Screening for Depression in Adults: A Summary of the Evidence for the U.S. Preventive Services Task Force

Michael P. Pignone, MD, MPH; Bradley N. Gaynes, MD, MPH; Jerry L. Rushton, MD, MPH; Catherine Mills Burchell, MA; C. Tracy Orleans, PhD; Cynthia D. Mulrow, MD, MSc; and Kathleen N. Lohr, PhD

Purpose: To clarify whether screening adults for depression in primary care settings improves recognition, treatment, and clinical outcomes.

Data Sources: The MEDLINE database was searched from 1994 through August 2001. Other relevant articles were located through other systematic reviews; focused searches of MEDLINE from 1966 to 1994; the Cochrane depression, anxiety, and neurosis database; hand searches of bibliographies; and extensive peer review.

Study Selection: The researchers reviewed randomized trials conducted in primary care settings that examined the effect of screening for depression on identification, treatment, or health outcomes, including trials that tested integrated, systematic support for treatment after identification of depression.

Data Extraction: A single reviewer abstracted the relevant data from the included articles. A second reviewer checked the accuracy of the tables against the original articles.

Data Synthesis: Compared with usual care, feedback of depression screening results to providers generally increased recognition of depressive illness in adults. Studies examining the effect of screening and feedback on treatment rates and clinical outcomes had mixed results. Many trials lacked power to detect clinically important differences in outcomes. Meta-analysis suggests that overall, screening and feedback reduced the risk for persistent depression (summary relative risk, 0.87 [95% CI, 0.79 to 0.95]). Programs that integrated interventions aimed at improving recognition and treatment of patients with depression and that incorporated quality improvements in clinic systems had stronger effects than programs of feedback alone.

Conclusion: Compared with usual care, screening for depression can improve outcomes, particularly when screening is coupled with system changes that help ensure adequate treatment and follow-up.

Depressive disorders are common, chronic, and costly. Prevalence rates from community-based surveys range from 1.8% to 3.3% for depression within the past month and 4.9% to 17.1% for lifetime prevalence (1, 2). In primary care settings, the point prevalence of major depression ranges from 4.8% to 8.6% (3). Depressive illness is projected to be the second leading cause of disability worldwide in 2020 (4). The substantial public health and economic significance of depression is reflected by its considerable effect on health care utilization and great monetary costs: $43 billion annually, of which $17 billion represents lost work days (5, 6).

Despite the high prevalence and substantial impact of depression, detection and treatment in the primary care setting have been suboptimal. Studies have shown that usual care by primary care physicians fails to recognize 30% to 50% of depressed patients (7). Because patients in whom depression goes unrecognized cannot be appropriately treated, systematic screening has been advocated as a means of improving detection, treatment, and outcomes of depression.

In 1996, the U.S. Preventive Services Task Force found insufficient evidence to recommend for or against routine screening for depression with standardized questionnaires (8). The Task Force recommended that clinicians "maintain an especially high index of suspicion for depressive symptoms in adolescents and young adults, persons with a family or personal history of depression, those with chronic illnesses, those who perceive or have experienced a recent loss, and those with sleep disorders, chronic pain, or multiple unexplained somatic complaints" (8). It also recommended physician education in recognizing and treating depression.

To help determine whether systematic, routine screening for depression in adults is warranted, we performed an updated systematic review for the U.S. Preventive Services Task Force. Specifically, we examined three key questions: 1) What is the accuracy of case-finding instruments for depression in primary care populations? 2) Is treatment of depression in primary care patients effective in improving outcomes? 3) Is routine systematic identification with case-finding questions

Table 1. Characteristics of Case-Finding Instruments Used To Detect Depression in Adults in Primary Care Settings*

Instrument	Items, n†	Time Frame of Questions	Score Range	Usual Cut-Point‡	Literacy Level§	Administration Time, min
Beck Depression Inventory	21	Today	0–63	Mild, 10; moderate, 20; severe, 30	Easy	2–5
Center for Epidemiologic Study Depression Screen	20	Past week	0–60	16	Easy	2–5
General Health Questionnaire	28	Past few weeks	0–28	4	Easy	5–10
Medical Outcomes Study Depression Screen	8	Past week	0–1	0.06	Average	<2
Primary Care Evaluation of Mental Disorders	2	Past month	0–2	1	Average	<2
Symptom-Driven Diagnostic System–Primary Care	5	Past month	0–4	2	Easy	<2
Zung Self-Depression Scale	20	Recently	25–100	Mild, 50; moderate, 60; severe, 70	Easy	2–5

* Adapted from reference 10.
† Item numbers for the Primary Care Evaluation of Mental Disorders and the Symptom-Driven Diagnostic System–Primary Care refer to depression questions only. Several instruments now have shortened versions as well.
‡ The cut-point is the number at or above which the test is considered positive.
§ "Easy" is a third- to fifth-grade reading level and "average" is a sixth- to ninth-grade reading level, according to the fog formula.

(screening), with or without integrated management and follow-up systems, more effective than usual care in identifying patients with depression, facilitating treatment of patients with depression, and improving clinical outcomes?

The results of the comprehensive review are available from the Agency for Healthcare Research and Quality (www.ahrq.gov/clinic/uspstfix.htm) (9). In brief, we found that several short, accurate, and easy-to-use instruments for detecting depression are available (9, 10) (**Table 1**). Brief instruments, including asking the patient two questions about the presence of depressed mood and anhedonia ("Over the past 2 weeks, have you felt down, depressed or hopeless?" and "Over the past 2 weeks, have you felt little interest or pleasure in doing things?"), appear to perform as well as longer instruments. Effective treatments, including pharmacologic and behavioral or counseling interventions, are available for depressed patients identified in primary care settings (9).

We also examined the evidence on whether screening for depression in primary care settings affects recognition, treatment, and clinical outcomes of adult patients with depression. In this article, we review the evidence pertaining to this overarching question.

METHODS

To identify relevant articles, we searched the MEDLINE database from January 1994 through August 2001 by using the Medical Subject Headings *depression* or *depressive disorders,* plus keyword searches for commonly used screening instruments. These terms were then combined with the Medical Subject Headings *mass*

screening or *sensitivity and specificity* or *primary health care* or *ambulatory care* or *family practice.* We supplemented these sources by searching the Cochrane database on depression, neurosis, and anxiety disorders; performing additional specific MEDLINE searches from 1966 to 1994; hand searching bibliographies; and querying experts.

We reviewed randomized trials conducted in primary care settings that examined the effect of screening for depression on identification, treatment, or health outcomes, including trials that tested integrated, systematic support for treatment after identification of depression.

Two of the authors independently reviewed the titles and abstracts of the articles identified by the literature searches and excluded ones on which they agreed that eligibility criteria were not met. When the initial reviewers disagreed, the articles were carried forward to the next review stage in which the investigators reviewed the full articles and made a final decision about inclusion or exclusion by consensus.

One reviewer abstracted the relevant information from each article into evidence tables. A second author checked these tables and noted discrepancies, which were then resolved by consensus. We calculated absolute differences in outcomes and 95% CIs by using Stata software, version 6.0 (Stata Corp., College Station, Texas), when these results were not presented in the original articles.

To summarize the effect of screening on clinical outcomes, we performed meta-analysis by using RevMan software (Cochrane Collaboration, 2000) and the DerSimonian and Laird random-effects model.

This evidence report was funded through a contract to the Research Triangle Institute–University of North Carolina Evidence-based Practice Center from the Agency for Healthcare Research and Quality. Staff of the funding agency contributed to the study design, reviewed draft and final manuscripts, and made editing suggestions.

RESULTS

The effect of routine screening of adult patients for depression in primary care was compared with usual care in 14 randomized trials in primary care settings (11–25). The main outcomes examined were differences in providers' rate of detection or recognition of depression, the proportion of patients with depression who were treated or referred for treatment, and clinical outcomes of depression. The screening interventions differed in intensity. Some trials provided feedback of screening results alone; others provided feedback and general or specific treatment advice to the providers; and some provided feedback and treatment advice and helped practices develop systematic means of improving the quality of treatment and follow-up. The trials, which were stratified by intensity of the intervention, are described below and summarized in **Tables 2** through **5**.

Effects of Screening and Feedback Alone

Johnstone and Goldberg applied the self-administered General Health Questionnaire to 1093 primary care patients and identified 119 with depression (14). These 119 patients were randomly assigned to immediate feedback of the results to the physician or to usual care. The groups did not differ significantly in mean General Health Questionnaire scores at 12-month follow-up, except for the subgroup of patients with severe depression, for whom feedback improved scores. Among all patients, the total amount of time spent depressed within 1 year decreased by approximately 2 months ($P < 0.01$).

Three trials evaluated feedback of Zung Self-Depression Scale scores to providers. Moore and colleagues asked 212 consecutive patients 20 to 60 years of age who attended a university-based family medicine residency clinic to self-administer the Zung Self-Depression Scale (15). The 96 patients who scored higher than 50 were randomly assigned to a group whose providers were given immediate written feedback of results or to a group whose providers received a generic note saying that their patients had been screened. The same note was affixed to the charts of patients in each group who had scored 50 or less. Recognition of depression, as assessed by chart audit, was 56% in the intervention group (28 of 50 patients) and 22% in the control group (10 of 46 patients). Prescription of treatment was not assessed.

Linn and Yager tested immediate written feedback of Zung Self-Depression Scale results compared with no screening in 74 consecutive new patients from a primary care clinic, using chart audit to assess outcomes (16). Depression was more likely to be diagnosed in patients assigned to the feedback group than in those receiving usual care (29% vs. 8%); treatment rates were low and similar in each group (13% vs. 8%). Neither Moore and colleagues nor Linn and Yager reported clinical outcomes.

Magruder-Habib and associates screened 800 Veterans Administration patients for depression in a primary care clinic (18). Research assistants administered the Zung Self-Depression Scale and used the Diagnostic Interview Schedule to confirm diagnosis according to criteria from the *Diagnostic and Statistical Manual of Mental Disorders,* third edition (DSM-III) (18, 26). The 100 patients who screened positive (excluding those with scores higher than 75 or history of depression) and met DSM-III criteria for major depression were randomly assigned to feedback of screening results or usual care; chart audit was used to assess outcomes. Patients whose physicians received feedback were three times more likely to be accurately identified as depressed at the index visit than were patients whose clinicians had not received such feedback (25% vs. 8%; difference, 17 percentage points [CI, 3 to 32 percentage points]). At 1 year of follow-up, 42% of the intervention group and 21% of the control group had been recognized as depressed. At 3 months of follow-up, more patients in the feedback group were being treated for depression, but the difference was not statistically significant (37% vs. 27%; difference, 11 percentage points [CI, −8 to 29 percentage points]). No clinical outcomes were measured (18).

Dowrick studied 116 patients who were initially rated "not depressed" by their usual general practitioners but had self-administered Beck Depression Inventory scores greater than 14 (20). The patients were randomly assigned to no feedback or feedback that was given to

Table 2. **Studies on the Effect of Screening and Feedback***

Author, Year (Reference)	Screening Instrument	Participants, *n*	Mode of Administration	Confirmatory Diagnostic Interview?	Feedback to Provider	Quality Rating†	
						Internal Validity	External Validity
Johnstone and Goldberg, 1976 (14)	GHQ	119	Self	Yes‡	Immediate feedback	Good	Fair
Moore et al., 1978 (15)	SDS	212	Self	No	Immediate written feedback	Good	Fair
Linn and Yager, 1980 (16)	SDS	150	Self	No	Immediate written feedback	Good	Good
Zung and King, 1983 (17)	SDS and immediate diagnostic interview	49	Psychiatrist	Yes‡	Immediate feedback	Fair	Poor
Magruder-Habib et al., 1990 (18)	SDS	100	Research assistant	Yes‡	Immediate written feedback	Good	Good
Callahan et al., 1994 (19) and 1996 (21)	CES-D	175, 222	Research assistant	Yes (HAM-D)‡	Feedback to schedule 3 additional visits within 3 months	Good	Fair
Dowrick, 1995 (20)	BDI	116	Self	No	Written feedback to provider 1 week after visit, plus chart note	Fair	Fair
Lewis et al., 1996 (22)	GHQ	681	Self	PROQSY group only	Immediate feedback on GHQ results; participants asked to complete PROQSY and, if positive, to schedule follow-up in 1 week	Fair	Fair
Reifler et al., 1996 (23)	SDDS	358	Self	Yes§	Providers given diagnostic worksheet at same visit for participants who screened positive	Good	Good
Williams et al., 1999 (11)	CES-D, blinded DSM-III-R	969	Self	Yes§	Immediate written feedback	Good	Good
Katzelnick et al., 2000 (12)	SCID and HAM-D	407	Telephone by research assistant	No	Immediate written feedback and additional support	Good	Good
Wells et al., 2000 (24)	Two-item instrument	1356	Research assistant	Yes (subset)§	Providers notified and asked to schedule visit within 2 weeks	Good	Fair
Whooley et al., 2000 (25)	GDS	2346	Research assistant	No	Intervention providers notified same day (before visit, 74%; after visit, 26%)	Fair	Fair
Rost et al., 2001 (13)	Sadness or anhedonia within 2 weeks	479 (189 not recently treated)	Nurse	Yes	Feedback to provider; nurse-centered follow-up weekly for 5 weeks	Fair	Good

* BDI = Beck Depression Inventory; CES-D = Center for Epidemiologic Study Depression Scale; DSM-III-R = *Diagnostic and Statistical Manual of Mental Disorders*, 3rd edition, revised; GDS = Geriatric Depression Scale; GHQ = General Health Questionnaire; HAM-D = Hamilton Depression Scale; PROQSY = self-administered computerized assessment; SCID = Structured Clinical Interview for DSM-IV; SDDS = Symptom-Driven Diagnostic System–Primary Care; SDS = Zung Self-Depression Scale.
† The definitions of the quality ratings are as follows. Good: Evidence includes consistent results from well-designed, well-conducted studies in representative populations that directly assess effects on health outcomes. Fair: Evidence is sufficient to determine effects on health outcomes, but the strength of the evidence is limited by the number, quality, or consistency of the individual studies; generalizability to routine practice; or indirect nature of the evidence on health outcomes. Poor: Evidence is insufficient to assess the effects on health outcomes because of limited number or power of studies, important flaws in their design or conduct, gaps in the chain of evidence, or lack of information on important health outcomes.
‡ Required before randomization.
§ Not related to randomization.

providers 1 week after the visit in which screening took place and noted in the chart for subsequent visits. At 1 year, rates of diagnosis and treatment of depression were higher in the intervention group than the control group, although the differences were not statistically significant. Clinical outcomes were not measured.

Reifler and coworkers studied 358 primary care patients by using the self-administered Symptom-Driven

Diagnostic System for Primary Care (23). The clinicians of intervention-group patients received results of the Symptom-Driven Diagnostic System for Primary Care; the clinicians of controls were not informed of the results. At 3 months, the research team observed no clinically or statistically significant differences in clinical outcomes, but the actual proportions of patients who were still depressed were not presented in the report.

Lewis and colleagues used the self-administered General Health Questionnaire or the General Health Questionnaire plus a computer-based diagnostic tool (PROQSY) to examine the effect of feedback to providers of positive scores on outcomes in low-income primary care patients in London (22). At 6 weeks, compared with General Health Questionnaire scores in controls, scores were improved in patients whose providers received feedback on the PROQSY results but not in those whose providers received only General Health Questionnaire results. When a General Health Questionnaire score greater than 1 was used to indicate current depression, patients who were screened with PROQSY were slightly less likely than controls to be depressed at 6 weeks (69% vs. 74%; difference, 5 percentage points [CI, −14 percentage points to 3 percentage points). At 6 months of follow-up, mean General Health Questionnaire scores did not differ between groups.

Williams and associates tested the effect of immediate provider feedback of results of the Center for Epidemiologic Study Depression Scale or a single question about depressed mood with no feedback (11). They confirmed the presence or absence of depression by using criteria from the Diagnostic Interview Schedule and DSM-III, revised (DSM-III-R) (27), but they did not use this information to determine eligibility for the trial.

Current depression was defined as meeting the DSM-III-R criteria for major depression or dysthymia or having minor depression (depressed mood or anhedonia plus one to three additional DSM-III-R symptoms). On the basis of chart reviews, current depression was recognized in 39% of patients whose providers received feedback from screening and in 29% of controls (difference, 10 percentage points [CI, −8 to 28 percentage points]). Rates of treatment were similar in each group. At 3 months, 37% of the intervention group and 46% of the control group met DSM-III-R criteria for depression (difference, −8 percentage points [CI, −21 to 4 percentage points]) (11).

Effects of Screening and Feedback with Treatment Advice

Zung and King screened 499 patients at one private physician's office by using the Self-Depression Scale screening test administered by a psychiatrist (17). Of the 60 patients who screened positive for depression, 49 had major depression according to DSM-III criteria. These 49 patients were randomly assigned to a group in which the provider received the results of screening ($n = 23$) or to usual care ($n = 26$). Patients identified as depressed were treated with alprazolam, a benzodiazepine drug that is currently not recommended for treating depression. At 4 weeks, follow-up data were available for

Table 3. Summary of the Effect of Feedback from Screening on Rates of Diagnosis*

Author, Year (Reference)	Participants with Diagnosis		Absolute Difference (95% CI)	P Value†
	Intervention Group	Control Group		
	% (n/n)		percentage points	
Johnstone and Goldberg, 1976 (14)‡	NR	NR	NR	NR
Moore et al., 1978 (15)‡	56 (28/50)	22 (10/46)	34 (16.7 to 52)	<0.001
Linn and Yager, 1980 (16)§	29 (7/24)	8 (4/50)	21 (1 to 41)	
Zung and King, 1983 (17)‖	NR	NR	NR	NR
Magruder-Habib et al., 1990 (18)‖	25 (12/48)	8 (4/52)	17 (3 to 32)	0.018
Callahan et al., 1994 (19)‖	32 (32/100)	12 (9/75)	20 (8 to 32)	0.002
Callahan et al., 1996 (21)‖	87 (111/128)	40 (38/94)	46 (35 to 58)	0.001
Dowrick, 1995 (20)‡	35 (18/51)	21 (13/63)	15 (−2 to 31)	
Lewis et al., 1996 (22)‡	NR	NR	NR	NR
Reifler et al., 1996 (23)	NR	NR	NR	NR
Williams et al., 1999 (11)‡	39 (30/77)	29 (11/38)	10 (−8 to 28)	>0.05
Katzelnick et al., 2000 (12)	NR	NR	NR	NR
Wells et al., 2000 (24)‡	NR	NR	NR	NR
Whooley et al., 2000 (25)‡	35 (56/162)	34 (58/169)	1 (−9 to 10)	>0.2
Rost et al., 2001 (13)	NR	NR	NR	NR

* All figures are rounded to nearest value. NR = not reported and cannot be calculated from available data.
† P values were not always reported.
‡ Denominator is patients who screened positive.
§ Denominator is all patients.
‖ Denominator is patients who screened positive and were confirmed to have major depression on diagnostic interview.

Table 4. **Summary of the Effect of Feedback from Screening on Rates of Treatment***

Author, Year (Reference)	Participants Treated		Absolute Difference (95% CI)	*P* Value†
	Intervention Group	Control Group		
	% (n/n)		*percentage points*	
Johnstone and Goldberg, 1976 (14)‡	NR	NR	NR	NR
Moore et al., 1978 (15)‡	NR	NR	NR	NR
Linn and Yager, 1980 (16)§	13 (3/24)	8 (4/50)	5 (−11 to 20)	
Zung and King, 1983 (17)‖	NR	NR	NR	
Magruder-Habib et al., 1990 (18)‖	3 months: 37 (18/48)	27 (14/52)	11 (−8 to 29)	>0.2
Callahan et al., 1994 (19)‖	26 (26/100)	8 (6/75)	18 (7 to 29)	0.002
Callahan et al., 1996 (21)‖	46 (58/127)	29 (27/94)	17 (4 to 30)	0.001
Dowrick, 1995 (20)‡	27 (14/51)	21 (13/63)	7 (−9 to 23)	
Lewis et al., 1996 (22)‡	NR	NR	NR	
Reifler et al., 1996 (23)‡	NR	NR	NR	
Williams et al., 1999 (11)‡	45 (35/77)	43 (16/38)	2 (NR)	>0.2
Katzelnick et al., 2000 (12)	82 (179/218)	32 (61/89)	50 (41 to 58)	<0.001
Wells et al., 2000 (24)‡	59 (NR)	50 (NR)	9 (NR)	0.006
Whooley et al., 2000 (25)‡	36 (59/162)	43 (72/169)	−6 (−17 to 4)	>0.2
Rost et al., 2001 (13)	69 (NR)	28 (NR)	41 (NR)	

* All figures are rounded to nearest percentage. NR = not reported and cannot be calculated from available data.
† *P* values were not always reported.
‡ Denominator is patients who screened positive.
§ Denominator is all patients.
‖ Denominator is patients who screened positive and were confirmed to have major depression on diagnostic interview.

21 intervention-group patients and 20 controls. The intervention patients were less likely than control patients to remain depressed after 1-month follow-up; 33% of intervention patients were still depressed versus 65% of controls when persistent depression was defined as a failure to improve by 12 or more points on the Zung Self-Depression Scale (difference, −32 percentage points [CI, −61 to −3 percentage points]).

Callahan and associates studied patients older than 60 years of age in an academic primary care setting that served low-income patients (19, 21). Research assistants initially screened potential participants by using the Center for Epidemiologic Study Depression Scale. Participants who scored 16 or higher were given the Hamilton Depression Scale. Patients who scored higher than 14 on the Hamilton Depression Scale underwent randomization by physician group, in which certain clinic sessions were randomly assigned to the intervention group and others to the control group. All physicians received an educational talk at baseline. Providers of intervention-group patients received feedback from screening plus individually targeted educational information and specific treatment recommendations. Physicians in the intervention group also were asked to schedule three specific visits for study patients to address depression.

Depression diagnoses were documented more frequently for intervention-group patients than for controls (87% vs. 40%) (21). Initiation of a treatment plan was more common among intervention patients (46% vs. 29%; difference, 17 percentage points [CI, 4 to 30 percentage points]) (19, 21). The proportion of patients who were still depressed at 6 months of follow-up (Hamilton Depression Scale >10) was 87% for intervention-group patients and 88% for controls (difference, −1 percentage point [CI, −11 to 9 percentage points]).

Whooley and colleagues studied the effect of screening with the Geriatric Depression Scale and feedback among patients older than 65 years of age in 13 practices in the Kaiser Permanente system (25). Research assistants screened patients on the day of a regularly scheduled clinic visit and gave same-day feedback (74% before visits and 26% after visits) to the providers in 7 intervention clinics; they gave no feedback to providers in 6 usual-care practices. All providers received an initial education session on management of depression. Intervention-group patients were offered a series of six weekly group educational sessions led by a nurse. Rates of recognition of depression were similar in each group, but prescription of antidepressant medication (on the basis of pharmacy database review) was higher among controls. Continued depression, defined as a Geriatric De-

pression Scale score greater than 6, was assessed 2 years after enrollment; data were available for 69% of patients. At 2 years of follow-up, 42% of intervention-group patients and 50% of controls were still depressed (difference, −8 percentage points [CI, −21 to 6 percentage points]) (25).

Effects of Integrated Interventions To Improve Recognition and Management of Depression

Wells and colleagues combined screening and a quality improvement program for depression treatment in 46 primary care clinics and measured the effect on treatment and outcomes of depression (24). Patients were enrolled if they screened positive on a two-question instrument. Patients received the Composite International Diagnostic Interview criterion standard examination, but participation was not based on its result. The investigators enrolled 1356 patients and followed them

for 12 months. Randomization was at the level of the practice, and the intervention included feedback of the results of the screening test and a request that the provider schedule a visit within 2 weeks. Intervention practices also received educational materials, assistance in treatment initiation and maintenance, and access to nurse-led medication follow-up or to cognitive–behavioral therapy.

At 12 months, the proportion of patients receiving appropriate treatment (defined as any appropriate antidepressant or at least one visit to a mental health provider) was higher in the intervention group than in the control group (59% vs. 50%; difference, 9 percentage points [CI not reported]; $P = 0.006$). On the basis of Center for Epidemiologic Study Depression score, intervention-group patients were less likely than controls to be depressed at 6 months (55% vs. 64%; difference, −9 percentage points [CI, −15 to −3 percentage points]).

Table 5. **Summary of the Effect of Feedback from Screening on Patient Outcomes***

Author, Year (Reference)	Outcome Measured	Outcome Data		Absolute Difference (95% CI)	P Value†
		Intervention Group Value	Control Group Value		
Johnstone and Goldberg, 1976 (14)‡	Mean duration of depression in 1 year	4.2 mo	6.3 mo	−2.1 mo (NR)	<0.01
Moore et al., 1978 (15)‡	NR	NR	NR	NR	
Linn and Yager, 1980 (16)§	NR	NR	NR	NR	
Zung and King, 1983 (17)‖	Percentage of participants with <12-point decrease on SDS at 1 month	33%	65%	−32 percentage points (−61 to −3)	0.04
Magruder-Habib et al., 1990 (18)	NR	NR	NR	NR	
Callahan et al., 1994 (19) and 1996 (21)‖	Percentage of participants with HAM-D ≥10 at 6 months	87%	88%	−1 percentage point (−11 to 9)	
Dowrick, 1995 (20)‡	NR	NR	NR		
Lewis et al., 1996 (22)‡	Percentage of participants who had not improved at 6 weeks (GHQ score ≥ 2)	69%	74.5%	−5 percentage points (−14 to 3)	
Reifler et al., 1996 (23)‡	Zung SDS score	¶	¶	¶	
Williams et al., 1999 (11)‡	Percentage of participants who were depressed at 3 months (DSM-III-R criteria)	37%	46%	−8 percentage points (−21 to 4)	
Katzelnick et al., 2000 (12)	Percentage of participants who were depressed at 12 months (HAM-D ≥ 7)	55%	72%	−18 percentage points (−27 to −8)	<0.001
Wells et al., 2000 (24)‡	Percentage of participants who were depressed at 6 months	55.4%	64.4%	−9 percentage points (−15 to −3)	0.005
Whooley et al., 2000 (25)	Percentage of participants who were depressed at 24 months (GDS ≥ 6)	42%	50%	−8 percentage points (−21 to 6)	>0.2
Rost et al., 2001 (13)	Mean change in CES-D score	21.7	13.5	8.2 (NR)	<0.05

* All figures are rounded to nearest percentage. CES-D = Center for Epidemiologic Study Depression Scale; DSM-III-R = *Diagnostic and Statistical Manual of Mental Disorders*, 3rd edition, revised; GDS = Geriatric Depression Scale; GHQ = General Health Questionnaire; HAM-D = Hamilton Depression Scale; NR = not reported and cannot be calculated from available data; SDS = Zung Self-Depression Scale.
† P values were not always reported.
‡ Denominator is patients who screened positive.
§ Denominator is all patients.
‖ Denominator is patients who screened positive and were confirmed to have major depression on diagnostic interview.
¶ No data were given; the investigators stated that there was "no difference for those screening positive for any disorder."

Katzelnick and associates compared the benefits of a systematic primary care–based depression treatment program for depressed "high utilizers" not already receiving treatment of depression (12). Using a health maintenance organization database, they defined eligible patients as those who had had ambulatory visits at a rate greater than the 85th percentile over 2 years. They then identified depressed patients by using a two-stage telephone screening process. Initial screening was performed by using the depression-specific portion of the Structured Clinical Interview for DSM-IV; patients who screened positive then completed the Hamilton Depression Scale and were eligible if their score was greater than 15 (28). The investigators randomly assigned practices to the intervention program or to usual care. Patients receiving usual care were notified that they had screened positive for depression and were counseled to see their physicians, but no feedback was given directly to providers. Intervention-group patients were invited to participate in a depression management program that consisted of patient education materials, physician education programs, telephone-based treatment coordination, and antidepressant medication treatment that was initiated and managed by the primary care physician. In an intention-to-treat analysis, patients who received the depression management program were significantly more likely than usual care recipients to fill a prescription for antidepressants in the first 6 months (82% vs. 32%; difference, 50 percentage points [CI, 41 to 58 percentage points]). At 1 year of follow-up, 55% of depression management program participants and 72% of usual care recipients (difference, −18 percentage points [CI, −27 to −8 percentage points]) were still depressed.

Rost and coworkers examined the effectiveness of a systematic approach to identification and treatment of depression within primary care practices (13). The researchers randomly assigned 12 practices to usual care or to a quality improvement intervention. They identified patients by using initial screening questions about anhedonia or depressed mood, followed by confirmatory diagnostic questions from the Inventory to Diagnose Depression. Usual-care recipients received no further treatment, whereas intervention recipients received materials designed to increase adherence to medical therapy and intervention staff were offered additional training. The intervention improved outcomes in patients who had not recently been treated for depression but not in

patients who had been recently treated for depression (mean change in Center for Epidemiologic Study Depression score, −8.2 [$P < 0.05$] vs. −3.5 [$P > 0.05$], respectively).

## Summary of the Effect of Screening and Feedback

Feedback of screening results to providers generally increases the recognition of depression, especially major depression, by a factor of two to three. The absolute increases in the diagnosis of depression range from 10 to 47 percentage points. In contrast, trials examining the effect of screening and feedback on treatment rates have had mixed results. In three studies, the documented rates of treatment were nearly equal in the intervention and control groups (11, 16, 20). Other studies, however, found improvements in the rate of treatment; increases in the prescription of antidepressant medication were more common than changes in mental health referrals.

The results of individual studies were also mixed with respect to the effect of screening on clinical outcomes: Some found positive results, whereas others did not (**Table 5**). The wide variation in interventions tested, outcome measures used, and timing of follow-up assessments hampered interpretation of overall results. Seven of 10 studies reported the proportion of patients who were still depressed at some time after initial screening. In these studies, the proportion of patients who were still depressed was lower in the intervention group than the control group, although results were significant in only 3 studies. Of the 3 studies that examined health outcomes but did not report the proportion of depressed patients, 2 had positive results for some outcomes (13, 14) and 1 reported no effect for any outcome (23).

We examined several potential factors to explain the mixed results. We found no consistent relationships between differences in outcomes and patient and provider characteristics, use of particular outcome measures, varying duration of follow-up, or trial quality.

The trials that we identified examined a range of strategies, including simple feedback of scores obtained from depression screening questionnaires; feedback given in the context of general education efforts for providers; feedback with treatment advice that may or may not have been tailored to specific patients; and integrated recognition and management approaches that relied on multiple system supports within the clinic to

Figure 1. **Meta-analysis of the effect of screening and feedback on the proportion of patients with persistent depression.**
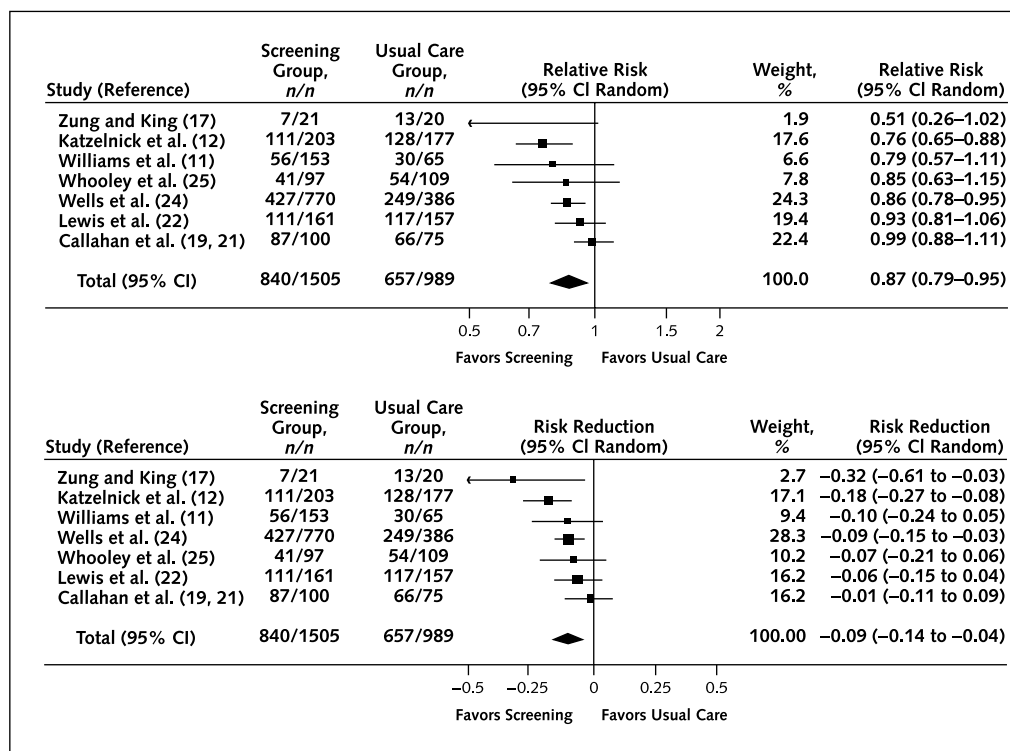
Top. Summary estimate of relative risk for persistent depression for screening compared with no screening. **Bottom.** Summary estimate of absolute risk reduction in persistent depression with screening compared with no screening.

ensure prompt, coordinated follow-up of diagnosis and treatment. Data from existing trials do not definitively rule in or rule out clinical benefits from less intensive interventions, such as feedback alone. Limited data suggest that delayed feedback of results, as provided by Dowrick (20), may be less effective than immediate feedback. Intensive, integrated identification and management that incorporated quality improvements in clinic systems have demonstrated clinical effectiveness in broad-based primary care clinic populations (12, 24).

Many trials that did not find a statistically significant difference in outcomes were not sufficiently powered to exclude clinically important differences in outcomes. For example, the point estimates of effect in the studies by Williams (11) and Whooley (25) and their colleagues, both of which were considered "negative" trials, were similar to the effect seen in the larger trial by Wells and associates (24), which had a statistically significant result and has been interpreted as a "positive" study. This finding suggests that the mixed results may

be explained in part by differences in adequacy of sample sizes among trials.

Meta-Analysis

Because many trials had insufficient power to exclude the possibility of clinically significant changes in clinical outcomes, we used meta-analysis to determine a summary estimate of effect.

We used a random-effects model to combine the seven trials that had sufficient data for meta-analysis (**Figure 1**). The summary relative risk for remaining depressed was 0.87 (CI, 0.79 to 0.95) for intervention recipients, suggesting that screening provided a 13% reduction in relative risk. The summary estimate of the risk difference was −9 percentage points (CI, −14 to −4 percentage points). We detected heterogeneity in the results for the outcome of reduction in relative risk ($P = 0.052$), in large part because of the strongly positive study by Katzelnick and associates (12).

Because of the heterogeneity in the full meta-analy-

Figure 2. **Meta-analysis of the effect of screening and feedback on the proportion of patients with persistent depression, excluding the study by Katzelnick and colleagues (12).**
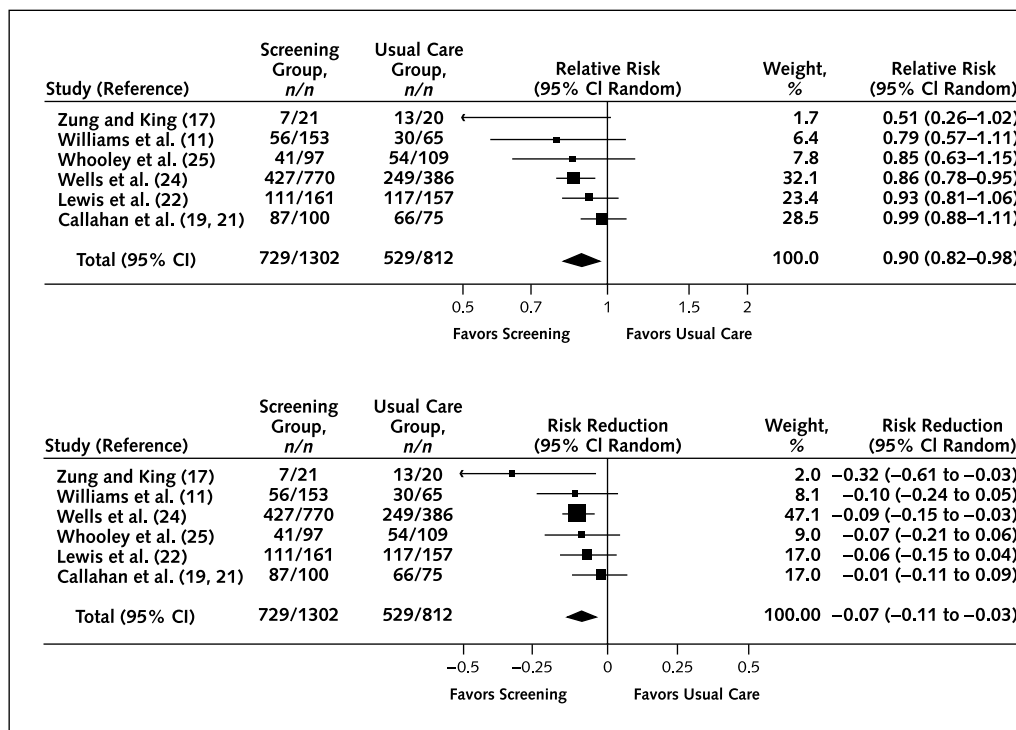
Top. Summary estimate of relative risk for persistent depression for screening compared with no screening. **Bottom.** Summary estimate of absolute risk reduction in persistent depression with screening compared with no screening.

sis, we performed an alternative analysis from which we excluded the latter trial (**Figure 2**). In this alternative analysis, the summary risk reductions with screening were slightly smaller (relative risk, 0.90 [CI, 0.82 to 0.98]; summary risk difference, −7 percentage points [CI, −11 to −3 percentage points]) but heterogeneity was reduced ($P = 0.16$).

DISCUSSION

Whether care that incorporates screening for depression is superior to care based on usual methods of case identification is controversial. Multiple studies have examined the effect of providing feedback on results of screening for depression to providers in primary care settings. The rates of detection and diagnosis of depression, which are based mainly on chart review or completion of a study-specific form, increased by 10% to 47% in most studies reporting these outcomes. The effect on the proportion of patients receiving treatment was mixed: Some studies showed large increases (18, 19, 24), whereas others found no significant effect (11, 16, 20, 25).

Some individual trials examining the effect of screening on clinical outcomes have found positive results, but others have not. Many studies have been underpowered to detect clinically important differences in effectiveness. When the results of trials reporting interpretable clinical outcomes are combined, summary estimates suggest that screening is associated with a 13% reduction in relative risk and a 9–percentage point absolute reduction in the proportion of patients with persistent depression. Heterogeneity in trial results was noted on statistical testing, in large part because of the large positive effect reported in a trial that involved depressed patients who had frequent clinic visits (12). However, an alternative analysis that excluded this trial, and hence had less heterogeneity, showed only slightly smaller benefit from screening. These findings suggest that screening is probably effective in primary care patients with depression who are not high utilizers.

If screening can increase the proportion of patients achieving remission by 9% at 6 months, approximately

11 patients with depression would need to be identified to produce 1 additional remission. If the prevalence of treatment-responsive depression in primary care patients is 10%, 110 patients would need to be screened to produce 1 additional remission after 6 months of treatment.

Other reviewers have also examined the value of screening for depression and have reached divergent conclusions. Gilbody and coworkers performed a systematic review of routinely administered questionnaires for anxiety and depression published through 2000 (29). They identified six studies, five of which were included in our review. They did not include the recent trials by Callahan (19), Williams (11), and Whooley (25) and their colleagues, nor did they include the newer trials that used integrated efforts to improve recognition and treatment systems (12, 13, 24). They concluded that routine questionnaires did not increase recognition, treatment, or outcomes of depression, but their failure to include several large, recent studies with positive outcomes limits the validity of their conclusions (30).

Kroenke and associates performed a systematic review of studies published through May 1998 that addressed diverse interventions to improve recognition and treatment of mental disorders (primarily depression and anxiety) in primary care (31). They identified 27 randomized trials of interventions; of the 11 trials that focused on depression, we included 7 in our review. Most interventions, including screening and feedback, improved recognition and treatment; about half of the studies showed improved outcomes. The researchers chose not to combine the results in a meta-analysis because the studies used different outcome measures.

Several recent cost-effectiveness analyses have addressed the question of whether a modest improvement in depression outcomes warrants the increased effort of screening and providing systematic support for treatment. Valenstein and coworkers developed a cost–utility model to examine the consequences of screening a hypothetical cohort of 40-year-old adults, using estimates derived from the literature (32). In the base case of their Markov model, they assumed a prevalence of major depression of 8%; a sensitivity and specificity for the detection of major depression of 84% and 85%, respectively; and a cost of screening of $5.00 per person. They also assumed that 35% of patients would have full remission without treatment and that rates of full remission in standard or enhanced care settings would be 45% and 50%, respectively. They estimated that one-time screening had a cost–utility ratio of about $45 000 per quality-adjusted life-year gained; annual screening had a cost of more than $100 000 per quality-adjusted life-year gained.

Using data on costs and effectiveness obtained directly from the trial by Wells and colleagues (24), Schoenbaum and coworkers (33) examined the cost–utility of the screening and treatment support program studied by Wells and colleagues. Relative to usual care, the enhanced program, which included one-time screening and support to improve treatment, yielded additional benefits at a cost of $10 000 to $35 000 per quality-adjusted life-year gained. In a similar analysis that used data obtained directly from the study by Katzelnick and associates (12), Simon and colleagues (34) found a cost per depression-free day gained of $51.84 (CI, $17.37 to $108.47).

Cost-effectiveness data from the two recent trials of systematic efforts to screen for depression and provide integrated support for treatment suggest that such programs can be implemented efficiently and can produce cost-effectiveness ratios similar to those of other commonly performed preventive services, such as screening mammography in women older than 50 years of age or treatment of mild to moderate hypertension. Further research is required to determine which components of these integrated programs are most effective and to determine whether more efficient means of delivering effective care are possible.

Requests for Single Reprints: Reprints are available from the Agency for Healthcare Research and Quality Web site (www.ahrq.gov/clinic /uspstfix.htm) or the Agency for Healthcare Research and Quality Publications Clearinghouse (800-358-9295).

Current author addresses are available at www.annals.org.

References

1. **Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, et al.** Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. Arch Gen Psychiatry. 1994;51:8-19. [PMID: 8279933]

2. **Robins LN, Regier DA.** Psychiatric Disorders in America: The Epidemiologic Catchment Area Study. New York: Free Pr; 1991.

3. Depression in Primary Care: Volume 1. Detection and Diagnosis. Depression Guideline Panel. Clinical Practice Guideline No. 5. Rockville, MD: U.S. Department of Health and Human Services; 1993.

4. **Murray CJ, Lopez AD.** Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. Lancet. 1997;349:1436-42. [PMID: 9164317]

5. **Greenberg PE, Stiglin LE, Finkelstein SN, Berndt ER.** The economic burden of depression in 1990. J Clin Psychiatry. 1993;54:405-18. [PMID: 8270583]

6. **Penninx BW, Guralnik JM, Ferrucci L, Simonsick EM, Deeg DJ, Wallace RB.** Depressive symptoms and physical decline in community-dwelling older persons. JAMA. 1998;279:1720-6. [PMID: 9624025]

7. **Simon GE, VonKorff M.** Recognition, management, and outcomes of depression in primary care. Arch Fam Med. 1995;4:99-105. [PMID: 7842160]

8. Guide to Clinical Preventive Services. 2nd ed. U.S. Preventive Services Task Force. Baltimore: Williams & Wilkins; 1996:541-6.

9. **Pignone M, Gaynes BN, Rushton JL, Mulrow CD, Orleans CT, Whitener BL, et al.** Screening for Depression. Systematic Evidence Review No. 6. Rockville, MD: Agency for Healthcare Research and Quality; 2001. Publication no. AHRQ02-S002.

10. **Mulrow CD, Williams JW Jr, Gerety MB, Ramirez G, Montiel OM, Kerber C.** Case-finding instruments for depression in primary care settings. Ann Intern Med. 1995;122:913-21. [PMID: 7755226]

11. **Williams JW Jr, Mulrow CD, Kroenke K, Dhanda R, Badgett RG, Omori D, et al.** Case-finding for depression in primary care: a randomized trial. Am J Med. 1999;106:36-43. [PMID: 10320115]

12. **Katzelnick DJ, Simon GE, Pearson SD, Manning WG, Helstad CP, Henk HJ, et al.** Randomized trial of a depression management program in high utilizers of medical care. Arch Fam Med. 2000;9:345-51. [PMID: 10776363]

13. **Rost K, Nutting P, Smith J, Werner J, Duan N.** Improving depression outcomes in community primary care practice: a randomized trial of the QuEST intervention. Quality Enhancement by Strategic Teaming. J Gen Intern Med. 2001;16:143-9. [PMID: 11318908]

14. **Johnstone A, Goldberg D.** Psychiatric screening in general practice. A controlled trial. Lancet. 1976;1:605-8. [PMID: 55891]

15. **Moore JT, Silimperi DR, Bobula JA.** Recognition of depression by family medicine residents: the impact of screening. J Fam Pract. 1978;7:509-13. [PMID: 690583]

16. **Linn LS, Yager J.** The effect of screening, sensitization, and feedback on notation of depression. J Med Educ. 1980;55:942-9. [PMID: 7441675]

17. **Zung WW, King RE.** Identification and treatment of masked depression in a general medical practice. J Clin Psychiatry. 1983;44:365-8. [PMID: 6643397]

18. **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. [PMID: 2179650]

19. **Callahan CM, Hendrie HC, Dittus RS, Brater DC, Hui SL, Tierney WM.** Improving treatment of late life depression in primary care: a randomized clinical trial. J Am Geriatr Soc. 1994;42:839-46. [PMID: 8046193]

20. **Dowrick C.** Does testing for depression influence diagnosis or management by general practitioners? Fam Pract. 1995;12:461-5. [PMID: 8826066]

21. **Callahan CM, Dittus RS, Tierney WM.** Primary care physicians' medical decision making for late-life depression. J Gen Intern Med. 1996;11:218-25. [PMID: 8744879]

22. **Lewis G, Sharp D, Bartholomew J, Pelosi AJ.** Computerized assessment of common mental disorders in primary care: effect on clinical outcome. Fam Pract. 1996;13:120-6. [PMID: 8732321]

23. **Reifler DR, Kessler HS, Bernhard EJ, Leon AC, Martin GJ.** 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. [PMID: 8951303]

24. **Wells KB, Sherbourne C, Schoenbaum M, Duan N, Meredith L, Unützer J, et al.** Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. JAMA. 2000;283:212-20. [PMID: 10634337]

25. **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. [PMID: 10840264]

26. Diagnostic and Statistical Manual of Mental Disorders: DSM-III. 3rd ed. American Psychiatric Association. Washington, DC: American Psychiatric Assoc; 1980.

27. Diagnostic and Statistical Manual of Mental Disorders: DSM III-R. 3rd ed., revised. American Psychiatric Association. Washington, DC: American Psychiatric Assoc; 1987.

28. **First M, Spitzer R, Gibbon M, Williams JB.** Structured Clinical Interview for DSM-IV Axis I Disorders—Patient Edition (SCID-I/P, Version 2.0). New York: New York State Psychiatric Institute; 1995.

29. **Gilbody SM, House AO, Sheldon TA.** Routinely administered questionnaires for depression and anxiety: systematic review. BMJ. 2001;322:406-9. [PMID: 11179161]

30. **Pignone M, Gaynes BN, Lohr KN, Orleans CT, Mulrow C.** Questionnaires for depression and anxiety. Systematic review is incomplete [Letter]. BMJ. 2001;323:167-8. [PMID: 11463699]

31. **Kroenke K, Taylor-Vaisey A, Dietrich AJ, Oxman TE.** Interventions to improve provider diagnosis and treatment of mental disorders in primary care. A critical review of the literature. Psychosomatics. 2000;41:39-52. [PMID: 10665267]

32. **Valenstein M, Vijan S, Zeber JE, Boehm K, Buttar A.** The cost-utility of screening for depression in primary care. Ann Intern Med. 2001;134:345-60. [PMID: 11242495]

33. **Schoenbaum M, Unützer J, Sherbourne C, Duan N, Rubenstein LV, Miranda J, et al.** Cost-effectiveness of practice-initiated quality improvement for depression: results of a randomized controlled trial. JAMA. 2001;286:1325-30. [PMID: 11560537]

34. **Simon GE, Manning WG, Katzelnick DJ, Pearson SD, Henk HJ, Helstad CS.** Cost-effectiveness of systematic depression treatment for high utilizers of general medical care. Arch Gen Psychiatry. 2001;58:181-7. [PMID: 11177120]