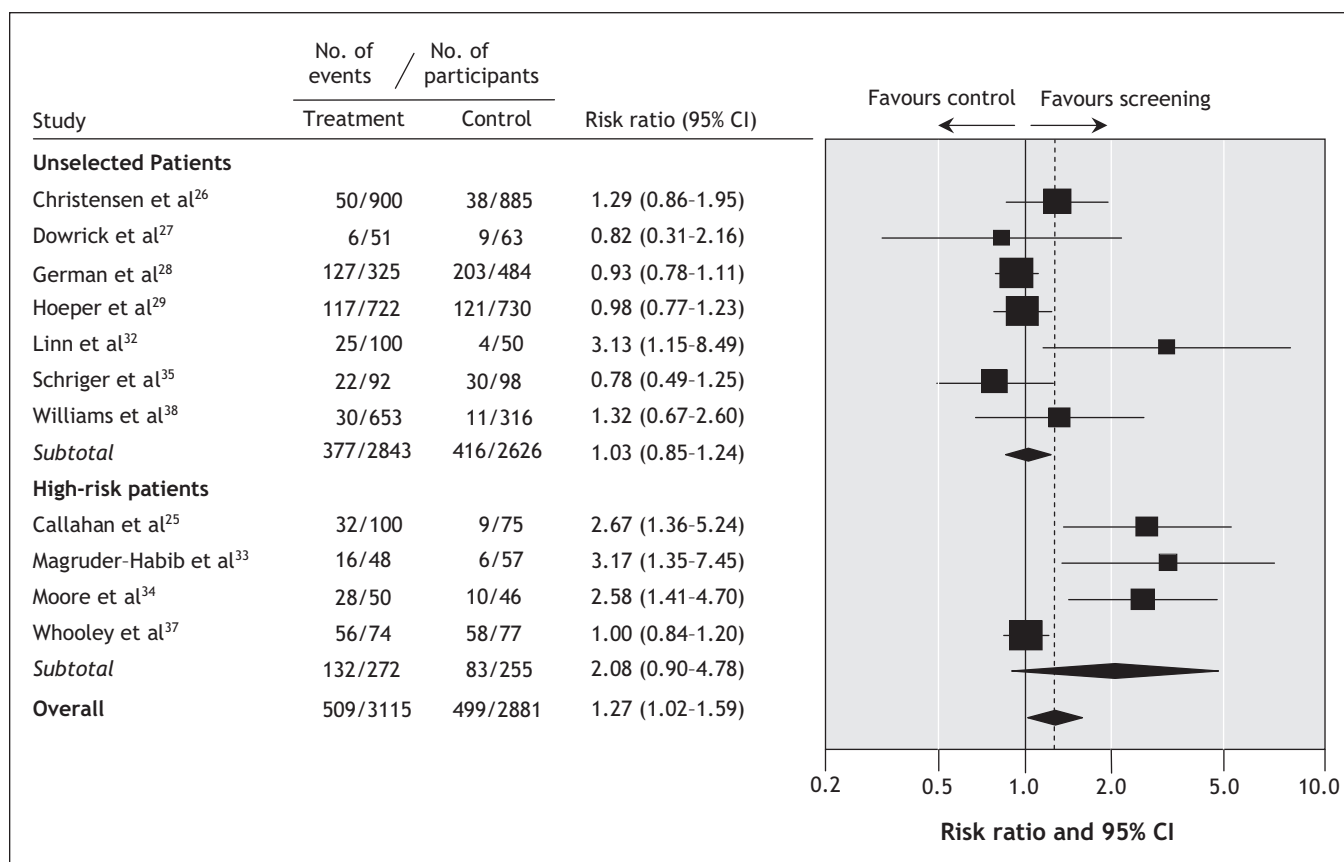
All of the included studies were described as randomized; however, few studies gave specific details of either the method of randomization or concealment of allocation. One study used a clustered design<sup>37</sup> but failed to account for clustering in the analysis of results, making it susceptible to unit-of-analysis error.<sup>40</sup>

### Data synthesis

#### Recognition of depression or anxiety

Eleven studies presented data on the effect of screening or case-finding instruments on the recognition of depression.<sup>25-29,32-35,37,38</sup> The rates of recognition were largely established by researchers scrutinizing medical records to see if the physician had made an entry about depression (Appendix 2, available online at [www.cmaj.ca/cgi/content/full/178/8/997/DC2](http://www.cmaj.ca/cgi/content/full/178/8/997/DC2)).

Visual inspection of the forest plots (Figure 2) and statisti-



**Figure 2:** Effect of screening and case-finding instruments on the recognition of depression by clinicians, by method of patient selection.

cal testing demonstrated moderate heterogeneity between studies ( $I^2 = 69\%$ ). We performed a random-effects pooling of the results, which showed that screening and case-finding instruments had a borderline positive impact on the rate of recognition of depression by clinicians (11 studies; relative risk [RR] 1.27, 95% CI 1.02 to 1.59). Exploration of the possible sources of heterogeneity showed that the most plausible explanation was the method of scoring and patient randomization (Figure 2). Selection and randomization of patients according to pre-existing scores above a cut-off value (high-risk patients) produced a larger effect size (ratio of RR 1.67, 95% CI 0.89 to 3.16) (Table 1). For unselected patients, screening and case finding instruments had no effect on depression recognition (RR 1.03, 95% CI 0.85 to 1.24).

We found that the size of the effect reported by studies that used depression-specific rating scales was larger than that reported by studies that used more broadly defined or mixed measures of depression or anxiety, although this was of borderline significance (ratio of RR 0.59, 95% CI 0.33 to 1.04,  $p = 0.06$ ). The overall effect size was the same in general hos-

pital (RR 1.38, 95% CI 0.79 to 2.43) and primary care settings (RR 1.30, 95% CI 0.99 to 1.70,  $p = 0.89$ ).

### Management of depression

Ten studies presented data on the impact of screening or case finding on the management of depression.<sup>24,25,27,28,31-33,35-37</sup> Among these studies, there was a borderline significant difference between the intervention and control groups for any intervention for depression (RR 1.30, 95% CI 0.97 to 1.76,  $I^2 = 81\%$ ) (Figure 3). Dividing the studies according to the method of selection did not reduce the overall level of between-study heterogeneity (from 81% to 61%), and this was a significant predictor of variation. Studies that randomized high-risk patients showed a larger effect size than studies that randomized unselected patients (meta-regression ratio of RR 1.37, 95% CI 0.64 to 2.94), although this difference was not significant ( $p = 0.37$ ) (Table 1).

There was no significant effect of screening on the prescription of antidepressants (RR 1.20, 95% CI 0.87 to 1.66,  $I^2 = 83\%$ ). There was a larger effect size among studies that randomized high-risk patients compared with those that ran-

**Table 1:** Random effects meta-analysis and meta-regression to explore clinical heterogeneity of the use of depression screening and case-finding questionnaires

Outcome (predictor variable)	Overall random effects meta-analysis; RR (95% CI)*	Random-effects meta-regression (95% CI)		Ratio of risk ratios (95% CI)†	p value‡	I <sup>2</sup> statistic,§ %
		Comparator	Reference			
<b>Recognition of depression</b>	1.27 (1.02-1.59)					69
Population (high risk v. unselected)		RR <sub>high risk</sub> 2.08 (0.90-4.78)	RR <sub>unselected</sub> 1.03 (0.85-1.24)	1.67 (0.89-3.16)	0.10	69
Instrument (nonspecific v. depression specific)		RR <sub>nonspecific</sub> 0.96 (0.85-1.09)	RR <sub>depression-specific</sub> 1.78 (1.12-2.84)	0.59 (0.33-1.04)	0.06	67
Setting (general hospital v. primary care)		RR <sub>general hospital</sub> 1.38 (0.79-2.43)	RR <sub>primary care</sub> 1.30 (0.99-1.70)	0.95 (0.45-2.02)	0.89	71
<b>Any intervention for depression</b>	1.30 (0.97-1.76)					81
Population (high risk v. unselected)		RR <sub>high risk</sub> 1.50 (0.89-2.53)	RR <sub>unselected</sub> 0.97 (0.81-1.18)	1.37 (0.64-2.94)	0.03	82
Instrument (nonspecific v. depression specific)		RR <sub>nonspecific</sub> 1.08 (0.89-1.32)	RR <sub>depression-specific</sub> 1.56 (0.87-2.79)	0.67 (0.30-1.50)	0.29	83
Setting (general hospital v. primary care)		RR <sub>general hospital</sub> 1.11 (0.84-1.46)	RR <sub>primary care</sub> 1.43 (0.91-2.23)	0.83 (0.39-1.75)	0.58	83
<b>Prescription of antidepressants</b>	1.20 (0.87-1.66)					83
Population (high risk v. unselected)		RR <sub>high risk</sub> 1.45 (0.91-2.31)	RR <sub>unselected</sub> 0.86 (0.64-1.15)	1.64 (0.55-4.89)	0.29	85
Instrument (nonspecific v. depression specific)		RR <sub>nonspecific</sub> 1.07 (0.79-1.45)	RR <sub>depression-specific</sub> 1.39 (0.65-2.99)	0.76 (0.24-2.45)	0.58	84
Setting (general hospital v. primary care)		RR <sub>general hospital</sub> 0.87 (0.63-1.19)	RR <sub>primary care</sub> 1.32 (0.87-2.01)	0.69 (0.19-2.52)	0.49	86
<b>Outcome of depression¶</b>	Standardized mean difference -0.02 (-0.25-0.20)					31

Note: CI = confidence interval.

\*Unless stated otherwise.

†Taken from random effects logistic meta-regression beta-coefficient of Log<sub>e</sub>RR.

‡Calculated using 1000 Monte Carlo simulations.

§I<sup>2</sup> statistic from overall random effects meta-analysis and from introduction of a predictive covariate using meta-regression (reduction from overall I<sup>2</sup> statistic suggests predictive covariate explains between-study heterogeneity).

¶No meta-regression performed.

domized unselected patients, although this was nonsignificant (meta-regression ratio of RR 1.64, 95% CI 0.55 to 4.89,  $p = 0.29$ ).

The use of a depression-specific instrument was not related to the rate of intervention for depression ( $p = 0.29$ ) or to the prescription of antidepressants ( $p = 0.29$ ). The effect of the intervention was also unrelated to study setting, with feedback being equally ineffective in primary care and general hospital settings ( $p = 0.58$ ).

### Effect of screening or case-finding on depression outcomes

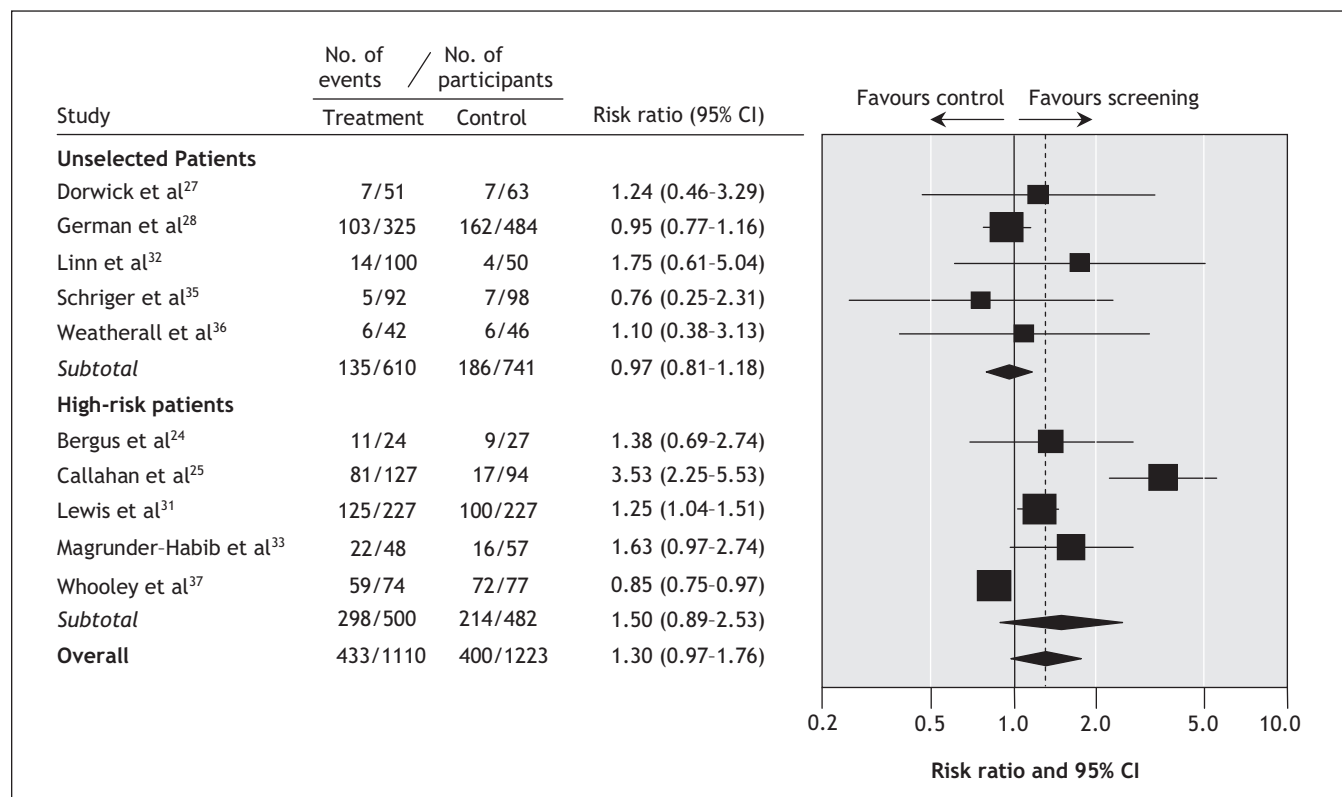
Seven studies<sup>24,25,27,30,31,37,38</sup> reported data on the impact of screening or case finding on the outcome of depression over time. Of these studies, 5 provided sufficient data to be pooled<sup>24,25,31,37,38</sup> (Appendix 2, available online at [www.cmaj.ca/cgi/content/full/178/8/997/DC2](http://www.cmaj.ca/cgi/content/full/178/8/997/DC2)). There was no overall impact of screening on depression outcomes (standardized mean difference  $-0.02$ , 95% CI  $-0.25$  to  $0.20$ ,  $I^2 = 31\%$ ) (Figure 4). There was low between-study heterogeneity, which we did not explore.

### Interpretation

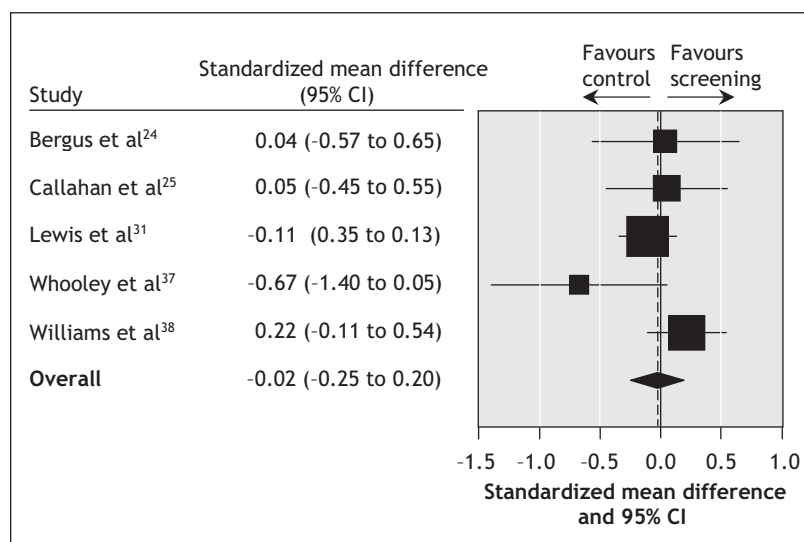
We found no substantial effect of screening or case-finding instruments on the overall recognition rates of depression, the management of depression by clinicians or on depression outcomes. These findings were true for both primary care and general hospital settings.

The finding that routinely administered screening or case-finding instruments for depression have little impact on the recognition of depression is a robust finding based on several large-scale studies. In a subset of studies that used the more complex 2-stage screening and feedback methods, there was some evidence of improved recognition. A further finding from our exploration of between-study heterogeneity is that depression-specific instruments seem to influence clinicians to a greater extent than less specific instruments, such as the General Health Questionnaire, that measure both depression and anxiety. It would seem that when information is specific and requires little additional computation on the part of the clinician, they may more readily integrate this information into their clinical decision-making process. Our finding that high-risk screening strategies might be more effective than unselected strategies might reflect an implicit decision-making process among clinicians, whereby they are more likely to act on the basis of information when there is strong positive likelihood that the information predicts the presence of a disorder. Among previously unselected patients, the prevalence of depression will be low ( $< 10\%$ ) and the post-test probability will be less than 50%, meaning that a positive screening test will be wrong more often than it is right. In this sense, our findings may reflect the Bayesian processes inherent in many clinical decisions.<sup>41</sup> These are areas that deserve further research and may point the way to finding an effective role for screening instruments in nonspecialist settings.

Despite our best efforts in summarizing these data, there are several limitations largely related to the primary studies in-



**Figure 3:** Effect of screening and case-finding instruments on the management of depression by clinicians, by method of patient selection.



**Figure 4:** Effect of screening and case-finding instruments on the outcome of depression at follow-up.

cluded. First, most studies did not report adequate concealment of allocation or the method of randomization; thus, we could not determine the susceptibility to bias.<sup>42</sup> Second, we were unable to account for some of the substantial heterogeneity that remained between studies. Further research should seek to identify other sources of clinical heterogeneity. Lastly, we should urge caution in drawing firm conclusions from any suggestive findings based upon our exploratory meta-regression analysis, since this involves making observational comparisons within randomized studies and the power of causal inference is therefore reduced.<sup>21</sup> These results should be considered hypothesis generating and further randomized trials are needed to test the robustness of these findings.

Previously, researchers have sought to apply systematic review methods to establish the effectiveness of screening for depression,<sup>8,9</sup> but with seemingly contradictory results.<sup>43</sup> The results presented in this review should be considered alongside those of a 2002 report by the US Preventive Services Task Force<sup>9</sup> and an updated report by the Canadian Task Force on Preventive Health Care based on the US review.<sup>4</sup> An Australian “review of reviews” by Hickie and colleagues in 2002 placed great emphasis on the results of the review by the US Preventive Services Task Force, but they did not include subsequent research or reviews included in our review.<sup>43</sup> Thus, to explore the reasons for this apparent divergence in results, we need to compare our methods and results with those of the US Preventive Services Task Force review.<sup>9</sup> We believe our findings are largely consistent with the reviews by the US Preventive Services Task Force and the Canadian Task Force on Preventive Health Care; however, the following differences should be noted. First, our review updates the reviews by the US Preventive Services Task Force and the Canadian Task Force on Preventive Health Care, and it includes 3 studies that were published after the other 2 reviews<sup>24,26,35</sup> and 3 studies that were not included in the US Preventive Services Task Force review.<sup>28,29,36</sup> Second, although the results were broadly

similar in both settings, our review focuses on screening in any setting, compared with the reviews by the US Preventive Services Task Force and Canadian Task Force on Preventive Health Care, which focused on primary care alone. Third, we excluded 1 study that had been included in both the US Preventive Services Task Force and Canadian Task Force on Preventive Health Care reviews because it did not meet our inclusion criteria.<sup>44</sup>

The most notable difference is that our review focuses on screening strategies alone and does not include studies in which screening was embedded within wider enhanced-care programs. The results of our review are, therefore, only relevant to stand-alone screening programs, for which we found clear evidence of limited or no benefit. The set of interventions reviewed by US Preventive Services Task Force and Canadian Task Force on Preventive Health Care included those in which screening was included as a part of enhanced patient care and clinician support (collaborative care and quality-improvement strategies).<sup>45–47</sup> Of particular importance was the inclusion of a large US study, the Partners in Care study,<sup>47</sup> whose results were strongly in favour of the active intervention (screening with collaborative care), which included face-to-face clinician education; computerized decision support; individualized treatment algorithms; psychotherapy or drug treatment; active support by a case manager; and regular consultation with a specialized mental health clinician (psychologist or psychiatrist). This study accounted for between 30% and 47% of the weighted information in the meta-analyses produced by the US Preventive Services Task Force.<sup>9</sup>

Complex enhanced collaborative care for depression improves the outcomes of depression, and these packages have been comprehensively reviewed elsewhere.<sup>10,13,48</sup> We do not question the effectiveness of these strategies, but it remains unclear whether screening is a necessary component of enhanced collaborative care for depression. From a recent review of the necessary components of collaborative care,<sup>10</sup> several other factors emerged as potentially necessary (and statistically significant) components including the use of coordinated patient follow-up, case managers with a mental health background and regular supervision of case managers.<sup>10</sup> Thus, the previous reviews by the US Preventive Services Task Force and Canadian Task Force on Preventive Health Care, which mixed studies of enhanced care (some including screening) and screening alone, do not provide reliable evidence on the effectiveness of screening. Many studies that include complex enhancements of care have not used screening as a recruitment strategy, but they have also reported positive results.<sup>49–51</sup> Further trials in this area should compare the relative effectiveness of enhancements of care with and without screening.

Our review complements those by the US Preventive Services Task Force and Canadian Task Force on Preventive Health Care, and it enhances our understanding of prior work and the available and more recent evidence. It helps us to better understand the role of screening in general and confirms

that screening without other systematic changes to improve depression management is unlikely to improve outcomes. This is of particular importance to policy makers who may have ignored the key recommendations of the US Preventive Services Task Force and Canadian Task Force on Preventive Health Care in the provision of additional management strategies for depression. In practice, screening strategies have often either been recommended for populations at high risk for depression, such as those with chronic illness, or adopted alone and without further enhancements of care.<sup>7</sup>

This article has been peer reviewed.

**Competing interests:** None declared.

**Contributors:** Simon Gilbody was the lead researcher and study guarantor. He, along with Trevor Sheldon and Allan House, contributed substantially to the study conception and design, the acquisition of data, and the analysis and interpretation of data. Each of the authors contributed to the drafting and revision of the manuscript and gave final approval of the version to be published.

**Acknowledgements:** An earlier version of this review was published in the Cochrane Library in 2005 [Gilbody S, House AO, Sheldon TA. Screening and case finding instruments for depression. *Cochrane Database Syst Rev* 2005;(4):CD002792]. The current paper represents an updated version that incorporates more recent trials, analysis and interpretation.

## REFERENCES

- Agency for Health Care Policy Research. *Depression in primary care*. Washington: US Department of Health and Human Services, 1993.
- Simon GE, Von Korff M. Recognition and management of depression in primary care. *Arch Fam Med* 1995;4:99-105.
- Katon W, Ciechanowski P. Impact of major depression on chronic medical illness. *J Psychosom Res* 2002;53:859-63.
- MacMillan HL, Patterson CJS, Wathen CN; The Canadian Task Force on Preventive Health Care. Screening for depression in primary care: recommendation statement from the Canadian Task Force on Preventive Health Care. *CMAJ* 2005;172:33-5.
- Agency for Healthcare Research and Quality. *Screening for Depression: Systematic Evidence Review Number 6*. Rockville (MD): The Agency; 2002.
- National Institute for Clinical Excellence. *Depression: core interventions in the management of depression in primary and secondary care*. London (UK): Her Majesty's Stationary Office; 2004.
- Gilbody S, Sheldon T, Wessely S. Should we screen for depression? *BMJ* 2006;332:1027-30.
- Gilbody SM, House AO, Sheldon TA. Routinely administered questionnaires for depression and anxiety: a systematic review. *BMJ* 2001;322:406-9.
- Pignone MP, Gaynes BN, Rushton JL, et al. Screening for depression in adults: a summary of the evidence for the US Preventive Services Task Force. *Ann Intern Med* 2002;136:765-76.
- Gilbody S, Bower P, Fletcher J, et al. Collaborative care for depression: a cumulative meta-analysis and review of longer-term outcomes. *Arch Intern Med* 2006;166:2314-21.
- Higgins JPT, Green S. *Cochrane handbook for systematic reviews of interventions 4.2.4 [updated March 2005]*. Chichester (UK): John Wiley and Sons; 2005.
- Von Korff M. Can care management enhance integration of primary and specialty care? *BMJ* 2004;329:605.
- Gilbody S, Whitty P, Grimshaw J, et al. Educational and organizational interventions to improve the management of depression in primary care: a systematic review. *JAMA* 2003;289:3145-51.
- Donner A, Klar N. Issues in the meta-analysis of cluster randomized trials. *Stat Med* 2002;21:2971-80.
- DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.
- Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med* 2000;19:3127-31.
- Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557-60.
- Beck AT, Steer RA. *The Beck depression inventory manual*. San Antonio (TX): The Psychological Corporation, Harcourt Brace Janovich; 1987.
- Goldberg DP, Williams P. *The user's guide to the General Health Questionnaire*. Windsor: NFER-Nelson; 1988.
- Spitzer RL, Williams JB, Kroenke K, et al. Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. *Am J Obstet Gynecol* 2000;183:759-69.
- Higgins JPT, Thompson SG. Controlling the risk of spurious findings from meta-regression. *Stat Med* 2004;23:1663-82.
- Rao JNK, Scott AJ. A simple method for the analysis of clustered binary data. *Biometrics* 1992;48:577-85.
- Adams G, Gulliford MC, Ukoumunne OC, et al. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *J Clin Epidemiol* 2004;57:785-94.
- Bergus GR, Hartz AJ, Noyes RJ, et al. The limited effect of screening for depressive symptoms with the PHQ-9 in rural family practices. *J Rural Health* 2005;21:303-9.
- Callahan CM, Dittus RS, Tierney WM. Primary care physicians' medical decision making for late life depression. *J Gen Intern Med* 1996;11:218-9.
- Christensen KS, Toft T, Frostholt L, et al. The FIP study: a randomised, controlled trial of screening and recognition of psychiatric disorders. *Br J Gen Pract* 2003;53:758-63.
- 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 PS, Shapiro S, Skinner EA. Detection and management of mental health problems of older patients by primary care providers. *JAMA* 1987;257:489-96.
- Hoepfer EW, Nycz GR, Kessler JD, et al. The usefulness of screening for mental illness. *Lancet* 1984;1:33-5.
- Johnstone A, Goldberg D. Psychiatric screening in general practice. *Lancet* 1976;1:605-12.
- Lewis G, Sharp D, Bartholomew J, et al. Computerized assessment of common mental disorders in primary care: effect on clinical outcome. *Fam Pract* 1996;13:120-6.
- Linn LS, Yager J. The effect of screening, sensitisation and feedback on notation of depression. *J Med Educ* 1980;20:942-53.
- 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.
- Moore JT, Silimperi DR, Bobula JA. Recognition of depression by family medicine residents: the impact of screening. *J Fam Pract* 1978;7:509-13.
- Schriger DL, Gibbons PS, Langone CA, et al. Enabling the diagnosis of occult psychiatric illness in the emergency department: a randomized, controlled trial of the computerized, self-administered PRIME-MD diagnostic system. *Ann Emerg Med* 2001;37:132-40.
- Weatherall M. A randomized controlled trial of the Geriatric Depression Scale in an inpatient ward for older adults. *Clin Rehabil* 2000;14:186-91.
- Whooley MA, Stone B, Soghikian K. Randomized trial of case-finding for depression in elderly primary care patients. *J Gen Intern Med* 2000;15:293-300.
- Williams JW, Mulrow CD, Kroenke K. Case-finding for depression in primary care: a randomized trial. *Am J Med* 1999;106:36-43.
- Zung WW, Magill M, Moore JT, et al. Recognition and treatment of depression in a family medicine practice. *J Clin Psychiatry* 1983;44:3-6.
- Divine GW, Brown JT, Frazer LM. The unit of analysis error in studies about physicians' patient care behavior. *J Gen Intern Med* 1992;7:623-9.
- Gill CJ, Sabin L, Schmid CH. Why clinicians are natural Bayesians. *BMJ* 2005;330:1080-3.
- Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.
- Hickie IB, Davenport TA, Ricci CS. Screening for depression in general practice and related medical settings. *Med J Aust* 2002;177(Suppl):S111-6.
- Reifler DR, Kessler HS, Bernhard EJ, et al. Impact of screening for mental health concerns on health service utilization and functional status in primary care patients. *Arch Intern Med* 1996;156:2593-9.
- Katzelnick DJ, Simon GE, Pearson SD, et al. Randomized trial of a depression management program in high utilizers of medical care. *Arch Fam Med* 2000;9:345-51.
- Rost K, Nutting PA, Smith J, et al. Improving depression outcomes in community primary care practice: a randomised trial of the QuEST intervention. *J Gen Intern Med* 2001;16:143-9.
- Wells KB, Sherbourne C, Schoenbaum M, et al. Impact of disseminating quality improvement programmes for depression in managed primary care: a randomized controlled trial. *JAMA* 2000;283:212-20.
- Gilbody S, Whitty P, Grimshaw JG, et al. Improving the recognition and management of depression in primary care. *Effective Health Care*. York (UK): The Royal Society of Medicine Press; 2002.
- Datto CJ, Thompson R, Horowitz D, et al. The pilot study of a telephone disease management program for depression. *Gen Hosp Psychiatry* 2003;25:169-77.
- Dietrich AJ, Oxman TE, Williams JW, et al. Re-engineering systems for the treatment of depression in primary care: cluster randomised controlled trial. *BMJ* 2004;329:602-9.
- Hunkeler EM, Meresman JF, Hargreaves WA, et al. Efficacy of nurse telehealth care and peer support in augmenting treatment of depression in primary care. *Arch Fam Med* 2000;9:700-8.

**Correspondence to:** Dr. Simon Gilbody, Professor of Psychological Medicine and Health Services Research, Department of Health Sciences, Hull-York Medical School, University of York, YO10 5DD UK; fax 44 1904 321320; sg519@york.ac.uk