Screening for Depression

Systematic Evidence Review

Submitted to:
Agency for Healthcare Research and Quality
6010 Executive Boulevard, Suite 300
Rockville, Maryland 20852

Submitted by:
Research Triangle Institute
3040 Cornwallis Road
P.O. Box 12194
Research Triangle Park, North Carolina 27709

Contract No. 290-97-0011
Task No. 3
RTI Project No. 6919-003

April 2002

Screening for Depression

Systematic Evidence Review

Mike Pignone, MD MPH
Bradley N. Gaynes, MD MPH
Jerry L. Rushton, MD MPH
Cynthia D. Mulrow, MD, MSc
C. Tracy Orleans, PhD
B. Lynn Whitener, DrPH
Catherine Mills, MA
Kathleen N. Lohr, PhD

for the
Research Triangle Institute - University of North Carolina
Evidence-based Practice Center

April 2002

Contents

EVIDENCE REPORT

Figures

Tables

Structured Abstract

Context. Depressive disorders are an important cause of morbidity and are common in primary care settings. Previous research suggests that depression is underrecognized and undertreated. Screening for depression in primary care settings may improve recognition, treatment, and outcomes of depressive disorders.

Objective. To review systematically the literature regarding the effectiveness of screening for depressive disorders in primary care settings.

Data Sources. We systematically searched MEDLINE from 1994 through 1999 using 2 Medical Subject Headings (MeSH) terms, depression and depressive disorders, and combined them with predefined strategies to identify diagnostic accuracy studies and randomized controlled trials of screening and treatment. We used the second edition of the *Guide to Clinical Preventive Services*, recent systematic reviews, and focused searches of MEDLINE from 1966 to 1994 to identify older articles of interest. We also used hand checking of bibliographies; a search of the Cochrane depression, anxiety, and neurosis database; and extensive peer review to identify articles not captured through our main search strategy.

Structured Abstract

Study Selection. Diagnostic accuracy studies were included if they reported sensitivity and specificity results based on evaluation against a criterion standard. Treatment studies were included if they were randomized trials in primary care populations. Screening outcome studies were included if they were randomized trials that reported outcomes of change in recognition or treatment of depression or change in health outcomes.

Two reviewers initially examined titles and abstracts of articles and excluded those that clearly did not meet inclusion criteria. Two reviewers then examined the full articles of the remaining studies and determined final eligibility by consensus.

Data Extraction. A single reviewer abstracted the relevant data from the included articles and entered them into evidence tables. A second reviewer checked the accuracy of the tables against the original articles.

Data Synthesis. Studies examining the diagnostic accuracy of depression screening instruments generally have found sensitivity results of 80% to 90% and specificity results from 70% to 85% in adults and 60% to 100% and 60% to 85% in children.

For adult primary care patients with major depression, treatment with pharmacotherapy or psychotherapy reduces symptom duration and severity. Pharmacotherapy and psychotherapy appear to produce a similar magnitude of effect. Approximately 4 patients must be treated to produce 1 additional clinical remission. Cognitive-behavioral therapy appears to reduce depression scores in children and adolescents. Data on pharmacotherapy for children and adolescents are mixed: tricyclic agents appear ineffective, and data for selective serotonin reuptake inhibitor drugs are inconclusive but promising.

Structured Abstract

Feedback of screening results to providers appears to increase recognition of depression in adults compared with usual care but its effect on treatment and clinical outcomes are mixed. Screening appears to be more effective when coupled with systematic efforts to ensure adequate treatment and follow-up. The effect of screening has not been evaluated in children.

Other than medication side effects, little evidence is available about the potential harms of screening and treatment of depression.

Conclusions. Accurate screening tests and effective therapies for depression are available. Screening for depression can improve outcomes compared with usual care in adults, particularly when coupled with efforts to ensure adequate treatment and follow-up. The effect of screening in children and adolescents is unknown.

Chapter 1. Introduction

Burden of Suffering

Depressive disorders are common, chronic, and costly. Lifetime prevalence levels from community-based surveys range from 4.9% to 17.1%.[1-3] In primary care settings, the prevalence of major depression is 6% to 8% (Table 1).[4] Longitudinal studies suggest that about 80% of individuals experiencing a major depressive episode will have at least 1 more episode during their lifetime, with the rate of recurrence even higher if minor or sub-threshold episodes are included.[5] Approximately 12% of patients who experience depression will have a chronic, unremitting course.[5] The substantial public health and economic significance of this chronic illness is reflected by the considerable utilization of health care visits and tremendous monetary costs: $43 billion (1990 dollars) annually, with $17 billion of that resulting from lost work days.[6]

The burden of suffering from depression is substantial. Suicide, the most severe of depressive sequelae, has a rate of approximately 3.5% among all cases with major depression, a risk that increases to approximately 15% in people who have required psychiatric hospitalization.[7] The specific risk for suicide associated with depressive disorders is elevated 12- to 20-fold compared to the general population.[8] The World Health Organization (WHO) identified major depression as the fourth leading cause of worldwide disease burden in 1990, causing more disability than either ischemic heart disease or cerebrovascular disease. Its associated morbidity is expected to increase; unipolar depressive illness is projected to be the second leading cause of disability worldwide in 2020.[9] Furthermore, depression

appears to contribute to increased morbidity and mortality from other medical disorders, such as cardiovascular disease.[10]

Both the chronicity and recurrence of depressive illness play a large role in depression's heavy disease burden. The more severe a depression becomes and the longer it lasts, the greater the likelihood that the depression will become chronic.[11] Consequently, early effective identification and management of depressive illness will not only decrease the substantial morbidity associated with the current episode but may also decrease the likelihood that the illness will become chronic, with its additional associated morbidity.[12]

Epidemiology of Depressive Illness in Adults

Major Depression

Depressive illness can have a variety of presentations, and these range in both severity and chronicity. Major depression is the most severe form; according to criteria of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV), it consists of an episode of at least 2 weeks in which an individual has 5 of 9 specific depressive symptoms, 1 of which must be depressed mood or anhedonia (loss of interest or pleasure).[13] These symptoms must cause clinically significant distress or impairment in social, occupational, or other important areas of functioning, a requirement which emphasizes the marked disability resulting from depressive illness. Major depression has a prevalence of 6% to 8% in the primary care setting, making it as common a presentation as hypertension.[4]

Dysthymia

Dysthymia, a chronic, low-grade depressive illness of at least 2 years' duration, has a prevalence of 2% to 4% in the primary care setting.[3] Although its symptoms are less severe, the morbidity associated with dysthymia is substantial.[14] The severe impact of the illness is reflected by the 17% of patients with dysthymia who make serious suicide attempts.[15] Furthermore, it is a risk factor for subsequent development of a major depressive episode.[16]

Sub-threshold Disorders

Sub-threshold disorders consist of depressive symptoms that are not severe enough to meet DSM-IV criteria but that still cause substantial disability.[13] They are as common as major depression in primary care settings. Presentations may include remitting major depressive episodes, evolving major depressive episodes, or episodes that will never reach criterion for a major depression.

Minor depression, an episodic sub-threshold disorder that is similar to major depression, consists of between 2 and 4 DSM depressive symptoms. It is not an officially recognized DSM-IV diagnosis but is included in DSM-IV as a type of "Depressive Disorder Not Otherwise Specified."[13] Minor depression is at least as common as major depression in primary care sites (point prevalence 8% to 10%).[17] Health-related quality-of-life measures, including physical health, disability, and social functioning, are significantly more impaired for people with minor depression than for people who are not depressed and only slightly better than those with major depression.[18] One-fifth of people with minor depression may progress to major depression within the year.[17]

Depression Severity in Primary Care

In general, depressive illness is less severe in primary care than in mental health settings. Patients have fewer psychiatric symptoms, a lower likelihood of a history of major depression, a lower likelihood of having received prior treatment, and a lower risk of psychiatric hospitalization.[19] The short-term prognosis is better, with a greater chance of recovery at 1 year follow-up[19] and a higher rate of response to treatment.[20] Furthermore, this improved prognosis may be independent of adequate treatment for depression.[21]

Epidemiology of Depressive Disorders in Children and Adolescents and Special Populations

Depressive disorders are common in childhood and adolescence. The prevalence of major depressive disorder (MDD) is 0.8% in preschool children, 2% in school-age children, and 4.5% in adolescents.[3]

Patients with co-occurring depressive and medical illnesses are a key subpopulation as they are at risk of not receiving potentially effective antidepressant therapies.[22] Those with other co-occurring psychiatric illnesses, including substance abuse and anxiety disorders, are at risk for persistent depressive illness.[23] Additionally, differences in depressive illness among different ethnic groups are an important but understudied area. Where the literature provides specific information, we will address the screening and treatment issues for these special populations throughout our review.

Health Care Interventions

Key Role of Primary Care Providers

Primary care practices play a substantial role in the assessment and management of depressive illness. As the initial provider seen by most patients entering the health care system, primary care physicians frequently offer the first opportunity for identification of depressive illness. They also provide the bulk of treatment for depression. People with depressive disorders are more likely to receive treatment from a primary care physician than a mental health professional,[24] and primary care physicians record approximately the same number of yearly patient visits for antidepressant prescriptions as do psychiatrists.[25] However, primary care physicians fail to recognize and treat 30% to 50% of adult depressed patients.[26,27] Multiple competing demands, complicated presentations, limited time, and minimal training make identifying and managing depressive illness in a primary care setting a challenging task.[28] Failure to detect depression may be greater for African American or Hispanic patients and for patients under 35 years.[29]

Interventions for depression include antidepressant medication, herbal therapies, psychosocial therapies, educational and quality improvement strategies, electroconvulsive therapy, and light therapy. The latter 2 are not first-line primary care treatments and will not be addressed in this review. General categories of therapeutic interventions are listed in Table 2.

Antidepressant medications include tricyclic antidepressants (TCAs), heterocyclic antidepressants, selective serotonin reuptake inhibitors (SSRIs), monoamine oxidase inhibitors (MAOIs), and other newer agents (such as norepinephrine-serotonin reuptake inhibitors). Alternative

herbal therapies, such as St. John's Wort, may also be effective in treating depression, but they are not reviewed here.[30]

Psychotherapy is defined as a formal, time-limited communication intervention. Specific forms of psychotherapy that have been studied in primary care populations include cognitive-behavioral therapy and problem-solving therapy. Each of these approaches is based on the theory that distorted thoughts and maladaptive coping strategies lead to depressive illness. Interpersonal therapy (IPT) conceives of depressive illness as an expression of dysfunctional or problematic relationships. Psychotherapies may vary in terms of how formally structured they are, how much contact time is required, and who provides the therapy. Supportive counseling, which may be offered by health care workers with relatively less training and is often based on Rogerian theory, is a less structured form of psychotherapy. Psychoanalytic psychotherapy has not been studied in primary care populations.

Prior Recommendations

In 1996, the U.S. Preventive Services Task Force (USPSTF) found insufficient evidence to recommend for or against routine screening for depression with standardized questionnaires. They recommended that clinicians maintain a 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 unexplained somatic symptoms."[31] The USPSTF also encouraged physician education in recognition and treatment of depression but did not issue a graded recommendation.

The American College of Physicians (ACP/ASIM) recently released guidelines on the use of pharmacotherapy for depression, but the ACP/ASIM does not have an official policy on routine screening in primary care.[32] The American Academy of Family Physicians also does not have a position on depression screening. Currently, the American Academy of Pediatrics (AAP) has no position statements or guidelines that specifically address the screening, diagnosis, and/or treatment of depression. AAP committees have encouraged pediatricians to include psychosocial questions about the child and family in routine medical interviews and to consider depression in specific groups including children with chronic medical disorders, adolescents considering suicide, victims of violence and natural disasters, and other high-risk groups. Even for these special groups, however, no specific screening instruments are recommended or discussed. In 1994, the Canadian Task Force on the Periodic Health Examination (now the Canadian Task Force on Preventive Health Care) examined the question of screening and recommended against performing routine screening.[33]

Analytic Framework and Key Questions

The Research Triangle Institute and University of North Carolina at Chapel Hill Evidence-based Practice Center (RTI-UNC EPC), together with members of the current USPSTF and other clinical and methodologic experts (Appendix A), sought to clarify issues concerning the screening for and treatment of depression by performing a systematic review of the relevant scientific literature on these topics. This systematic evidence review (SER) specifically updates Chapter 49 of the second *Guide to Clinical Preventive Services* produced in 1996 by the previous USPSTF.[31] A glossary of commonly used

abbreviations and acronyms for screening instruments, therapies, and other terms used in this SER can be found in Appendix B.

For prevention to be effective, 3 requirements must be met. First, a reliable and feasible screening process must be available that can accurately identify primary care patients with depression. Second, effective treatment must be available that can improve outcomes for depressed patients. Third, treatment in those detected by screening must improve outcomes compared with usual care in the absence of screening. Our approach to producing this SER on screening for depression takes these 3 issues into account, as discussed with respect to the analytic framework and key questions (below).

Analytic Framework

The analytic framework for this SER is depicted in Figure 1. People with unrecognized depression undergo screening for depression. Screening can correctly classify patients with depression as "depressed" or patients without depression as "not depressed," or it can make false-negative or false-positive mistakes. Patients correctly identified as depressed may then undergo treatment, which may lead to improved scores on depression screening instruments and may also reduce morbidity and mortality, and improve quality of life. Treatment may also have adverse effects, including medication side effects or unnecessary treatment for patients who would have an uncomplicated, nondisabling episode in the absence of treatment. Trials of screening may increase the identification of depression, increase the proportion of depressed people who are treated, or improve indices of depressed mood when compared with usual care.

Key Questions

Based on the analytic framework, we developed 3 key questions:

1. What is the accuracy of screening instruments for depression in primary care populations?
2. Is treatment of depression in primary care patients (with pharmacologic therapy, psychotherapy, combinations of the 2, or educational interventions) effective in improving outcomes?
3. Is screening more effective than usual care in identifying patients with depression, facilitating treatment of patients with depression, and improving outcomes?

The key questions include the direct effects of screening on detection, treatment, and outcomes (Key Question No. 3) and the 2 main links in the screening "chain"—namely, the ability of the test to detect depressed patients (Key Question No. 1) and the availability of effective treatment for patients who would be detected by screening (Key Question No. 2). Because our initial survey of the evidence regarding the direct effects of screening suggested that data to answer this question were limited and inconclusive, we decided to examine the evidence for each of the main links in the screening chain as well.

The linkage between studies that examine only diagnostic accuracy and studies that examine only treatment is difficult to study directly because the spectrum of patients included in each type of study may be different. We attempted to examine the evidence for each question that would most likely be generalizable to the patients screened in primary care settings.

Table 1. Prevalence of Depressive Illness

Group Studied	Condition	Prevalence (%) Point	Lifetime
Community	Major depression	Men, 2-3 Women, 4-9	Men, 7-12 Women, 20-25
	Dysthymia	---	Men, 2.2 Women, 4.1
	Depression NOS	11	
Primary care settings	Major depression	4.8-8.6	---
	Dysthymia	2.1-3.7	---
	Minor depression	8.4-9.7	---
Patients with medical illness	Clinically significant depression	12-16	---

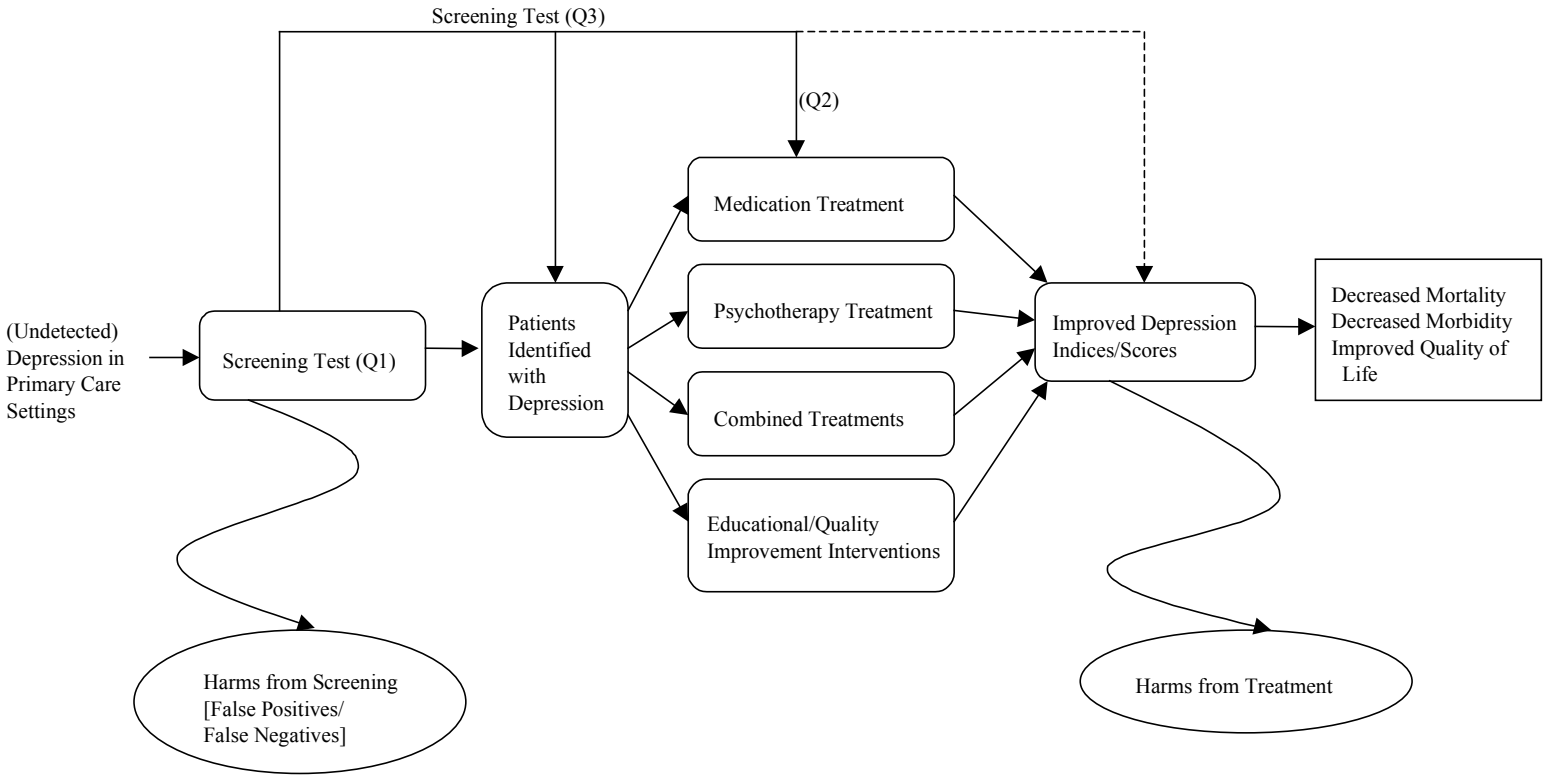
Source: Depression Guideline Panel, 1993.[3]

--- Indicates prevalence rates not available

Table 2. Treatment Interventions for Patients Identified with Depression in Primary Care

Type	Examples
Medication	Tricyclic antidepressants (TCAs)
	Heterocyclic antidepressants
	Selective serotonin reuptake inhibitors (SSRIs)
	Monoamine oxidase inhibitors (MAOIs)
	Norepinephrine-serotonin reuptake inhibitors, including reversible monoamine oxidase inhibitors (RIMAs)
Psychotherapy	Cognitive-behavioral therapy (CBT) and problem-solving therapy
	Interpersonal therapy (IPT)
	Supportive therapy (by a social worker or health visitor)
	Psychoanalytic psychotherapy (referral to psychiatrist)
Alternative therapies	St. John's Wort
Electroconvulsive therapy	
Light therapy	

Figure 1: USPSTF Analytic Framework for Depression



Q1. What is the accuracy of screening instruments for depression in primary care populations?

Q2. Is treatment of depression in primary care patients (with pharmacologic therapy, psychotherapy, combinations of the 2, or educational interventions) effective in improving outcomes?

Q3. Is screening more effective than usual care in identifying patients with depression, facilitating treatment of patients with depression, and improving outcomes?

Chapter 2. Methods

This chapter of the SER documents the procedures that the RTI-UNC Evidence-based Practice Center (EPC) used to develop this report on screening for depression among adults and children. We document the literature search (eg, inclusion and exclusion criteria, relevant Medical Subject Headings [MeSH terms]) and briefly describe the procedures followed in abstracting data from included articles, developing evidence tables, analyzing the literature, and subjecting the draft to a robust peer review process.

In all these steps, EPC staff collaborated with 2 members of the USPSTF who acted as liaisons for this topic; they are co-authors of the SER. This collaboration took place chiefly by e-mail and numerous conference calls. Steps in the development of this SER were presented at USPSTF meetings in May and September 1999 and February 2000, where the EPC staff, USPSTF liaisons, and the full Task Force were able to discuss the analytic framework and key questions, literature search strategy, results, and implications of the findings.

Literature Search Strategy

To identify articles relevant to the questions of screening and treatment of depression, the EPC staff searched the MEDLINE database from 1994 to 1999 and used recent systematic reviews. We supplemented these sources by searching the Cochrane database on depression, neurosis, and anxiety disorders; conducting additional specific MEDLINE searches from 1966 to 1994; and hand-searching

bibliographies of systematic reviews, relevant original articles, the second edition of the *Guide to Clinical Preventive Services,*[31] and the 1993 Clinical Practice Guideline on Depression from the Agency for Health Care Policy and Research (now the Agency for Healthcare Research and Quality).[3]

Inclusion/Exclusion Criteria

We prospectively established eligibility criteria for all searches. Table 3 presents these criteria. We restricted the search to articles published in English and excluded nonpublished studies, those published in abstract form only, letters, and editorials.

Diagnosis articles were identified by searching for studies with information about diagnostic accuracy, particularly sensitivity and specificity. We included only those articles that compared the screening instrument with a criterion standard. For articles on therapy, we restricted the search to randomized controlled trials (RCTs) and meta-analyses of RCTs. For articles on direct effects of screening and feedback, we included randomized trials and before-and-after studies of identification, treatment, or health outcomes.

We also used the second edition of the USPSTF *Guide to Clinical Preventive Services,*[31] as well as systematic reviews, meta-analyses, and evidence-based practice guidelines that addressed screening and treatment of depression, to identify key articles that appeared earlier than the 1994 or 1995 period. Finally, we reviewed the bibliographies of included articles to detect any important articles that may have been missed at other steps. Table 4 documents the results of the 2 main literature searches.

Literature Reviewed

Two EPC staff 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 EPC team members reviewed the full articles and made a final decision about inclusion or exclusion. Table 5 summarizes the results of the literature searches and reviews of abstracts.

Literature Synthesis and Preparation Of Systematic Evidence Review

Data Abstraction and Development of Evidence Tables

Reviewers entered study design and outcomes data from the articles on screening accuracy, screening outcomes, and treatment onto paper abstraction forms. These data were used to construct evidence tables.

To characterize the quality of the included studies, the internal and external validity for each article were rated in the evidence tables using criteria developed by the USPSTF Methods Work Group. Apart from grading individual articles, we also rated the aggregate internal validity and external validity as well as the coherence (agreement of the results of the individual studies) for each of the key questions in the analytic framework. Appendix C presents the Work Group's detailed criteria for grading individual articles and rating aggregate validity and consistency of the articles reviewed.

In addition to these general criteria, we developed specific guidelines for this report. In diagnostic accuracy studies, we required that the studies had performed verification of screening results against an accepted criterion standard. Studies in which no criterion standard was used were excluded from this report. Studies that reported the results for only the portion of the sample that received the criterion standard were considered to have potential for spectrum bias and were also rated "fair."

For treatment studies, the failure to report results by intention-to-treat led to a grade of "fair" if the difference in sample size at the beginning and end of the trial was greater than 20% overall or if the drop-out rate was significantly different between the intervention and control groups.

Screening outcomes studies were included if they examined the impact of screening and feedback versus usual care on the diagnosis, treatment, or outcomes of depression.

Peer Review Process

We conducted a broad-based, external review of the draft SER. Outside reviewers were representatives of key primary care professional associations that have formal liaison ties to the USPSTF, a representative of the Canadian Task Force on Preventive Health Care, representatives of other professional societies, clinical experts in the area of depression, staff of the Agency for Healthcare Research and Quality, and representatives of other relevant federal agencies. Appendix A lists the names and affiliations of all peer reviewers.

Table 3. Depression: Inclusion and Exclusion Criteria

Category	Inclusion	Exclusion
General Inclusion and Exclusion Criteria		
Databases	MEDLINE + Cochrane	Other databases
Languages	English only	Other languages
Populations	Humans only	Animal studies
Study design	Original data	Letters, editorials, and non-systematic reviews that have no original data
Diagnostic Accuracy Inclusion and Exclusion Criteria		
Publication date	January 1994-December 1999	
Study design	Must have criterion standard	
Outcomes of interest	Sensitivity and specificity	
Study population	Primary care or community settings (including long-term care)	Hospital settings Psychiatry clinics
Adult Pharmacologic Therapy Inclusion and Exclusion Criteria		
Publication date	1994-December 1999	
Study design	Randomized controlled trials	
Study population	Primary care or community settings	Hospital settings Psychiatry clinics Children and adolescents
Adult Psychotherapy Inclusion and Exclusion Criteria		
Publication Date	1966-December 1999	
Study design	Randomized controlled trials	
Study population	Primary care or community settings Hospital settings Children and adolescents (including long-term care)	
Child and Adolescent Treatment Inclusion and Exclusion Criteria		
Publication date	1966-December 1999	
Study design	Randomized controlled trials	
Study population		

Table 4. Screening for Depression: Search Strategy Results

Step	Search Strategy for Screening	Number of Articles
1	Explode depression	26,043
2	Explode mass screening	41,430
3	Explode "sensitivity and specificity"	79,063
4	Explode reproducibility of results	46,916
5	2 or 3 or 4	153,961
6	Beck depression	1,393
7	CES-D	360
8	Diagnostic Interview Schedule	677
9	General Health Questionnaire	994
10	Hamilton Rating Scale	921
11	Hopkins Symptom Checklist	170
12	HSCL	86
13	SCL-90	550
14	Medical Outcomes Study	368
15	MHI-5	6
16	Mental Health Inventory	33
17	MADRS	214
18	Montgomery-Asberg	275
19	PRIME-MD	32
20	SCID	3,780
21	SDDS-PC	0
22	Zung	493
23	6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 22	9,626
24	Explode primary health care or explode family practice or explode ambulatory care	74,560
25	23 and 24	610
26	2 or 3 or 4	153,961
27	1 and 26	574
28	25 or 27	1,172
29	Limit 28 to (human and English language)	1,097

CES-D indicates Center for Epidemiology Study Depression Scale; HSCL, Hopkins Symptomatic Checklist; SCL-90, Symptom Checklist 90; MHI-5, Mental Health Index; MADRS, Montgomery-Asberg Depression Rating Scale; PRIME-MD, Primary Care Evaluation of Mental Disorders; SCID, Structured Clinical Interview for DSM-IIIR (or –IV); SDDS-PC, Symptom Driven Diagnostic System – Primary Care.

Table 4. Screening for Depression: Search Strategy Results (continued)

Step	Search Strategy for Psychotherapy/Treatment	Number of Articles
1	Explode depression (prevention and control, diet therapy, drug therapy, therapy)	8,576
2	Explode psycotherapy	77,340
3	Explode depression or explode depressive disorder	57,351
4	2 and 3	4,516
5	1 or 4	11,831
6	Limit 5 to (human and English language)	9,076
7	Limit 6 to randomized controlled trial	703
8	Explode randomized controlled trial or explode random allocation or explode single-blind method or explode double-blind method	103,349
9	6 and 8	503
10	7 or 9	971

Step	Search Strategy for Adult Psychotherapy	Number of Articles
1	Explode depression or depressive disorder	57,351
2	Explode psychotherapy	77,340
3	1 and 2	4,516
4	Limit 3 to (human and English language)	3,765
5	Limit 4 to randomized controlled trial	363
6	Explode randomized controlled trial or explode random allocation or explode single-blind method or explode double-blind method	91,676
7	4 and 6	82
8	5 or 7	389

Step	Search Strategy for Child Treatment	Number of Articles
1	Explode depression or depressive disorder	57,351
2	Limit 1 to (human and English language)	47,838
3	Limit 2 to randomized controlled trial	2,465
4	Explode randomized controlled trial or explode random allocation or explode single-blind method or explode double-blind method	103,349
5	2 and 4	2,449
6	3 or 5	3,549
7	Limit 6 to (newborn infant < birth to 1 month > or infant < 1 to 23 months > or preschool child < 2 to 5 years > or child < 6 to 12 years > or adolescence < 13 to 18 years >)	735

Table 5. Screening for Depression: Summary Results from Literature Searches and Reviews

Search and Review Results	All Searches
Number of Abstracts	
From literature search	1,942
From supplemental search	193
Reviewed	2,135
Excluded at abstract review phase	1,671
Included for full article review	464
Number of Articles	
Excluded after full review	202
Included in this SER	192
Included in Evidence Tables	70

SER indicates systematic evidence review.

Chapter 3. Results

We included detailed information, including demographic characteristics of the study population, descriptors of study design and setting, diagnoses and conditions of interest, criterion standard used for measurement (for screening topics), numerous outcome measures, and indicators of quality in the Evidence Tables in Appendix D. The tables cover, respectively, screening accuracy (41 entries in Evidence Table 1);[17,34-73] pharmacologic treatment (7 entries covering 9 publications in Evidence Table 2);[74-82] psychotherapeutic treatment (13 entries covering 15 publications in Evidence Table 3);[74,77-90] screening outcomes (13 entries in Evidence Table 4).[91-103] Some articles appear in more than one Evidence Table. (See the main glossary in Appendix B and the specialized glossary in Appendix D for abbreviations.)

Key Question 1: Accuracy of Screening Tests for Depression

For screening to be effective, then reliable, accurate, feasible, and acceptable screening methods must be available. On the advice of the USPSTF liaisons, we focused the review on diagnostic accuracy and the ability of the instruments to classify patients correctly as depressed or well. We comment briefly on feasibility and acceptability but have not systematically reviewed the literature in those areas.

Screening Accuracy in Adults

Multiple reliable depression screening instruments are available for adults.[3,104] Numerous studies have examined the diagnostic accuracy of screening tests for depression in adults. We identified 33 articles that had been published from January 1994 to August 1999 and 8 older articles published from 1966 to December 1993 that examined the sensitivity and specificity of 13 different screening instruments against a criterion standard for the diagnosis of depression. The following sections examine several aspects of the diagnostic performance of the screening tests in different populations, including community, general practice, or primary care patients, the elderly, children and adolescents, and special populations. This information is then used to estimate the diagnostic consequences of screening for depression in these different populations.

As with all screening procedures for diagnostic tests, a positive screen for depression does not make a diagnosis of a depressive illness. Unlike many other disorders, depression has no universally accepted criterion standard. Several diagnostic instruments have been used to define the presence or absence of depression (Table 6). The most feasible standard in primary care is most likely a comparison of the patient's symptoms with criteria listed in the *Diagnostic and Statistical Manual (DSM)* particularly DSM-IV for depressive illnesses.[13] A specific DSM-IV Primary Care Version has been tailored to be a useful aid in diagnosing mental disorders in primary care.[105]

After confirming that a patient who screens positive meets the diagnostic criterion for a specific depressive illness, the clinician must consider other potential causes of depression (such as hypothyroidism, depression due to medication or substance use, vitamin deficiencies, or electrolyte imbalances). Additionally, the clinician must take into account other psychiatric illnesses that can

present with depressive symptoms (Table 7). Such considerations would require additional history collection and possibly laboratory tests. Should 1 of the additional causes of depressive illness be identified, first steps at treatment may be directed at this underlying etiology. Otherwise, treatment for the depressive illness (whether in the primary care setting or by referral to a mental health professional) can be initiated.

The 41 studies in Evidence Table 1 (listed in alphabetical order by author) (Appendix D) include 24 studies of adults in community or primary care settings, 12 articles that address screening in older adults, and 5 studies performed in special populations. The primary screening instruments used in these studies are the Center for Epidemiologic Study Depression scale (CES-D), used as the main instrument in 13 studies; the Geriatric Depression Scale (GDS), used in 6 studies; the General Health Questionnaire (GHQ), used in 4 studies; the Beck Depression Inventory (BDI) used in 3 studies; the Zung self-depression screener (SDS) used in 3 studies; the Symptom Driven Diagnostic System – Primary Care (SDDS-PC) used in 2 studies; the Self Care-D, used in 2 studies; the depression screening module of the Medical Outcomes Study (MOS) used in 2 studies; and 6 instruments that were used in 1 study each. Table 8 describes the basic characteristics of these instruments.

The majority of the identified studies (23/34) examined sensitivity and specificity for major depressive disorder, defined by a variety of criterion standards, many of which are based on DSM-III or DSM-IIIR criteria. Eight studies examined screening accuracy for depression without specifying a specific disorder. One study each specifically examined screening accuracy for minor depression, subsyndromal depressive disorders, "depression NOS," or a "significantly depressed state." Three studies could not be characterized. Some studies used more than 1 disease definition.

Chapter III: Results

Older Studies

Mulrow et al[104] systematically reviewed the performance of screening tests for depression conducted between 1966 and February 1994. They identified 15 published and 4 unpublished articles that met their inclusion criteria, which required that the outcome status of at least 50% of the subjects be verified by an acceptable criterion standard examination. Eleven of these articles met our inclusion criteria as well and appear in Evidence Table 1.[39,40,42,43,47,60,63,69,73,97,102]

To summarize performance, Mulrow et al[104] calculated the average sensitivity and specificity for the included articles (based on the usual cut-points for each instrument) and constructed a summary receiver operating characteristics (ROC) curve. The overall sensitivity was 84% (95% confidence interval [CI], 79% to 89%), and overall specificity was 72% (95% CI, 67% to 77%). These values translate to a positive likelihood ratio (LR) of about 3 and a negative LR of 0.2. Results did not differ substantially based on the degree of verification bias. The included instruments were easy to administer and complete, and they had been written at either easy (third to fifth grade) or average (sixth to ninth grade) literacy levels.

General Primary Care Populations

We identified 23 newer articles that Mulrow et al had not included. Six of the 23 newer studies were conducted in primary care settings in nonelderly or mixed populations.[36,51,61,62,65,71] Klinkman et al[51] found that the CES-D had a sensitivity of 81% and specificity of 72% for scores above 15, compared with a gold standard diagnosis based on a Structured Clinical (Diagnostic) Interview (SCID) for DSM-IIIR or -IV. Parkerson and Broadhead[61] found a similar level of performance for the Duke AD screener: 81% sensitivity and 64% specificity for scores greater than 30. Salokangas et al[62] found that The Depression Scale (DEPS) performed reasonably well (sensitivity, 74%; specificity, 85% for scores

greater than 8). Bashir et al[36] tested the GHQ in a random sample of British general practice attenders and found a sensitivity of 76% and a specificity of 74%. Steer et al[65] reported that the BDI performed extremely well (sensitivity, 97%; specificity, 99%) against a less rigorous criterion standard, the mood module of the Primary Care Evaluation of Mental Disorders (PRIME-MD).

The study by Whooley et al[71] deserves special comment. They examined the performance of multiple screening tests, including the CES-D, BDI, and MOS, as well as a new two-item screener that included only questions about depressed mood and anhedonia, in a population of veterans (97% men) from an urgent care setting. The two-item screener (sensitivity, 96%; specificity, 57%; area under the ROC curve, 0.82) performed nearly as well as the CES-D and MOS (area under the ROC curve, 0.89 for each). Shorter versions of the CES-D and BDI also performed well.

Overall, these newer studies had sensitivity and specificity results similar to those found by Mulrow et al.[104] Sensitivity with some of the newer short screeners was slightly improved, with specificity similar to that of older instruments.

Elderly Populations

Twelve newer studies (Evidence Table 1) specifically examined the performance of depression screening instruments in older adults, including 6 using the GDS, 3 using the Self Care-D, and 3 using the CES-D (Table 9). The age limits used to define "elderly" varied; 1 study included adults older than 50 years of age, another enrolled only those older than 75 years, and others fell in between. The settings included community-based recruitment, primary care clinics, geriatric assessment clinics, patients' homes, and a nursing home.

Each of these screening instruments demonstrated relatively good test performance characteristics (Table 9), with sensitivities generally 80% to 95% and specificities of 70% to 85%. Each

instrument showed modest variation between studies. In general, confidence intervals were not calculated for the sensitivity and specificity estimates, and few studies calculated area under the ROC curves. Two studies, Gerety et al[45] and Lyness et al, [56] compared the GDS and CES-D instruments; both found that the GDS performed better. In Gerety et al,[45] the area under the ROC curve was 0.91 for the GDS and 0.85 for the CES-D. According to Lyness et al, [56] each instrument had similar performance for major depression, but the GDS performed better for "minor" depression. None of the studies compared the Self Care-D with either the CES-D or the GDS.

Special Populations

We identified 5 studies of depression screening in special populations that met our inclusion criteria (Evidence Table 1). Geisser et al[44] tested the CES-D in a pain clinic. The criterion standard was a clinical interview with a psychologist using DSM-IV criteria. They found a 33% prevalence, a sensitivity of 82%, and a specificity of 73% using a score of 27 or greater to define a positive screen.

Holcomb et al[48] examined the performance of the BDI in an obstetrics and gynecology setting. They used the Diagnostic Interview Schedule (DIS) as a gold standard and found an 11% prevalence of current depression. A BDI score of 16 or greater had 83% sensitivity and 89% specificity for depression.

Irwin et al[50] used the CES-D in a community-based sample of adults with known physical illness. They compared their screening results against the SCID as a criterion standard. Scores of greater than or equal to 4 had 99% sensitivity and 84% specificity for depression.

Leung et al (1998) studied the performance of the Zung SDS in Chinese family practice patients in Taiwan.[53] This team reported that SDS scores of greater than or equal to 55 had 67% sensitivity and

90% specificity for depression when compared against a diagnosis by a physician using DSM-IV criteria.

Lustman et al (1997) examined the BDI in patients with diabetes, using the DIS as a criterion standard.[55] The prevalence for major depression was 37%, and a BDI score greater than or equal to 13 had 85% sensitivity and 88% specificity.

Summary of Screening Accuracy in Adults

Several depression screening instruments appear to detect depression effectively. Recent research has shown that shorter screening tests, including simply asking 2 questions about depressed mood and anhedonia, appear to detect a large majority of depressed patients; in some cases, they perform better than the original instruments from which they had been derived.

In general, sensitivity results were good to excellent and specificity results were moderate to good; with commonly used cut-points, typical values were 80% to 90% for sensitivity and 70% to 85% for specificity. If the prevalence of major depression is estimated to be between 5% and 15% in primary care settings, the positive predictive value (probability of depression after a positive test) would be 25% to 50% (Table 10). Thus, more than half of patients who screen positive will be false positives for major depression. Some of these "false positives" may be patients with minor depression or dysthymia. People with positives screens require further diagnostic questioning before clinicians apply a diagnostic label and suggest a treatment plan.

One problem with depression screening instruments is that continuous data (ie, scores on the instruments) are dichotomized into positive and negative results at an arbitrary cut-off value and then used to calculate sensitivity and specificity (as well as positive and negative likelihood ratios) for that

cut-off. With this approach, valuable information is lost because all scores above the threshold are counted equally (similarly, all below the threshold are also treated the same).

Some studies in this report partially overcome this problem by providing information on area under the ROC curve, which quantitates overall performance by producing a score between 0.50 (no information) and 1.0 (perfect information). An even more useful technique is to calculate stratum-specific likelihood ratios (SSLRs) for ranges of scores on an instrument. The SSLR for the result of the screen is multiplied by the pre-test odds to give the post-test odds. Furukawa et al[106] calculated SSLRs for the CES-D using data from Japanese psychiatric hospitals and clinics. Scores of 0 to 29 were associated with an SSLR of 0.35; scores of 30 to 49 were associated with an SSLR of 2.3; and scores over 50 were associated with an SSLR of 11.7 (Table 11).

Another difficulty in measuring the accuracy of screening instruments comes when trying to interpret specificity. Instruments used in some studies to detect major depression may count subjects with subsyndromal depressive illnesses as false positives. A true measure of specificity would count as false positives only those patients who are free from any significant depressive illness but who screened positive, because patients with subsyndromal illnesses may also benefit from treatment or more careful observation. Patients with other important and treatable disorders such as substance abuse, anxiety disorders, complicated grief reactions, or bipolar disorders may also be counted as false positives in some studies, but they might well be identified by the more careful and in-depth assessment that would presumably follow a positive screen. If, however, treatment for depression is initiated on only the basis of screening positive, then patients with other related illnesses may receive suboptimal care.

Using Risk Factors to Identify Patients with Depression

Because the prevalence of depression is only 5% to 10% in primary care settings, some experts have suggested that the presence of known risk factors for depression be used to determine who should or should not be screened—a strategy of selective screening. Although, intuitively appealing, most common risk factors for depression perform relatively poorly in discriminating patients who are depressed from those who are not depressed. Conde et al[107] demonstrated that most common risk factors have positive likelihood ratios (LR) between 1 and 2 and negative likelihood ratios between 0.5 and 1, suggesting low predictive ability (Table 12).

Other factors, such as a previous history of depression or concurrent diagnosis of panic disorder or generalized anxiety disorder, have positive LRs greater than 10; their presence warrants further investigation for depression, perhaps including a diagnostic interview. Their absence, however, does not significantly change the likelihood of depression.

Depression screening tools have a positive LR of approximately 3 and a negative LR of 0.2, demonstrating that they perform better than most of the common demographic risk factors. Based on these data, a strategy of selective screening does not appear to be superior to simply performing (or asking the patient to perform) one of the brief screening tools. In patients with previous depression or a current anxiety or panic disorder, directly proceeding to a full diagnostic interview may be warranted instead of initial screening.

Screening Accuracy in Children and Adolescents

The identification of depression in children and adolescents has not been as well studied as in adults. Increasing recognition of the important burden of depressive illness and its sequelae in children

and adolescents has led to greater attention to means to identify, prevent, and treat mood disorders in this vulnerable population.

Depressive illnesses may have different clinical characteristics and presentations in children and adolescents than in adults. Child and adolescent psychiatrists have developed several structured diagnostic interviews that have been used to characterize and diagnose depression in youth, but they are too long and complex for routine use by primary care providers. Apart from the DSM, these include versions of the Child Assessment Schedule (CAS), Diagnostic Interview for Children and Adolescents (DICA), Diagnostic Interview Schedule for Children (DISC), and Schedule for Affective Disorders and Schizophrenia for School-age Children (K-SADS).[108] These instruments are often used as criterion to make the diagnosis of depression.

The use of different criterion standards is critical to the appraisal of screening test performance as these standards have their own limitations with regard to sensitivity and specificity that affect the evaluation of screening tools.

Only a small number of studies have addressed screening test performance in ambulatory, nonpsychiatric pediatric populations that are generalizable to primary care. The screening tools that have been evaluated most commonly are reviewed below and summarized in Table 13.

Beck Depression Inventory (BDI)

Two studies looked at performance of the BDI in outpatient samples referred for psychiatric care;[109,110] most subjects were adolescents. Sensitivity was 48%, 86%, and 89% with corresponding specificities of 87%, 82%, and 88%. Positive predictive values were high (63%, 83%, and 93%) because of the high prevalence of depression in these referred patients.

Three studies used the BDI in general school samples of adolescents. The largest study included 1,704 Oregon high school students and used a BDI of ≥11 for females and ≥15 for males to assign a diagnosis of current depression (according to DSM-III criteria).[111] Sensitivity was 84%; specificity, 81%. Positive predictive value was 10% and negative predictive value 99.5%. A small sample of 49 adolescents from a school population was a part of a study using the BDI to identify DSM-III major depression.[112] Using a cut-off of 16, the investigators reported 100% sensitivity and 93% specificity for the BDI. Prevalence of depression was 10% (5/49 adolescents); positive predictive value was 61%. The third study of adolescent students used a BDI of ≥16 to assess lifetime history of DSM-III major depression and dysthymia.[113] For depression, sensitivity and specificity were 77% and 65%, respectively. Prevalence of depression was 4%; positive predictive value was 8%. For dysthymia, sensitivity and specificity were 71% and 64%, respectively. Prevalence of depression was 5%; positive predictive value was 10%.

Finally, the only study conducted in a general primary care setting used a version of the BDI, the Beck Depression Inventory for Primary Care (BDI-PC) to assess major depression during 100 adolescent health maintenance examinations.[114] A BDI-PC cut-off of 4 yielded a sensitivity of 91%, a specificity of 91%, and a positive predictive value of 56% for the population with a high prevalence of 11%.

Center for Epidemiological Studies - Depression Scale (CES-D)

The CES-D is a 20-item scale developed for adults. The CES-D in children did not correlate well with the Children's Depression Inventory (CDI) and did not discriminate depressed and nondepressed patients adequately for use in children.[115]

Two studies have described CES-D screening accuracy for depression in large school-based samples of adolescents. Roberts et al[111] looked at CES-D scores in the Oregon sample that also used the BDI. Investigators applied a cut-off of 22 for males and 24 for females to identify current depression (DSM-III criteria) in 1,704 adolescents. Sensitivity was 84%; specificity was 75%; and positive predictive value was 8%.

Garrison et al[116] used a subsample of 332 students identified in a larger survey of adolescents in the Southeastern United States. Using various cut-off points, the researchers found that optimal screening characteristics for depression occurred at a cut-off point of 12 for males and 22 for females. For males, sensitivity was 85%, specificity was 49%, and positive predictive value was 13%. For females, sensitivity was 83%, specificity was 77%, and positive predictive value was 25%. Screening performance of the CES-D was also assessed for dysthymia using a cut-off of 16 for males and 20 for females. For males, sensitivity was 75%, specificity was 67%, and positive predictive value was 14%. For females, sensitivity was 100%, specificity was 67%, and positive predictive value was 8%.

The CES-D also has a version for children, the CES-DC. In 1 study of the CES-DC using a cut-off of 15, Fendrich et al[117] found the CES-DC to have a sensitivity of 71% and a specificity of 57%.

Other Screening Instruments

In a population of adolescents referred for psychiatric care, Angold et al[118] tested the Short Mood and Feeling Questionnaire in a mixed sample of 173 children and adolescents. They used the DISC as a criterion standard and found sensitivity of 70% and specificity of 85%.

Several other screening instruments have been used in children and adolescents, but most have not been used to screen a primary care sample of pediatric patients. These other tests include the Children's Depression Inventory (CDI), Child Depression Scale (CDS), Children's Self-report Rating Scale (CSRS), Depression Self-Rating Scale (DSRS), and Reynolds Adolescent Depression Scale (RADS). Studies of these scales have reported validation in psychiatric inpatient and referred samples, and so these instruments may be useful in some settings. However, the studies either do not report data in primary care populations or do not describe test performance results to address use as general screening tools.[119]

The Pediatric Symptom Checklist (PSC) and Child Behavior Checklist (CBCL) have been shown to be feasible to implement in primary care practice and have relatively good sensitivity and specificity as a general screen of mental health needs. These tests may increase awareness of unrecognized psychosocial problems; however, they do not appear to perform well in identifying specific individual diagnosis such as depression.[120,121]

Special Populations

Children with comorbid psychopathology or chronic medical illness and other pediatric subpopulations have been reported to have a higher prevalence of depressive disorders than the general population. Special populations may be candidates for targeted screening, but few studies report screening accuracy results. Sensitivity and specificity in psychiatric inpatient or outpatient groups are

generally similar to the results presented above, although predictive values will be higher because of the higher prevalence of depression.[122]

One study in a population of chronically ill pediatric patients (an important subset of pediatric patients with higher prevalence of depressive disorders) evaluated test performance of the CDI, the PSC, and the CBCL.[123] The authors found a high prevalence of depression and mental disorders and relatively good specificity of the measures at detecting depression, anxiety, or both (78% to 96%). They concluded, however, that low sensitivity of the tests (26% to 55%) limited their clinical usefulness for this patient population. Other, better performing depression scales have not been tested in children with chronic illnesses.

Summary of Screening Accuracy in Pediatric Populations

The existing literature suggests that screening instruments for depression in adolescents that have been tested in community or primary care settings perform reasonably well. They produce sensitivity values ranging from 75% to 100% and specificity values from 70% to 90%, values similar to those found in adults, although there are fewer studies and fewer total subjects. Fewer data are available for children. The prevalence of disease and the positive predictive value in children are quite low, but the values rise in adolescents. Like adults, those who screen positive should undergo a more rigorous diagnostic interview before being labeled as depressed.

Key Question 2: Outcomes of Treatment for Depression in Primary Care Settings

Treatment of depression in primary care patients can involve antidepressant medication, psychotherapy, or a combination of the 2. Additionally, educational and quality improvement

interventions directed at the patient, clinician, or care system have been applied to improve the effectiveness of treatment for depression. As part of examining whether screening for depression is beneficial in the primary care setting, we sought to determine whether treatment for depressive disorders in primary care patients can improve outcomes, including depressive severity, functional status, and health care utilization. We first address the evidence for treatment of adults, including the elderly and special populations, and then examine the evidence for children and adolescents.

Treatment of Depression in Adults

The Depression Guideline Panel of the Agency for Health Care Policy and Research (AHCPR) systematically reviewed literature published through December 1990 and performed a meta-analysis on 7 of the 24 extant randomized controlled trials (RCTs) conducted with primary care patients, all of which were pharmacologic interventions.[3] Only 1 of the 7 studies involved a selective serotonin reuptake inhibitor (SSRI); the preponderance of medication trials involved tricyclic antidepressants (TCAs) (4 studies) or heterocyclic agents (5 studies). The overall drug efficacy was 57.8%; the placebo response rate (included in 3 of the studies) was 35.6%.

We updated the AHCPR review using 3 more recent systematic reviews and a search of articles published from 1966 through December 1999. We included articles that provided clinical outcome measures and had been performed in a primary care setting. The systematic reviews included a review of treatment in primary care, which examined 28 articles;[124] a review of the treatment of dysthymia with 15 articles;[125] and a review of treating depression in patients with physical illness that identified 18 articles.[126] One study had been included in a review by Mulrow et al[124] and in a separate Cochrane review by Lima and Moncrieff;[125] another study[127] was included in both of the reviews by Mulrow and

her colleagues[124,128] and the Gill and Hatcher[126] Cochrane review. These first 2 reviews involved antidepressant trials and did not include studies of psychotherapy. Gill and Hatcher assessed trials involving antidepressant drugs, 3 of which had had a concomitant psychotherapy. As their analysis was limited to the effects of antidepressants, we will report its results only in regard to antidepressant outcomes. In addition to the articles from these reviews, our literature searches identified 19 other trials of treatment for depression. Data from these trials are included in the Evidence Tables in Appendix D.

Across all these sources, we identified a total of 78 studies for review in this SER (the 59 articles from the 3 previous systematic reviews, plus the 19 newly identified articles). Of these 78 studies, 73 directly tested an antidepressant or psychotherapeutic treatment (or both): 60 tested an antidepressant alone, 5e involved both an antidepressant and psychotherapeutic intervention (3 of which looked at the effects of a combined intervention), and 8 tested a psychotherapy intervention alone.

The remaining 5 studies involved educational or quality improvement interventions. Four involved multidisciplinary collaboration and education directed at the patient, clinician, and system of care.[129-132] One assessed the effect of drug counseling and information leaflets for patients on medication adherence and depressive severity.[133]

In the following sections, we examine the outcome of various forms of interventions for depression, including antidepressant medications, psychotherapy, and educational or quality improvement interventions.

Pharmacologic Interventions

Details about pharmacological treatment studies that met our inclusion criteria can be found in Evidence Table 2 (Appendix D). The discussion below is presented first for large-scale reviews (for

which data have not been provided in Evidence Tables) and then for other studies; for the latter, the 6 entries in the Evidence Table (which cover 8 publications) are presented in reverse chronological order.

Results from Large-Scale Reviews. Mulrow and colleagues[124] completed a systematic review from 1980 through January 1998 that evaluated RCTs involving depressed primary care patients that compared the efficacy of "newer" antidepressants to that of other pharmacologic or psychosocial interventions or to placebo. They identified 28 trials involving 5,940 primary care patients; these covered major depression (14 studies), dysthymia (2 studies), or another form of depressive illness ("depression requiring treatment," "depressive illness," "endogenous depression," or mixed anxiety-depression, 12 studies). Average response rates were 63% for newer agents and 35% for placebo (rate ratio,1.6; 95% CI, 1.2 to 2.1). This magnitude is equivalent to that noted in the AHCPR Depression Guideline Panel review. The response rate was similar for newer and older agents (rate ratio, 1.0; 95% CI, 0.9 to 1.1). The drop-out rate because of adverse effects was significantly lower for newer agents than for the TCAs (8% vs 13%; absolute risk reduction [ARR], 4%; 95% CI, 0% to 7%) although the overall drop-out rate did not differ.

Although response rates appeared similar across different depressive disorders in the Mulrow et al review, there were too few studies in each group to exclude a modest difference. The most frequent diagnosis of interest, as noted above, was major depression; the remaining other forms of depressive illness may include dysthymia, minor depression, or some additional subthreshold depressive illness. Only 2 trials clearly addressed dysthymia, making conclusions about its pharmacologic treatment in primary care settings less clear.

A recent systematic review by Lima and Moncrieff[125] of all RCTs comparing drugs and placebo for dysthymia from 1966 through January 1997 may provide additional important information for that condition. The review identified 15 studies involving 1,964 patients, with trial duration ranging from 4

to 12 weeks. One study had been conducted in a general practice;[134] the remainder had been performed in a mixture of community, inpatient, and outpatient mental health care settings. The analysis made no distinction among the different settings. Antidepressants were 56% more likely to reduce dysthymic symptoms than placebo (risk ratio [RR], 1.56; 95% CI, 1.43 to 1.67). Treatment response did not differ by class of antidepressant. Patients treated with TCAs were more likely to report adverse events than those on placebo, but they were not significantly more likely to drop out.

Gill and Hatcher[126] recently reviewed all RCTs published through June 1998 that had examined antidepressant interventions in depressed patients who also had a physical illness. Settings were not limited to primary care. The 18 studies in this review involved a total of 838 patients. Study subjects had a wide range of medical illnesses (5 studies examined patients with human immunodeficiency virus [HIV] infection; 3 with stroke, 2 with cancer, 2 with mixed medical diagnoses; and 1 each with diabetes, head injury, heart disease, lung disease, multiple sclerosis, and renal disease). Patients could be diagnosed as depressed by any criterion. Those treated with antidepressants were significantly more likely to improve (52%) than those given placebo (30%) (odds ratio [OR], 0.37; 95% CI, 0.27 to 0.51). Six of the 18 trials involved a diagnosis of major depression by structured clinical interview; for this subgroup, the effect was similar (OR, 0.35; 95% CI, 0.22 to 0.55).

Results from Additional Trials. We identified 6 additional RCTs for depressive illness in primary care involving the use of antidepressants (Evidence Table 2). Five of these studies reported benefit for antidepressant intervention compared to either placebo[74,76,77] or usual care;[78,81] 1 study compared a cognitive-behavioral therapy (CBT) with antidepressant treatment and with a combination of the 2 interventions.[75] Five studies involved patients with diagnoses of major depression.[74,75,77,78,81] [74,78,135] Strict intention-to-treat analyses were conducted in 4 of the trials;[74,76,78] [75] Mynors-Wallis et al[77] and Scott and Freeman[81] analyzed only those subjects who received at least some treatment.

Appleby and colleagues[74] compared fluoxetine (an SSRI) to placebo (both with either 1 or 6 CBT sessions) for women screened originally on obstetrics wards who had postpartum major or minor depression 6 to 8 weeks after delivery. Patients with major depression were in the majority in each group (60.5% for fluoxetine and 56.8% for placebo). No distinction was made in the analysis between those with major and minor depression. Of note is that a substantial proportion of women who fulfilled study criteria did not enter the trials; of 188 with confirmed diagnoses of depression, only 87 agreed to enter the trial. The fluoxetine group averaged a 66.9% decrease in Hamilton Depression Rating Scale (HAM-D) scores at 12 weeks compared to a 54.0% decrease for the placebo group. The statistical significance of this difference was not reported. Among subjects completing treatment (70% of the randomized sample), treatment appeared to lead to significant improvement, with the fluoxetine group having a 78% decrease in HAM-D scores compared to a 61% decrease in the placebo group (*P* = "significant"). The fluoxetine and CBT treatments did not appear to interact significantly, and no advantage was found for those receiving both interventions.

Schulberg et al[78] compared primary care patients receiving the TCA nortriptyline alone to those receiving only interpersonal psychotherapy (IPT) and to those receiving usual care. All subjects had a rigorously diagnosed major depressive disorder that used a three-stage assessment. Of 7,652 waiting-room patients completing the CES-D screen, 1,492 scored above a cut-off of 22 and were not currently being treated for a mood disorder. These patients were eligible for the next phase, consisting of diagnostic confirmation using the DIS Depression section;[136] of the 1,059 patients completing this section, 678 (64%) met the criterion for a major depression. Of these 678 patients, 403 (59%) completed the third stage, in which a consultation-liaison psychiatrist confirmed the depression of major depression and confirmed protocol eligibility. Psychiatrists judged 283 (70%) of those they evaluated as protocol eligible; 276 of these agreed to a randomized treatment assignment.

Patients in the nortriptyline group had weekly or biweekly visits until the acute phase of treatment had ended and monthly visits thereafter. Of those treated with nortriptyline, 48% had recovered at 8 months, as had 18% of those treated with usual care. There was no significant difference in outcome between the medication and the psychotherapy intervention (48% with nortriptyline, 46% with IPT).

Mynors-Wallis et al[77] compared amitriptyline (also a TCA) or psychotherapy (problem-solving therapy) to placebo in patients with major depression. As with the other treatment arms, the amitriptyline group was offered 6 treatment sessions over 3 months, and treatments were usually given at the patient's home or local health center. All 3 groups had 3.5 hours of contact time (about 35 minutes per session). An intention-to-treat analysis was not performed, as outcomes were measured only for those attending 4 or more sessions. Recovery by 12 weeks was seen for 52% of the patients receiving amitriptyline, 60% of those receiving problem-solving therapy, and 27% of those receiving placebo.

Scott and Freeman[81] compared amitriptyline prescribed by a psychiatrist, cognitive therapy provided by a psychologist, or counseling given by a social worker to usual care for patients with major depression. The amitriptyline group averaged approximately 240 minutes (4 hours) of total contact time over the 16-week course of treatment; the usual care group (treated by general practitioners) averaged 50 minutes. An intention-to-treat analysis was not performed; of those randomized to antidepressant treatment, 5 (16%) never began the intervention and were not included in the results. Each of the 4 groups had marked improvement of their symptoms over the four-month study period: 58% of the amitriptyline group had recovered at 16 weeks, compared to 48% of the usual care group.

Malt et al[76] compared sertraline (an SSRI) and mianserin (a newer heterocyclic agent) to placebo for patients with 2 weeks of depressive symptoms that were judged to be "severe enough to require

treatment." Patients were seen weekly for the first month and then with a gradually lengthening follow-up interval for a total of 10 visits over a 24-week period. This study employed an effectiveness design that attempted to reproduce more accurately the clinical situation in primary care by not excluding patients with concomitant medical illness and not excluding those experiencing a placebo response. Clinically significant responses occurred in 61% of those receiving sertraline, 54% of the mianserin group, and 47% of the placebo group. The number needed to treat (NNT) for sertraline was 7. Of note, 86% to 89% of all subjects met criteria for a major depressive episode, although only 18% of all subjects were considered profoundly depressed on the Clinical Global Impression (CGI) scale.

Mynors-Wallis and colleagues compared 6 sessions of medication-only alone treatment (provided by research general practitioners [GP], [137] not the patients' usual GP) to 6 sessions of problem-solving (PS) psychotherapy (by a trained research GP or a trained research nurse) and to a combination of medication and psychotherapy treatment [75]. A usual care or placebo group was not included. GPs referred subjects with a depressive illness requiring treatment; those included had had at least 4 weeks of probable or definite major depression as confirmed by Research Diagnostic Criteria.[138] The number of actual contact hours for the medication-only group was not given. Each of the 4 groups showed substantial improvement. In an intention-to-treat analysis, 67% of the medication-only group had recovered (HAM-D $\leq$7) at the end of the 12-week treatment course; 56% remained recovered at the 1-year mark. The medication group did not differ significantly from either the problem-solving groups or the combination treatment group. Although not statistically significant, the medication-only and combination treatment groups lost 17% of their patients to follow-up, compared to 36% of the PS-GP group and 22% of the PS-nurse group.

Of note, all but 1 of the trials included in this review were efficacy trials, conducted under ideal conditions with much closer and more frequent follow-up than is routine in primary care. Such results may not generalize to normal primary care practice. Simon et al[139] initially randomized patients to SSRI (fluoxetine) or tricyclic (desipramine or imipramine) antidepressant treatment and then allowed subsequent antidepressant management to be undertaken by the primary care physician. In this effectiveness trial, the proportion of patients continuing the original medication was significantly higher for the fluoxetine group (80% over the 6-month period) than for either the desipramine group (52% overall) or the imipramine group (57% overall) ($P< 0.001$ for each comparison at one-month, three-month, and six-month follow-up), although the proportion in each group continuing any antidepressant was similar at each assessment. These findings suggest that patients are more likely to switch treatment from tricyclic agents than from SSRIs.

Psychotherapy Interventions

Evidence Table 3 (Appendix D) presents information on 13 studies of psychotherapy (covering 15 publications); the entries appear in alphabetical order. We present the discussion below in terms of studies on major depression, minor depression, dysthymia, and/or other depressive conditions.

Major Depression. Eleven of the 13 studies of psychotherapy involved patients with major depressive disorders (Evidence Table 3). The five studies that also included medication trials are described with respect to the medication efficacy in the previous section; the outcomes of psychotherapy are described below. As shown in Table 14, the more effective interventions tended to have a more highly structured intervention than is typically the case; that is, the more effective approaches were well formulated, limited in time, and standardized in application, and they tended to have clearly defined goals and stages. Only 6 studies used intention-to-treat approaches.[74,75,78,85,88,89] The studies are

reviewed below in the order of decreasing magnitude of effect and decreasing stringency of outcome measures (eg, recovery is more stringent than reduction in depressive severity).

Mynors-Wallis and colleagues[77] compared 6 treatment sessions of well-structured PS therapy, guided by a treatment manual and provided by either an experienced psychiatrist or trained GPs, against usual care. As noted earlier, all groups (including the pharmacologic arm) had 3.5 hours of contact time. No intention-to-treat analysis was done. At 12 weeks, 60% of those in the PS group had recovered compared to 27% of those in the usual care arm.

Holden et al[84] compared counseling by health visitors to usual care in a trial involving women with postpartum major or minor depression. The health visitors had limited training and provided 8 weekly sessions of an unstructured, supportive intervention of at least 30 minutes duration. The therapy was not administered according to any standardized manual or approach. Approximately two-thirds of each group had patients with major depression at the start of the trial. No intention-to-treat analysis was performed: 55 women were randomized; of these, 50 completed the trial and were included in the results. At 13 weeks, 69% of the health visitor group and 38% of the usual care group had neither major or minor depression as assessed by Research Diagnostic Criteria. The results did not distinguish between major and minor depression.

Scott et al[89] compared six 30-minute cognitive therapy sessions to usual care. No manual was used, but the treatment was relatively well structured and a random sample of psychotherapy tapes were reviewed to ensure quality. In an intention-to-treat analysis, at 7 weeks 62% of the group randomized to cognitive therapy had recovered, as had 33.3% of those with usual care. Follow-up was also assessed at 58 weeks in a treatment-completer analysis, and the psychotherapy arm had significantly lower depressive severity (HAM-D=6.1) than the usual care arm (HAM-D=10.7). Of note was the large attrition rate at 1-year follow up (16/28 in cognitive therapy group, 8/24 in usual care group).

Katon et al[85] evaluated a brief CBT intervention as part of a multi-faceted primary care intervention for major or minor depression that included on-site education and consultation for physicians about antidepressant and behavioral treatment of depression. Analysis was intention-to-treat. The psychotherapy intervention was geared toward improving medication adherence and consisted of 4 to 6 meetings with a psychologist for a total of 2.5 to 3.5 hours plus 4 telephone contacts. Outcomes for patients involved in this program were compared to outcomes for patients receiving usual care for the same conditions. For major depression, 70.4% of those receiving the multi-faceted intervention involving CBT had a greater than 50% decrease in depressive severity at 4 months compared to 42.3% of those in the usual care group. The effect size was smaller for minor depression (66.7% improved with therapy, 52.8% with usual care) and did not reach statistical significance.

Mynors-Wallis and colleagues[75] compared 6 sessions of well-structured PS therapy by a trained GP to 6 sessions by a trained nurse, to antidepressant medication alone, and to a combination of the medication and PS therapy. Therapy was provided in either the patient's home or the local health center. The first PS sessions lasted 1 hour; subsequent sessions lasted 30 minutes. Patients receiving PS therapy alone had a mean number of 4.6 treatment sessions (2.8 hours total contact time); those receiving combination treatment had a mean number of 5.2 PS treatment sessions (3.1 hours contact in addition to medication management time).

After 3 months of treatment, 51% of the PS-GP group and 54% of the PS-nurse group had recovered (HAM-D $\leq$7), compared to 67% of the medication alone group and 60% of the combination group. At 1-year follow-up, 62% of the PS-GP group had recovered, as had 56% of the PS-nurse group, 56% of the medication alone group, and 66% of the combination group. As described before, the 4 groups did not differ significantly in terms of rate of recovery, suggesting that combination treatment for

routine depressive illness in primary care is no more effective than a single intervention, and that outcomes will not differ between PS therapy delivered by a trained GP and that delivered by a trained nurse. These findings are in contrast to recent research in specialty settings suggesting benefit for combination in certain situations, such as preventing recurrence of depression in a geriatric psychiatry setting[140] and in treating chronic depression in an outpatient psychiatry setting.[141]

Schulberg et al[78] compared 16 weeks of IPT delivered by doctoral-level, experienced therapists using a well-structured, standardized protocol to usual care. In an intention-to-treat analysis, 46% of those randomized to the IPT group recovered as did 18% of the usual care group.

Ross and Scott[88] tested individual cognitive therapy (consisting of 12 sessions lasting 45 minutes over 3 months) or group cognitive therapy (12 sessions lasting 90 minutes over 3 months) to usual care. All treatment was delivered by the same experienced social worker; it is unclear if the treatment was structured. All groups appeared to improve. Following the 3-month intervention period, those receiving cognitive therapy appeared to have significantly greater reductions in depressive severity than usual care (32% reduction on HAM-D vs 17%, $P < 0.01$; intention-to-treat analysis). The individual and group forms of treatment did not differ significantly. For the subset of patients who had been assessed 12 months after completing treatment, benefits appeared to be maintained, although no usual care group was available for comparison.

The Appleby et al[74] study did not distinguish between patients who developed major or minor depression postpartum. Those receiving 6 sessions of minimally structured CBT totaling 3.5 hours by a nonspecialist with minimal training experienced a 64% decrease in the HAM-D score at 12 weeks compared to a 57.7% decrease for those receiving a single, 1-hour CBT session from a nonspecialist. These results were slightly less robust than the pharmacologic intervention. Again, significance was not reported for the intention-to-treat analysis. For the patients completing treatment (30% attrition), HAM-

D scores decreased by 76% for those with 6 sessions, a significantly greater decrease than the 66% drop for the 1-session group.

Scott and Freeman[81] compared cognitive therapy delivered by a psychologist or supportive counseling delivered by a social worker to usual care. Neither the cognitive treatment nor the counseling was provided according to a formal manual or otherwise clearly structured. Analysis was not intention to treat. Over the 16-week course, the cognitive intervention averaged nearly 7.75 hours and the social work counseling more than 12 hours, compared to less than 1 hour by the general practitioners. At 16 weeks there was no difference in percentage recovered between the cognitive therapy group and usual care (41% vs 48%), but the social work group (72%) produced substantially higher rates of recovery.

Teasdale et al[90] compared up to 20 one-hour sessions of cognitive therapy (mean 15.2 hours) delivered by doctoral-trained, experienced psychologists to usual care for patients with major depression in a primary care setting. The investigators ensured adequacy of treatment by tape review and did not employ a structured manual. Analysis was not done on the basis of intention to treat. Immediately post-treatment, patients in the therapy group averaged a greater change in depressive severity on the BDI than did the usual care group (22 point decrease vs 11.5 point decrease, P <0.01). This benefit was not apparent at follow up three months after completing treatment. Of note, contact time for the therapy group was substantially greater than for usual care.

Blackburn and colleagues[83] compared the outcomes for patients receiving either a pharmacologic intervention (the TCA amitriptyline) or only cognitive psychotherapy to outcomes for patients receiving combined cognitive psychotherapy and amitriptyline; all patients had a diagnosis of major depression. Psychologists performed 12 to 20 sessions of therapy; no manual was used and the degree to which the treatment was structured is unclear. An intention-to-treat analysis was not done, and allocation to

therapists was not randomized. Subjects in all 3 groups showed benefit, but patients in the cognitive therapy group and the combined treatment group tended to have a greater decrease in depressive severity than the amitriptyline-only group. Specifically, using a more than 50% decrease in depressive severity immediately post-treatment as the outcome of interest, 81.8% of the combined group, 72.7% of the cognitive therapy group, and 55% of the medication group achieved that outcome (overall chi-square test was not significant). The cognitive therapy and combined groups appeared to have substantially more visits than the medication-only group. Attrition during the trial was 27%.

Minor Depression. Two studies assessed the benefits of counseling for patients with minor depression. Miranda and Munoz[87] compared a CBT approach consisting of 8 weekly 2-hour sessions by doctoral-level psychologists following a specific protocol (according to a formal manual) to usual care in primary care medical patients. Over the subsequent year, the cognitive therapy group had a greater reduction in depressive severity and missed fewer medical appointments. The sample (n=150) consisted of patients with minor depression (33%, n=49) and patients with other subthreshold depressive symptoms. The attrition rate for the full sample was large; 20% of those randomized attended none of the 8 sessions, and 37% of the sample attended fewer than half of the sessions.

Lynch et al[86] compared telephone counseling (consisting of 6 weekly 20-minute phone sessions of PS therapy conducted by student therapists with minimal experience) to usual care for the treatment of minor depression. The therapy was relatively structured and was based on an existing PS therapy model. The sample size was small (n=29). The telephone counseling group had more drop-outs than usual care (4/15 vs 1/14) and an intention-to-treat analysis was not done. The counseling group had a significant 4.7-point drop (from a baseline of 15.6) in its HAM-D score immediately following the intervention, whereas the usual care group had no significant change in depressive severity (from 12.4 to 13.3).

We found no RCTs of psychotherapy for dysthymia in either primary care or psychiatric settings.

Educational and Quality Improvement Interventions

Five studies examined health care delivery strategies that did not directly involve traditional medication or psychotherapeutic interventions.[129,132,133] Katon and colleagues[129] tested a "Collaborative Care" model that included patient education, on-site consultation for patients, active collaboration with primary care physicians, and increased frequency and intensity of primary care visits. At 4 months, significantly more patients with major depression who received care through the collaborative care model had a greater than 50% decrease in depressive severity than did patients on usual care (74.4% vs 43.8%). The authors reported no significant difference for patients with minor depression (60% vs 67.9%).

Llewellyn-Jones and colleagues[132] tested a "Shared Care" model for "depressed" patients involving caregiver education, health education and promotion for patients, and improved communication between general practitioners and staff at a single elderly residential care facility. Their design examined control and intervention groups in a serial fashion. The intervention was "population-based" in that it was targeted to the entire living facility. Participation was variable: only 62% of either study group had general practitioners who attended the provider education program. The intervention itself was relatively inexpensive. Compared with patients in the control group, patients in the multi-faceted intervention group had a significantly greater reduction in their GDS scores (by 1.87 points) and were more likely to move to a "less depressed" state (45% vs 31%).

Peveler et al[133] tested the benefits of 2 sessions of counseling about antidepressant medication adherence, or the provision of an information leaflet about adherence, versus usual care in a population with "depressive illness." No difference in depressive symptoms as measured by the Hospital Anxiety

and Depression Scale[142] was found between treatment groups overall. However, among patients with major depression who received higher doses of medication, those in the counseled group had significantly lower final depression scores than those with usual care (4.0 vs 5.9).

Katzelnick et al[131] compared the benefits of a systematic, primary care-based depression treatment program for depressed "high utilizers" not in active treatment. This depression management program (DMP) consisted of patient education materials, physician education programs, telephone-based treatment coordination, and antidepressant medication treatment initiated and managed by the patients' primary care physician. Those receiving the DMP were compared to a usual care arm in an intention-to-treat analysis. The DMP group was significantly more likely to fill 3 or more antidepressant prescriptions in the first 6 months (69.3% vs 18.5%, vs, $P< 0.001$) and had significantly greater improvement in HAM-D depressive severity scores at 1 year (-9.2 vs -5.6, $P< 0.001$), with this benefit beginning by 6 weeks into the study. Additionally, at 1 year, intervention patients were more improved on mental health, social functioning, and general health self-report measures ($P< 0.05$ for each domain). Of note, mean visits counts in the DMP increased by 1.6 visits, whereas mean visits counts decreased in the usual care group by 2.0 visits ($P=0.02$).

Simon et al[130] compared a program of feedback only and 1 of feedback plus care management to usual care in primary care patients with recently diagnosed depressive illness. The feedback-only intervention consisted of feedback and algorithmic recommendations to doctors at 8 and 16 weeks based on data from computerized records of pharmacy and visits. The feedback plus care management group additionally provided to patients 2 later telephone monitoring contacts (at 8 and 16 weeks), which were followed by more sophisticated feedback to the doctor based on information received during the phone call.

In an intention-to-treat analysis compared to usual care, the care management group had a higher probability of receiving at least moderate doses of antidepressants (OR, 1.00; 95% CI, 1.23 to 3.22). The care management group also had a significantly higher probability of showing a 50% decrease in depression severity (OR, 2.22; 95% CI, 1.31 to 3.75) and a significantly lower probability of persistent major depression (OR, 0.45; 95% CI, 0.24 to 0.86) at 6 months. Meanwhile, relative to usual care, the feedback-only group showed no difference on receiving at least moderate doses of antidepressants (data not provided), the probability of a 50% decrease in severity (OR, 1.12; 95% CI, 0.73 to 1.73), or the probability of major depression at follow-up (OR, 0.89; 95% CI, 0.55 to 1.46).

Conclusions about Therapies for Adults

Effective treatments for depressive illness in primary care are available. Antidepressant medications for major depression are clearly effective compared with placebo. Most of these results have come from structured efficacy trials with selected populations, although more recent studies using usual-care comparison groups and real-world settings have produced similar effects.[76-78]

Antidepressant interventions for dysthymia are probably effective in primary care patients; although only 2 studies have been performed in primary care settings, evidence from multiple sites (inpatient psychiatric hospitals, outpatient psychiatric clinics, primary care practices, and the community) show a similar magnitude of effect. The evidence regarding the benefit of antidepressant medication for minor depression is limited. The 1 trial addressing this question (the Collaborative Care model,[129] in which improved medication prescription and adherence was part of the intervention) did not find a statistically significant benefit with antidepressants, but it may have been underpowered to detect a modest but clinically important effect (10% to 15%).

Tricyclic agents and newer agents (including SSRIs) have similar efficacy. The newer agents, however, have fewer side effects and are less likely to have side effects that lead to drop-out. Total drop-out rates, however, did not differ. Of note, the 1 effectiveness study (which most closely represented actual practice in primary care by allowing naturalistic follow-up and management by primary care physicians) [139] found that the drop-out rates for the tricyclic-treated patient were much higher than those for the SSRI-treated patients. For patients with major depression, greater side effects lead to significantly higher drop-out rates from treatment, although similar drop-out rates were not noted for patients with dysthymia.

Psychotherapeutic interventions appear as effective as antidepressant interventions for major depression, with a similar magnitude of effect. In general, the more effective psychotherapeutic interventions had greater structure to their treatments. Relative to pharmacologic interventions, psychotherapeutic interventions were clearly more time intensive. Four studies used between 4 and 6 sessions totaling 3 to 4 hours for their interventions.[74,85,89,135]

Evidence on the effectiveness of psychotherapeutic intervention for patients with minor depression is limited, although the results of 2 studies using well-structured interventions suggest potential benefit. [86,87] No evidence exists concerning the use of psychotherapy alone for dysthymia.

Few studies have examined the effect of combining medications and psychotherapy. Two studies involving combined treatments did not find a significant incremental benefit when compared to a single active intervention.[74,83] However, 2 recent trials in psychiatry clinic settings suggest that combination therapy may improve long-term outcomes.[140,141]

Treatment of Depression in Children and Adolescents

Treatment of depression has been less studied in adolescents and children than in adults. Nevertheless, recent trials and systematic reviews have increased the knowledge of the efficacy of different forms of treatment for depression. In this section we review the evidence for treatment of depression in adolescents and children with psychotherapy and pharmacotherapy. We reverse the order of discussion (relative to that for adults) because psychotherapy has been a comparatively more important intervention for children in the past. Before considering treatment options, we discuss options for preventing depression in this age group.

Preventing Depression

One method of reducing the impact of depression is to treat risk factors and symptoms before they lead to a full episode of major depression. Some studies, described below, provide limited evidence on this approach for children and adolescents (eg, intervening with children with subclinical depression or providing assistance with coping skills for children at risk of depression).

Jaycox et al[143] reported reduction in depressive symptoms in the Penn Prevention Program, a prospective cohort study of 142 children ages 10 to 13 years. They used CBT to teach coping strategies to 69 "at-risk" children in a treatment group. At-risk children were selected based on depressive symptoms and reports of parental conflicts. The treatment group was compared to 73 control children who did not receive any intervention. Children were not randomized to intervention. Outcomes were assessed after 6 months using the Children's Depression Inventory (CDI). In the treatment group, the percentage of children who were moderately depressed (CDI ≥15) decreased significantly from 24% to

15% ($P \leq 0.05$); in the control group the change in percentage of depressed subjects was not significant (24% to 23%, $P=0.36$). Based on self-reported depressive symptoms in the 6 months following the intervention, 23% of the children in the treatment group and 44% of the control group reported moderate depressive symptoms ($P \leq 0.05$).

Clarke et al[144] were able to demonstrate positive results in adolescents with depressive symptoms at risk for developing a DSM-IIIR-defined episode of depression. The intervention consisted of assessment of symptomatology by the CES-D with subsequent K-SADS diagnosis of depression or dysthymia. The investigators randomly assigned 172 adolescents with subclinical depression to a usual-care control group or an after-school cognitive psychotherapy group. Total incidence of major depression or dysthymia during follow up was 18 of 70 children (25.7%) in the control group and 8 of 55 (14.5%) for the intervention group.

Lamb et al[145] conducted a school-based program designed to promote coping among rural adolescents with depressive symptoms. The study surveyed 222 students ages 14 to 19 years and identified a subgroup of subjects with moderate to high Reynolds Adolescent Depression Scale (RADS) scores who could be randomly assigned to treatment or control groups. The treatment consisted of 8 weeks of group sessions using coping techniques and role-playing tasks. Four students dropped out of the treatment group; 1 left the control group. The investigators found that 87% of the intervention group and 61% of the control group improved on RADS scores. These results were significant for females ($P=0.032$) but not for males.

Psychotherapy

Psychotherapy has been the mainstay of treatment for children diagnosed with depression. Various forms of psychotherapy and counseling have been used. CBT is the method that has been studied most rigorously and been shown to be effective.

Reinecke et al[146] recently reviewed evidence on CBT in a systematic review and meta-analysis. The authors identified 6 controlled clinical trials with 14 post-treatment control comparisons and 10 follow-up control comparisons covering 217 subjects.[147-151] All studies were conducted in adolescents ages 10 to 19 years. All but 1 study recruited subjects in schools and used group therapy sessions. The interventions lasted 5 to 8 weeks and included 6 to 14 sessions with follow-up periods of 1 to 3 months. Outcomes were based on different depression scales.

The overall pooled effect size (a measure of change in standard deviations) at post-treatment was -1.02 (95% CI, -1.23 to -0.81) and for follow-up data -0.61 (95% CI, -0.88 to -0.35). Negative effect size scores indicated a decrease in combined depression measures and improvement of symptoms in terms of standard deviations. Thus, CBT appears to be effective in reducing depressive symptoms among adolescents. Treatment gains seem to be maintained after completion of therapy. The results of this meta-analysis were consistent with other meta-analyses of psychotherapy for depression in children and adolescents.[152,153]

Pharmacotherapy

Tricyclic Antidepressants. Two recent systematic reviews have examined the use of TCAs in children and adolescents. Hazell et al[166] published a meta-analysis of 12 RCTs comparing the efficacy of TCAs with placebo in depressed children ages 6 to 18 years.[154-166] All studies but 1 suggested greater improvement in the TCA group than in the placebo group, but the difference was statistically significant

in only 1 study. Six studies presented results as a change in scales of depressive symptoms using the CDI, Children's Depression Rating Schedule-Revised (CDRS-R), K-SADS, or Depressive Adjective Checklist (DACL). Effect size in the 6 studies ranged from -0.29 to 1.57 with a pooled effect size of 0.35 standard deviations (95% CI, -0.16 to 0.86). The authors concluded that the trend toward improvement in depression on TCAs versus placebos was not statistically significant and likely not clinically significant. They did note the important placebo effect (in some trials more than 50% of subjects improved).

Geller et al[172] conducted a systematic review of TCA use in children and adolescents for various indications including depression.[160,162,165,167-172] They reviewed double-blind, placebo-controlled trials and reported no significant improvement with treatment of depression using TCAs compared with placebos in 6 studies. One of the studies in the review produced mixed results based on different outcome rating scales.[168] Another study demonstrated improved outcomes on intravenous clomipramine versus placebo;[169] however, a study focusing on intravenous medication is not applicable to ambulatory care treatment. We found no additional RCTs using TCAs for treatment of depression in children or adolescents in our updated literature search. Thus, it appears that TCAs are ineffective for treating depression in children and adolescents.

In addition to considering efficacy, the important side effects of TCAs, including sudden death and fatal overdose potential, must be considered in any discussion of management of patients in this age group.

Selective Serotonin Reuptake Inhibitors. SSRIs are a relatively new therapy for the treatment of children and adolescents with depression. Favorable anecdotal clinical experiences and open trials have reported improvement in depression for pediatric patients on fluoxetine,[173-177] sertraline,[178-180]and paroxetine.[181-183] Recent clinical trials (discussed below) have added to the evidence.[184,185] To date,

however, no studies in children or adolescents have been conducted in the primary care setting. Efficacy studies, clinical experience, and case reports suggest that overdose potential and side effects are lower in pediatric subjects than in adults; however, more subtle effects on neurobiology and behavior are unknown at this time.

Simeon et al[184] published a placebo-controlled, double-blind study of fluoxetine. The study included 40 inpatients and outpatients ages 13 to 18 years with unipolar depression defined by HAM-D scores of ≥20. The intervention consisted of a 1-week placebo period for all subjects followed by 8 weeks of either fluoxetine titrated to 20-60 mg daily dose or placebo. Thirty-two patients were followed for a mean of 24 months with the HAM-D and other behavioral symptom scales and clinical measures. It is not clear if the 8 drop-outs were included in the final results. Results were not reported in sufficient detail to calculate effect size. In general, most adolescents on fluoxetine or on placebo improved. Fluoxetine treatment was superior to placebo in many clinical measures, but the differences were not statistically significant.

Emslie et al[185] conducted the first double-blind, randomized, placebo-controlled clinical trial of fluoxetine in children and adolescents. The study included 96 children ages 7 to 17 years with nonpsychotic major depression diagnosed by DSM-IIIR criteria from a structured clinical interview, depression scales, and consensus team diagnosis. All subjects participated in a 1-week placebo run-in period. Patients were randomized to placebo or 20 mg of fluoxetine every morning for 8 weeks. Thirty-six patients did not complete the full 8-week trial following randomization: 5 because of side effects (4 in the treatment group, 1 in the placebo group); 5 because of protocol violation (3 in the treatment group, 2 in the placebo group), and 26 because of a lack of efficacy (7 in the treatment group, 19 in the placebo group). Of the 60 patients who completed the 8-week trial, 25 of 34 (74%) responded to treatment and

15 of 26 (58%) responded to placebo. Differences in raw scores of the Clinical Global Impressions (CGI) and the CDRS-R were also significant among patients who completed 5 or more weeks of the trial. Although many of the subjects improved, only 31% of the original 48 treatment patients and 23% of the 48 placebo patients had a remission of depression to minimal symptomatology (CDRS-R ≤28). The NNT based on this result is 13 depressed children treated with fluoxetine to achieve clinical remission in 1 patient.[185]

Several studies that are under way or planned to evaluate SSRIs in depressed children and adolescents should add to the growing body of evidence on treatment. In addition, the Texas Medication Algorithm Project (TMAP)[186,187] and other groups such as the American Academy of Child and Adolescent Psychiatrists (AACAP)[188] have proposed treatment guidelines that feature SSRIs as first-line therapy for pediatric patients with depression. At present, most of the recommendations have focused on psychiatric care and do not describe the role of primary care providers in pharmacotherapy.

Combination Therapy. Clinical experience and expert opinion suggest that combination therapy may improve long-term outcomes especially for complex patients with comorbid disorders. No randomized trials in children or adolescents are available to describe the efficacy of combination therapy with multiple medications or pharmacotherapy and psychotherapy versus monotherapy with medication or counseling alone.

Additional Considerations

This review has attempted to describe generally the identification and management of children who present in primary care, but special patient populations should be considered. Gender, age, and ethnicity are important variables in existing studies that may limit generalizability of results. Many of

the above studies did not have large numbers of minorities or patients of lower socioeconomic status. Most of the positive studies and results are based on adolescents and older children. Aside from case reports and series, very few data are available about interventions in young children. Individual characteristics should be considered before the results on any larger population are generalized or applied to a specific patient.

Finally, children with poor health and chronic illnesses have been reported to have higher rates of depression and mental health problems. It is very important to consider depression and comorbid effects on chronic medical conditions in terms of adherence to medical treatment, functionality, and outcomes. However, in pediatric patients with chronic illness, screening tools for depression appear to lack sensitivity and predictive value and thus cannot be recommended for routine use.[123] In addition, studies are not yet sufficient to document treatment effectiveness in these patients.

Conclusions for Children and Adolescents

Data on prevention of depressive disorders in school and community settings provide support for intervention on selected youths with depressive symptoms, although no studies have described this type of intervention in primary care settings. The approach most relevant to primary care involves early recognition of depressed patients, proper identification and diagnosis, and facilitation of effective treatment.

Treatment of depression in adolescents with CBT or SSRIs appears to be effective. Whether these results can be generalized to primary care settings or to children is unclear. TCAs are not effective for treatment of depression in children and adolescents. The comparative efficacy of psychotherapy alone, medications alone, or combined treatments in children or adolescents is unknown.

Key Question 3: Screening Outcomes

The effect of giving health care providers the results of a screening test for depression has been compared with usual care in 14 randomized or quasi-experimental trials in primary care settings. Detailed study characteristics and results for these 14 trials can be found in Evidence Table 4 in Appendix D. In this section of the SER, we describe and compare the main findings from these trials and attempt to understand the effect of screening (compared with usual care) on the diagnosis, treatment, and outcomes of depression in primary care settings.

Overview of Screening Outcome Studies

Several different screening instruments have been tested as a means of providing feedback to providers. Four studies used the Zung SDS;[95,97,98,103] 3 papers from 2 studies used the CES-D;[91,92,102] 3 studies used the GHQ (which contains items about depression as well as other psychiatric conditions);[94,96,99] 1 study each used the BDI,[93] the SDDS-PC,[100] the GDS[82] and a 2 item screener.[101] The results of these studies are summarized in Table 15a-15d.

Eight papers from 7 studies[82,91,92,93,95,97,98,102] examined the effect of feedback of screening results on the rate of diagnosis and recognition of depressive disorders; another group of 8 studies[82,91,92,93,95,97,101,102] examined the effect on prescription of treatment for depressed mood. A different set of 9 trials directly examined the effect of screening on patient health outcomes, including changes in depression severity, duration, number of depressive symptoms, or health care

utilization.[91,92,94,96,100-103] In the next sections, we examine in depth the effect of screening feedback on diagnosis, treatment, and outcomes of depression.

Results of Screening Outcomes Studies

Effect of Screening on Recognition and Diagnosis of Depression

Seven studies examined the effect of screening, compared to usual care, for the diagnosis of depression (Table 15a and 15b). Moore et al,[98] Linn and Yager,[95] and Magruder-Habib et al[97] all used the Zung SDS screener. Callahan et al[91,92] and Williams et al[102] used the CES-D. Dowrick used the BDI.[93] Whooley et al[82] used the GDS. In each study, the detection of depression was assessed by chart audit.

Moore et al[98] screened consecutive patients, 20 to 60 years of age, at a university-based family medicine residency program. All patients were asked to self-administer the Zung SDS. The intervention patients' providers received feedback about SDS results greater than 50; providers of patients who scored below 50 and of all control patients simply received notice that their patients had been screened. No attempt was made to confirm the diagnosis using a criterion standard. Of 212 subjects in the trial, 96 scored above 50 (45%). Recognition of depression, defined by any notation in the chart, was 56% for cases in the intervention group (28/50) and 22% in the control group (10/46). The difference between intervention and control groups was similar for "severe depression," but rates of detection in both groups were higher (73% vs 37%). Effect on treatment rate and outcomes was not described.

Linn and Yager,[95] in testing the self-administered Zung SDS, randomized 150 consecutive new patients from a primary care clinic to feedback or no feedback. They found that patients assigned to the

feedback group were more likely to have depression diagnosed (29% vs 8%) than the no-feedback patients, but they did not employ any criterion standard.

Magruder-Habib et al[97] screened 800 Veterans Administration primary care clinic patients for depression. Research assistants administered the Zung SDS and used the DIS to confirm diagnosis with DSM-III criteria. Patients with SDS scores greater than 75 were excluded from randomization. The 100 patients who screened positive and met DSM-III criteria for major depression were then randomized to feedback or usual care. Those patients whose physicians received feedback were 3 times as likely to be accurately identified as depressed at the outset than were those whose clinicians had not received such feedback (25% vs 8%). At 1-year follow-up, 42% of the intervention patients, but only 21% of the controls, had been recognized as depressed.

Callahan et al[91,92] conducted an RCT of feedback from screening plus targeted educational information and treatment recommendations for patients over age 60 years in an academic primary care setting that served a low-income population. Potential subjects were screened by research assistants using the CES-D and HAM-D depression scales. Those patients scoring above the threshold for diagnosis were eligible to be randomized. Randomization was by physician, with certain clinic sessions randomly assigned to the intervention and others to control. All physicians received an educational talk at baseline.

Two articles appear to report results from this study. The first article, based on a 175-patient sample, found that patients in the intervention group were more likely than the control group to have a new notation of depression in their charts (32% vs 12%).[91] In the second paper, additional analyses on a larger sample size (n=222) found higher rates for documentation of depression (87% vs 40%).[92]

Williams et al[102] used the CES-D or a single question about depressed mood to examine the effect of feedback to providers for adult primary care patients. Most patients were able to complete the

single question (90%) and the CES-D (54%) without assistance. The presence or absence of depression was later confirmed using the DIS and DSM-IIIR criteria.[102] Current depression was defined as either meeting the DSM-IIIR criteria for major depression or dysthymia or having minor depression (depressed mood or anhedonia plus 1 to 3 additional DSM-IIIR symptoms). Based on chart reviews, current depression was recognized in 39% of patients whose providers received feedback from screening and in 29% of controls. This difference of 10% in the rate of recognition did not reach statistical significance.

Dowrick[93] randomized 116 patients who were initially rated "not depressed" by their usual general practitioners but had BDI scores greater than 14. Feedback was provided 1 week after the visit in which screening took place and was noted in the chart for subsequent visits. The study was powered to detect a 30% difference in the level of diagnosis after feedback. There was a higher level of depression diagnosis at 1 year in the feedback group (35% vs 21%; OR for detection, 2.10; 95% CI, 0.84 to 5.28), but the difference did not reach statistical significance.

Whooley et al[82] randomized primary care clinics to screening with physician feedback versus no screening or feedback for patients over age 65 years screened with the GDS. No criterion standard was applied. They found no difference in the rate of diagnosis of depression at 2 years.

In conclusion, feedback of screening results to providers increases the recognition of depression, especially major depression, by a factor of 2 to 3 in all cases except for the trial by Whooley et al.[82] The absolute increases in the diagnosis of depression range from 10% to 47%, with larger differences for major depression. Recognition and diagnosis of minor depression, when assessed, were generally low in both intervention and control groups.

Effect of Screening on Treatment of Depressed Patients

The effect of feedback of screening results on the proportion of depressed patients who receive treatment was examined in 7 studies (Tables 15a and 15c). Treatment generally included prescription of pharmacologic antidepressant therapy or referral to mental health services. Most studies evaluated treatment by chart audit; some used pharmacy databases. Actual patient adherence was not directly measured.

In contrast to recognition and diagnosis, the effect on rates of treatment was mixed. In 3 studies (Linn and Yager;[93] Dowrick;[95] Williams et al[102]), the documented rates of treatment were nearly equal in the intervention and control groups (Table 15c). Other studies, however, found improvements in the rate of treatment, with increases in the prescription of antidepressant medication more common than changes in mental health referrals. Callahan et al, [91,92] using a stepped program of treatment recommendations in addition to the feedback, found a difference of 17% to 18% in the initiation of a treatment plan and an increase in 12% for the rate of antidepressant prescription (P=0.01). Magruder-Habib et al[97] found an initial difference of 24% in the rate of treatment, although at 1 year it declined to a difference of 14% (56% vs 42%). The Williams et al[102] study also did not find an overall difference in treatment.

Wells et al[101] studied the effect of combining screening and a quality improvement program for depression treatment in 46 primary care clinics and measured its impact on treatment and outcomes of depression. Patients were enrolled if they screened positive on a 2-question screener. Patients received the Composite International Diagnostic Interview (CIDI) criterion standard examination, but participation was not based on its results. Randomization was at the level of the practice, and the intervention included feedback on the results of the 2-item screener. Intervention practices also received educational materials and assistance with quality improvement in treatment initiation and maintenance plus access to nurse-led medication follow-up or to cognitive-behavioral therapy. The investigators

screened 27,000 patients, identified 3,918 as potentially eligible, and randomized 1,356 patients. Subjects were followed for 12 months. The proportion of patients receiving appropriate treatment was increased in the intervention group at 6 months (50.9% vs 39.7%) and at 12 months (59% vs 50%, P=0.006).

Effect of Screening on Depression Outcomes

The effect of screening and feedback on depression outcomes was measured in 8 studies (Tables 15a and 15d).

Johnstone and Goldberg[94] applied the GHQ to 1,093 primary care patients and identified 119 cases of depression. These 119 subjects were randomly assigned to feedback of the results to the physician or to usual care. The investigators found no difference in mean GHQ scores at 12-month follow-up, but they did see a larger improvement with feedback among the subset of subjects with severe depression. For all patients, the mean duration of the first episode of depression and the total amount of time depressed were decreased by approximately 2 months (P <0.01).

Zung and King[103] screened 499 patients at a single private physician's practice. Of the 60 who screened positive, 49 were confirmed to have major depression using DSM-III criteria and were randomized to feedback and treatment with the benzodiazepine alprazolam (n=23) or to usual care (n=26). Four weeks later, outcome data were available for 20 patients in each group. The feedback and treatment group was more likely than controls to improve by at least 12 points on retesting with the Zung scale (66% vs 35%, P <0.05).

In Callahan et al[91,92] no improvements in HAM-D score emerged among those who received feedback of screening results.

Reifler et al[100] used the SDDS-PC, followed by a depression-specific diagnostic module, in 358 primary care patients. The 186 intervention patients had a lower mean number of visits than the 172 controls (3.7 vs 5.3, P=0.06), but other outcomes including SF-36 or SDS scores did not differ.

Lewis et al[96] used the GHQ and a computer-based diagnostic tool (PROQSY) to examine the effect of feedback of positive scores on outcomes in low-income primary care patients in London. Compared with GHQ scores for controls at 6 weeks, GHQ scores were lower for patients whose providers received feedback on the PROQSY results but not for those who received only GHQ results. The differences were attenuated and nonsignificant at 6-month follow-up.

Williams et al[102] found a statistically nonsignificant difference of 9% in the proportion of subjects still depressed at 3 months. The rate of recovery (patients with 1 or no DSM-IIIR criteria), however, was higher in the intervention than control groups (48% and 27%, respectively; P <0.05).

Whooley et al[82] found little difference in the proportion of patients depressed on the GDS after 24 months of follow-up: 42% for intervention patients and 50% for controls (P=0.3).

Wells et al[101] found statistically significant increases in the proportion of intervention patients (intervention practices received feedback of screening results and a quality improvement intervention) who were not depressed at 6 and 12 months and in the rate of job retention. Based on CES-D scores, intervention subjects were less likely to be depressed at 6 months than controls (55% vs 64%, P=0.001) and at 1 year (55% vs 61%, P=0.04). Among patients initially employed, 90% were still working, as compared with 85% of controls.[101]

Based on the results of Wells et al[101], approximately 10 to 12 patients identified as being depressed by screening would need to be treated to produce 1 additional remission. Twenty patients would need to be treated to preserve 1 patient's job. If depression is present in 5% to 10% of primary

care patients, 100 to 200 patients would need to be screened to produce 1 additional remission at 6 months.

Conclusions about Screening and Feedback

In summary, multiple studies have examined the effect of providing feedback of depression screening results to providers in primary care. The rate of detection and diagnosis of depression, based mainly on chart reviews or the completion of a study-specific form, increased by 10% to 47% in the 6 studies reporting this outcome. The effect on treatment was more variable. Four of the 8 studies reporting this outcome found small, nonsignificant increases in the proportion of patients treated for depression.[93,95,102] Magruder-Habib et al[97] found a much larger increase (24%), and Callahan et al[91] noted increases in antidepressant prescribing but not referral for counseling or psychiatric care. Wells et al also noted a 10% increase in appropriate treatment, which was statistically significant.

The effects of depression screening on clinical outcome of depression were also mixed. Two small, older trials found large improvements in major depression.[94,103] Two larger, well-designed trials found moderate improvements (9%) in remission from depression in a population with a mixed set of diagnoses.[101,102] Four other studies found small or no improvements in outcomes.[82,91,92,100]

Thus, although the effect of screening on diagnosis appears robust, improvements in more distal variables such as treatment and outcomes are not as consistent or as large. Translating the increased rates of detection with screening into improved outcomes may require that particular attention be paid to initiation and maintenance of effective therapy, perhaps in the form of a quality improvement effort or other programs systematically designed to provide appropriate care.

Demonstrating improvements in clinical outcomes (as measured by the proportion still depressed, for example) requires large samples. Studies with smaller sample sizes may be unable to demonstrate statistically significant results despite finding clinically significant differences in recovery.

Major depression appears more responsive to intervention with screening and feedback than minor depression, although the Wells et al[101] study suggests that outcomes can be improved for all subjects with sufficient attention to treatment. The appropriate outcome measure for minor depression differs from major depression, so failure to demonstrate changes in the proportion of patients depressed may not be a fair test for patients with subsyndromal illnesses.

Screening Outcomes for Children and Adolescents

No studies have examined the overarching question of treatment outcomes for children or adolescents identified by primary care providers using targeted screening or clinical suspicion. A large part of the literature focuses on development of screening measures and reliability testing; it does not provide information to assess screening accuracy or sample a general ambulatory population that generalizes to primary care settings. No randomized trials in children or adolescents evaluate the effects of screening for depression on outcomes of recognition, diagnosis, or treatment. No studies in pediatric patients have linked an initial screening assessment for depression with subsequent treatment and demonstrated improved patient outcomes as a direct result of screening. Some studies have shown that screening instruments, especially the relatively brief general measures such as the CBCL and PSC, may increase recognition of mental disorders and referrals; however, there is no evidence that these general screens of psychopathology can improve outcome of depressed children or adolescents.

Brief screens for depression, such as versions of the BDI and CES-D, have been used in children and adolescents. However, their predictive value in general populations with relatively low prevalence of depression may limit their effectiveness and usefulness as a screen for all pediatric primary care

patients. Targeted screening or use of measurement instruments on patients with suspected psychiatric disorder can improve diagnostic accuracy, but whether selective screening produces improved outcomes compared to usual care remains untested.

In addition to specific measures of depression, 2 general instruments that seek to identify psychosocial issues have been extensively researched and implemented in primary care.

Table 6. Diagnostic Instruments for Depression

Diagnostic Instrument Criterion Standards	Description	Application	Time Required to Diagnose Depressive Illness	Training Required	Feasibility in Primary Care Setting for Diagnosing Depression
DSM-IV diagnosis by a mental health professional[13]	List of specified diagnostic criteria as guideline for identifying specific psychiatric disorders	Clinical interview; used in both clinical and research settings	Few minutes	Minimal, can be learned with clinical experience; can be applied by primary care physician	Medium-High for common diagnoses such as major depression, dysthymia, minor depression
Structured Clinical Interview for DSM-IV (SCID)[189]	A semi-structured research diagnostic interview designed for making DSM diagnoses	Primarily research instrument administered by clinically trained interviewers; designed for a patient population	5-15 minutes	Moderate-High; depressive sections can be administered by trained primary care clinicians	Low
Diagnostic Interview Schedule (DIS)[136]	A fully structured research interview created to provide current and lifetime DSM diagnoses	Primarily research instrument self-administered or administered by "lay" interviewers; designed for epidemiologic research in a community	5-15 minutes	Moderate-High; not designed for primary care setting	Low
Composite International Diagnostic Interview (CIDI)[190]	A fully structured research interview created to provide current and lifetime DSM diagnoses; derived from DIS, with improved diagnostic accuracy and wider cross-cultural applicability	Primarily research instrument self-administered or administered by "lay" interviewers; designed for epidemiologic research in a community	5-15 minutes	Moderate-High; not designed for primary care setting	Low
Research Diagnostic Criteria (RDC)[138]	A set of diagnostic criteria similar to the DSM criteria	Research criteria for clarifying diagnoses	Few minutes	Minimal; can be learned with clinical experience	Medium-High (similar to DSM criteria)

DSM indicates Diagnostic and Statistical Manual of Mental Disorders (-III, third edition; -IIIR, third edition revised; -IV, fourth edition)

Table 7. Other Psychiatric Illnesses Producing Depressive Symptoms

Psychiatric Illness	Typical Symptoms
Bipolar disorder	Past or current presence of one or more manic episodes, usually accompanied by major depressive episodes
Panic disorder	Recurrent unexpected panic attacks about which there is persistent concern
Substance-related disorders	Maladaptive pattern of substance use manifested by recurrent and significant adverse consequences related to repeated use of substances (abuse), or a cluster of cognitive, behavioral, and physiological symptoms indicating continued substance use despite significant substance-related problems (dependence)
Substance-induced mood disorder	Depressive episode in which a substance (such as a drug of abuse, a medication, or a toxin) is judged to be etiologically related to the mood disturbance
Adjustment disorder	Depressive symptoms in response to a psychosocial stressor not meeting criteria for major depression
Bereavement	Depressive symptoms in reaction to loss of a loved one which (1) are present for two months or less and (2) do not cause marked functional impairment, morbid preoccupation with worthlessness, suicidal ideation, psychotic symptoms, or psychomotor retardation

Source: Adapted from DSM-IV.[13]

Table 8. Characteristics of Case-Finding Instruments for Adults Used to Detect Depression in Primary Care Settings

Instrument	Items, n†	Time Frame of Questions	Score Range	Usual Cut-point ‡	Literacy Level §	Administ-ration Time, min
BDI	21	Today	0-63	10 mild 20 moderate 30 severe	Easy	2-5
CES-D	20	Past week	0-60	16	Easy	2-5
GHQ	28	Past few weeks	0-28	4	Easy	5-10
MOS-D	8	Past week	0-1	0.06	Average	<2
PRIME-MD	2	Past month	0-2	1	Average	<2
SDDS-PC	5	Past month	0-4	2	Easy	<2
Zung SDS	20	Recently	25-100	50 mild 60 moderate 70 severe	Easy	2-5

Source: Adapted from Mulrow et al, 1995.[104]

BDI indicates Beck Depression Inventory; CES-D, Center for Epidemiologic Study Depression Screen; GHQ, General Health Questionnaire; MOS-D, Medical Outcomes Study Depression Screen; PRIME-MD, Primary Care Evaluation of Mental Disorders; SDDS-PC, Symptom Driven Diagnostic System – Primary Care; Zung SDS, Zung Self-Assessment Depression Scale.

† Item numbers for the PRIME-MD and SDDS-PC refer to depression questions only. Several instruments now have shortened versions as well.
‡ Cut-point is the number at or above which the test is considered positive.
§ Easy equals third- to fifth-grade reading level; average equals sixth- to ninth- grade reading level according to Fog Formula.

Table 9. Screening Accuracy in Geriatric Populations

Author	Test / Cut-point	Sensitivity	Specificity
D'Ath et al, 1994[41]	GDS / ≥5	91	72
Gerety et al, 1994[45]	GDS / ≥11	89	68
Neal and Baldwin, 1994[59]	GDS / ≥11	83	80
Van Marjwick et al, 1995[68]	GDS / ≥3	67	73
Arthur et al, 1999[34]	GDS / ≥3	100	72
Hoyl et al, 1999[49]	GDS / ≥5	94	82
Beekman et al, 1997[37]	CES-D / ≥20	93	73
Lewisohn et al, 1997[54]	CES-D / ≥12	76	77
Lyness et al, 1997[56]	CES-D / ≥21	92	87
Bird et al, 1987[38]	Self-Care D / ≥6	77	98
Upadhyaya and Stanley, 1997[67]	Self-Care-D / ≥5	95	86
Banerjee et al, 1998[35]	Self-Care-D / ≥8	90	53

CES-D indicates Center for Epidemiologic Study Depression scale; GDS, Geriatric Depression Scale

Table 10. Probability of Major Depression after a Positive Screening Test

Sensitivity / Specificity Estimates*	Pretest Probability 5%	10%	15%
90% / 85%	24%	40%	50%
90% / 80%	19%	33%	44%
84% / 72%*	13%	24%	34%
80% / 70%	12%	23%	32%

*Estimate from meta-analysis by Mulrow et al, 1995.[104]

Table 11. Stratum-Specific Likelihood Ratios (SSLRs) for CES-D

CES-D Score	SSLR with 95% CI	Post-test Odds If Pre-test Probability is 10%
0 - 29	0.35 (0.25 to 0.49)	3.7%
30 - 49	2.3 (1.8 to 3.1)	20.3%
50+	11.7 (3.1 to 44.0)	56.5%

CES-D indicates Center for Epidemiologic Study Depression scale.

Table 12. Diagnostic Value of Risk Factors for Major Depression

Risk factor	Prevalence (%)	OR / RR (95%CI)	LR + / LR -
Family history of depression	2.2	2.4 (1.4 to 4.1)	1.7 / 0.7
Female gender	71	1.8 (1.0 to 3.4)	1.2 / 0.6
Unmarried	41	1.6 (1.0 to 2.5)	1.3 / 0.8
Unemployed	4.6	2.1 (0.9 to 5.1)	2.0 / 0.9
Alcohol abuse	15	2.3 (1.3 to 3.9)	1.9 / 0.8

Source: Adapted from Conde et al 1998.[107]

Table 13. Studies of Screening for Depression in Pediatric Populations

Source	Instrument	Population	Diagnostic Gold Standard Criteria	Sensitivity	Specificity
Friedman and Butler, 1979[191]	CDI ≥19	N=40 Ages 8-13 years Psychiatric sample and normals	Ability to discriminate referred and nonreferred patients in the sample	88	90
Barrera and Garrison-Jones, 1988[112]	BDI Cut-off 16	N=49 Ages 12-18 years Community sample	Children's assessment Schedule DSM-III Major Depression	100	93
Fendrich et al, 1990[117]	CES-D-C Cut-off 15	N=166 Ages 12-18 years Referred psychiatric sample	DSM-III diagnosis of depression or dysthymia	71	57
Kashani et al, 1990[109]	BDI ≥16	N=100 Adolescents referred to counseling	DICA by nonpsychiatrists	48	87
Whitaker et al, 1990[113]	BDI ≥16	N=135 Ages 13-18 years Community sample	Depression: DSM-III diagnosis in lifetime Clinician in field interview	77	65
Whitaker et al, 1990[113]	BDI ≥16	N=135 Ages 13-18 years Community sample	Dysthymia: DSM-III Diagnosis in lifetime Clinician in field interview	71	64
Ambrosini et al, 1991[110]	BDI >13	N=122 Outpatients Psychiatric referral population	Current depression, K-SADS IIIR interview	86	82
Ambrosini et al, 1991[110]	BDI >13	N=53 Outpatients Psychiatric referral population	Depression, K-SADS IIIR Interview	89	88
Garrison et al, 1991[116]	CES-D Males (cut 12) Females (cut 22)	N=332 Ages 11-17 years Community sample	Depression: DSM-III criteria and CAS< 61	M† 85 F† 83	M 49 F 77

Table 13. Studies of Screening for Depression in Pediatric Populations (continued)

Source	Instrument	Population	Diagnostic Gold Standard Criteria	Sensitivity	Specificity
Garrison et al, 1991[116]	CES-D Males (cut 16) Females (cut 20)	N=332 Age 11-17 years Community sample	Dysthymia: DSM-III criteria & CAS<61	M† 75 F† 100	M 67 F 67
Roberts et al, 1991[111]	BDI ≥11 female ≥15 male	N=1,704 Ages 15-18 years Community sample	Current depression, DSM-III	84	81
Roberts et al, 1991[111]	CES-D ≥24 female ≥22 male	N=1,704, Ages 15-18 years Community sample	DSM-III Depression, Current	84	75
Angold et al, 1995[118]	SMFQ (cut 12)	N=173 Ages 6-17 years Mixed primary care and psychiatric sample	DISC	70	85
Winter et al, 1999[114]	BDI-PC ≥4	N=100 Ages 12-17 years Pediatrics office sample	Primary Care Evaluation of Mental Disorders Mood module [PRIME-MD, MM]	91	91

BDI indicates Beck Depression Inventory; BDI-PC, BDI for Primary Care; CDI, Children's Depression Inventory; CES- D, Center for Epidemiologic Study Depression scale; CES-D-C, CES-D for children; CAS, Child Assessment Schedule; DICA, Diagnostic Interview for Children and Adolescents; DSM, Diagnostic and Statistical Manual of Mental Disorders (-III, third edition; -IIIR, third edition revised; -IV, fourth edition); K-SADS, Schedule for Affective Disorders and Schizophrenia for School-age Children; SMFQ, Short Mood and Feeling Questionnaire.

Cut indicates cut-point.

†M = Male; F = Female.

Table 14. Studies of Psychotherapy in Patients with Major Depressive Disorders

Author	Structure	Contact Hours	Training	Outcome Measure	Response Rate: Intervention/ Control
Mynors-Wallis et al, 1995[77]	High	3.5	Medium-High	Recovery at 12 weeks (HAM-D ≤7 or BDI ≤8)	60%/27%‡
Holden et al, 1989[84]	Low	>4	Low	Recovery at 13 weeks (by Research Diagnostic Criteria*)	69%/38%‡ (MajD and MinD combined)
Scott, et al, 1997[89]	Medium	3	High	Recovery at 7 weeks (by NIMH criteria)	63%/33%
Katon et al, 1996[85]	High	2.5-3.5 +4 PC	High	≥50% improvement in SCL-20 depression score	MajD: 70.4%/42.3% MinD: 66.7%/52.8%
Mynors-Wallis et al, 2000[75]	High	2.8-3.1	Medium-High	Recovery at 52 weeks (HAM-D ≤7)	PS-GP: 62%/ § PS-N: 56%/ § Antidep: 56% / § PS+Antidep: 66%/ §
Schulberg et al, 1996[78]	High	16	High	Recovery at 8 months (by NIMH criteria)	46%/18%
Ross et al, 1985[88]	Low	9-18	High	% reduction in depressive severity (by HAM-D)	32%/17%
Appleby et al, 1997[74]	Low	3.5	Minimal	% reduction in depressive severity (by HAM-D)	64.0%/57.7%
Scott and Freeman, 1992[81]	Low	7.7-12.1	High	Recovery at 16 weeks (HAM-D<7)	Cog: 41%/48%‡ SW: 72%/48%‡
Teasdale et al, 1984[90]	Low	15.2	High	Decrease from baseline depressive severity (by BDI)	Post-tx: 22 pts/11.5 pts‡ 3 months post-tx: 19 pts/19 pts‡
Blackburn et al, 1981[83]	Low	12-20	High	≥50% decrease in BDI or HAM-D	Cog: 72.7%‡ Cog+Antidep: 81.8%‡ Antidep: 55.0%‡

Antidep indicates antidepressant; BDI, Beck Depression Inventory; Cog, cognitive therapy; HAM-D, Hamilton Depression Rating Scale; SCL-20, Symptom Checklist 20; MajD, major depression; MinD, minor depression; NIMH, National Institute of Mental Health; PC, private clinic; PS-GP, problem solving by general practitioner; PS-N, problem solving by nurse; SW, social worker providing supportive counseling; tx = treatment.

‡Not intention-to-treat analysis.
§No placebo or usual care comparison group.

Table 15a. Studies Examining the Effect of Screening and Feedback

Author / Year	Screening Instrument	Total Number of Subjects	Mode of Administration	Confirmatory Diagnostic Interview?	Feedback Provided
Johnstone and Goldberg, 1976[94]	GHQ	119	Self	Yes*	Immediate feedback
Moore et al, 1978[98]	SDS	212	Self	No	Immediate written feedback
Linn and Yager, 1980[95]	SDS	150	Self	No	Immediate written feedback
Zung and King, 1983[103]	SDS and immediate dx interview	49	Psychiatrist	Yes*	Immediate feedback
Magruder-Habib et al, 1990[97]	SDS	100	Research assistant	Yes*	Immediate written feedback
Callahan et al, 1994[91] 1996[92]	CES-D	175 222	Research assistant	Yes (HAM-D)*	Feedback to schedule 3 additional visits within 3 months
Dowrick, 1995[93]	BDI	116	Self	No	Written feedback to provider 1 week after visit plus chart note

* Required prior to randomization.
† Not related to randomization.

Table 15a. Studies Examining the Effect of Screening and Feedback (continued)

Author / Year	Screening Instrument	Total Number of Subjects	Mode of Administration	Confirmatory Diagnostic Interview?	Feedback Provided
Lewis et al, 1996[96]	GHQ	681	Self	PROQSY group only	GHQ group results provided immediately to provider; PROQSY group subjects were asked by provider to complete PROQSY within 1 week and schedule a follow-up visit
Reifler et al, 1996[100]	SDDS	358	Self	Yes†	Providers received diagnostic module worksheet at same visit for those screening positive
Williams et al, 1999[102]	CES-D, Blinded DSM-IIIR	969	Self	Yes†	Written results provided immediately to provider
Wells et al, 2000[101]	Two-item screener	1,356	Research assistant	Yes (subset)†	Providers notified and asked to schedule visit within 2 weeks
Whooley et al, 2000[82]	GDS	2,346	Research assistant	No	Intervention subjects notified same day: before visit, 74%; after visit, 26%

BDI indicates Beck Depression Inventory; CES-D, Center for Epidemiologic Study Depression scale; DSM, Diagnostic and Statistical Manual of Mental Disorders (-III, third edition, -IIIR third edition revised, -IV, fourth edition); dx, diagnosis; GDS, Geriatric Depression Scale; GHQ, General Health Questionnaire; HAM-D, Hamilton Depression Rating Scale; PROQSY, self-administered computerized assessment; SDS, Zung Self-Depression Scale; SDDS, Symptom-Driven Diagnostic System for Primary Care.

* Required prior to randomization.
† Not related to randomization.

Table 15b. Summary of the Effect of Screening and Feedback on Rates of Diagnosis

Author / Year	Percentage with Diagnosis		
	Intervention Group	Control Group	*P*-Value
Johnstone and Goldberg, 1976[94]		NR*	
Moore et al, 1978[98]	56%	22%	*P*<0.05
Linn and Yager, 1980[95]	29%	8%	NR
Zung and King, 1983[103]		NR	
Magruder-Habib et al, 1990[97]	25%	7.7%	*P* ≤0.05
Callahan et al, 1994[91]	32%	12%	*P*=0.002
1996[92]	87%	40%	*P*=0.001
Dowrick, 1995[93]	35%	21%	NR
Lewis et al, 1996[96]		NR	
Reifler, et al, 1996[100]		NR	
Williams et al, 1999[102]	39%	29%	*P* >0.05
Wells et al, 2000[101]		NR	
Whooley et al, 2000[82]	35%	34%	*P*=0.96

*NR indicates not reported.

Table 15c. Summary of the Effect of Screening and Feedback on Rates of Treatment

	Treatment		
Author / Year	**Intervention Value**	**Control Value**	***P*-Value**
Johnstone and Goldberg, 1976[94]		NR*	
Moore et al, 1978[98]		NR	
Linn and Yager, 1980[95]	13%	8%	NS*
Zung and King, 1983[103]		NR	
Magruder-Habib et al, 1990[97]	3 month: 37.5%	26.9%	*P*=0.05
	6 month: 45.8%	36.8%	
Callahan et al, 1994[91]	26%	8%	*P*=0.01
1996[92]	46%	29%	*P*=0.001
Dowrick, 1995[93]	27%	21%	NS*
Lewis et al, 1996[96]		NR	
Reifler, et al, 1996[100]		NR	
Williams et al, 1999[102]	45%	43%	NR*
Wells et al, 2000[101]	59%	50%	*P*=0.006
Whooley et al, 2000[82]	36%	43%	*P*=0.3

*NR indicates not reported; NS, not significant.

Table 15d. Summary of the Effect of Screening and Feedback on Rates of Patient Outcomes

Author / Year	Outcomes				Quality Ratings	
	What Measured	Intervention Value	Control Value	*P*-Value	Internal Validity	External Validity
Johnstone and Goldberg, 1976[94]	Mean months of depression in 1 year	4.2	6.3	*P*< 01	Fair	Fair
Moore et al, 1978[98]		NR*			Good	Good
Linn and Yager, 1980[95]		NR			Good	Good
Zung and King, 1983[103]	% with >12 point decrease on SDS at 1 month	66%	35%	*P*< 05	Fair	Fair
Magruder-Habib et al, 1990[97]		NR			Good	Fair
Callahan et al, 1994[91]	% with HAM-D <10 at 6 months	13%	12%	NR	Good	Fair
Callahan et al, 1996[92]		NR			Good	Fair
Dowrick, 1995[93]		NR			Fair	Fair

Table 15d. Summary of the Effect of Screening and Feedback on Rates of Patient Outcomes (continued)

Author / Year	Outcomes				Quality Ratings	
	What Measured	Intervention Value	Control Value	*P*-Value	Internal Validity	External Validity
Lewis et al, 1996[96]	Mean GHQ at 6 months	25.4 PROQSY 26.8 GHQ	25.9	*P*=0.12	Good	Fair
Reifler, et al, 1996[100]	Zung scale score	No difference for those screening positive for any disorder			Good	Good
Williams et al, 1999[102]	% depressed at 3 months DSM-IIIR criteria	37%	46%	*P*=0.19	Good	Good
	% with <1 DSM-IIIR criteria symptoms (generally 3 months)	48%	27%	95% CI for diff. (1-41%)		
Wells et al, 2000[101]	% depressed at 6 months	55.4%	64.4%	*P*=0.005	Good	Good
	% depressed at 12 months	54.5%	61.4%	*P*=0.04		
Whooley et al, 2000[82]	% depressed at 24 months (GDS ≥6)	42%	50%	*P*=0.30	Fair	Good

DSM indicates Diagnostic and Statistical Manual of Mental Disorders (-III, third edition; -IIIR, third edition revised; -IV, fourth edition); GDS, Geriatric Depression Scale; GHQ, General Health Questionnaire; HAM-D, Hamilton Depression Rating Scale; PROQSY, self-administered computerized assessment; SDS, Zung Self-Depression Scale.

*NR indicates not reported.

Chapter 4. Discussion

Major Findings

Depressive disorders are common in primary care settings and cause substantial morbidity and mortality. Multiple studies have documented that depressive disorders are often unrecognized or undertreated by "usual care" or nonsystematic approaches to diagnosis and therapy.[26,27] The overlap between symptoms of depression and symptoms of physical illnesses may lead to unnecessary tests and treatments in an attempt to diagnose or treat complaints that are actually caused by the depression itself. Failure to recognize and treat depression can lead to increased or prolonged disability, morbidity, and mortality, at least for those patients with more severe illnesses. Patients with less severe illnesses appear less likely to be detected but also may benefit less from treatment.

In our systematic review, we have addressed several key questions and subquestions concerning screening accuracy in various populations, pharmacotherapy and psychotherapy in adults and children, and screening outcomes in adults. The quality of evidence on these questions is summarized in Table 16, which shows our ratings (of good, fair, or poor, as defined in Appendix C) for 3 important measures relating to a body of evidence for a given key question or linkage in the analytic framework (Figure 1, Chapter 1) — aggregate internal validity, aggregate external validity, and coherence (ie, consistency).

In our systematic review, we have shown that brief, accurate, and feasible screening tests are available for detecting depressive disorders in adults and the elderly. This reflects good evidence across

the board for adults and good or fair-to-good evidence for the elderly. Recently tested shorter instruments appear to perform about as well as the longer versions evaluated in the previous 2 editions of the *Guide to Clinical Preventive Services*.[31,192] Among the elderly, specific scales appear to improve detection compared with general scales.

In addition, effective pharmacologic and psychotherapeutic treatments are available for adult primary care patients with major depression (good evidence in all 3 measures). Treatment for adults with dysthymia also appears effective although the amount of data from primary care populations is smaller than for major depression. The available data for treatment of adults with dysthymia and minor depression are less well developed but suggestive of benefit as well. Educational interventions designed to improve the quality of care have shown success in improving treatment initiation, adherence, and outcomes.[129]

The accuracy of screening tests for depression in adolescents and children has been less well studied in primary care settings but available data suggest similar levels of performance (fair-to-good evidence on all 3 measures). Treatment for adolescents with cognitive-behavioral therapy has been shown to improve depression,[146] with evidence judged to be good for internal validity and coherence and fair for external validity. The data for selective serotonin reuptake inhibitors (SSRIs) are mixed but suggestive of benefit, but tricyclic agents appear to be ineffective. The quality of the pharmacotherapy evidence for the pediatric age group is quite mixed, however (fair internal validity, good external validity, but poor coherence).

The overarching question for adult patients—whether screening and subsequent treatment is superior to treatment based on usual means of diagnosis—is controversial; evidence on aggregate internal and external validity is good, but the level of coherence in findings across these studies is only fair to poor. Data from several trials suggest that patients who are screened are more likely to be

recognized as depressed and sometimes are more likely to be treated. The effect of screening on clinical outcomes, however, has been mixed when compared with the usual care provided in studies. Further support beyond identification appears to improve treatment adherence and outcomes. The recent study by Wells et al[101] suggests that a simple 2-question screener, when coupled with a quality improvement process, can improve outcomes over 6 to 12 months in patients with a spectrum of depressive disorders.

Benefits and Harms

The potential benefits of screening and treatment of depressive disorders include reduced morbidity and mortality, improved quality-of-life functioning, and employment. It may also lower expenditures on unnecessary health care. The potential harms of screening include false-positive screening results, the adverse effects of treatment, the adverse effects and costs of treatment for patients who are incorrectly identified as being depressed, and the potential adverse effects of labeling.

The trade-offs between benefits and harms are an important component of the decision to screen or not to screen for depression. We currently have insufficient information about the harms of screening (false positives and labeling) to create a balance sheet to inform the decision to screen.

Future Research Needs

Despite a wealth of new studies concerning screening accuracy and the effectiveness of treatment for depression, key elements of the evidence base for depression screening remain insufficiently developed. The limitations are greater for children and adolescents than for adults, as reflected in Table 16. Nonetheless, additional research is needed across the age spectrum and in special populations, including the underserved and minority groups. For the adult and geriatric populations, especially those in primary care settings, further research on the identification and treatment of dysthymia and subsyndromal or "minor" depression will be a major step forward.

For all ages, outcomes to be considered should include persistent depressive symptoms and associated disability as well as appropriate outcome measures including functional status and quality of life. Such disability measures are a key element in documenting improvement for depressive illness and in reducing its staggering disease burden. In addition, investigators should examine health care utilization and ensure that their studies are sufficiently powered for detecting modest but clinically important differences.

Considerable additional research is needed for children and adolescents in both screening and treatment, particularly in primary care settings. The question of whether SSRIs are effective in adolescents and children should be addressed in additional sufficiently powered trials that are analyzed by intention to treat. Determining whether simple screening instruments can be accurate and are feasible for application in primary care settings or schools remains an important investigative step.

For both children and adults, more research is required about the harms of screening; these include issues relating to labeling, inappropriate treatment, failure to make the correct diagnosis, and

unnecessary costs. We also need better information about the optimal means of addressing a positive screening test in real-world settings, including the ability of primary care providers to conduct further diagnostic assessment, initiate treatment, and optimally use psychiatric referral.

Table 16: Summary of the Quality of Evidence for Key Questions

	Key Question	Aggregate Internal Validity	Aggregate External Validity	Coherence
1A	Screening Accuracy in Adults	Good	Good	Good
1B	Screening Accuracy in Elderly	Good	Fair - Good	Good
1C	Screening Accuracy in Children/ Adolescents	Fair - Good	Fair - Good	Fair - Good
2A	Pharmacotherapy in Adults	Good	Good	Good
2B	Psychotherapy in Adults	Fair	Good	Good
2C	Pharmacotherapy in Children	Fair	Good	Poor
2D	Psychotherapy in Children	Good	Fair	Good
3	Screening Outcomes in Adults	Good	Good	Fair-Poor

References

1. Kessler RC, McGonagle KA, Zhao S, et al. Lifetime and 12-month prevalence of DSM-IIIR psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Arch Gen Psychiatry.* 1994;51:8-19.

2. Robins L, Regier D. *Psychiatric Disorders in America*. New York: Free Press; 1991.

3. Depression Guideline Panel. *Depression in Primary Care: Volume 1. Detection and Diagnosis. Clinical Practice Guideline, Number 5.* Rockville, Maryland: U.S Department of Health and Human Services; 1993. AHCPR No. 93-0550.

4. Katon W. The epidemiology of depression in medical care. *Int J Psychiatry Med.* 1987;17:93-112.

5. Judd LL. The clinical course of unipolar major depressive disorders. *Arch Gen Psychiatry.* 1997;54:989-991.

6. Greenberg PE, Stiglin LE, Finkelstein SN, Berndt ER. The economic burden of depression in 1990. *J Clin Psychiatry.* 1993;54:405-418.

7. Blair-West GW, Mellsop GW, Eyeson-Annan ML. Down-rating lifetime suicide risk in major depression. *Acta Psychiatr Scand.* 1997;95:259-263.

8. Harris EC, Barraclough B. Suicide as an outcome for mental disorders. A meta-analysis. *Br J Psychiatry.* 1997;170:205-228.

9. Murray L, Stanley C, Hooper R, King F, Fiori-Cowley A. The role of infant factors in postnatal depression and mother-infant interactions. *Dev Med Child Neurol.* 1996;38:109-119.

10. Musselman D, Evans D, Nemeroff C. The relationship of depression to cardiovascular disease. *Arch Gen Psychiatry.* 1998;55:580-592.

11. Consensus Development Panel. NIMH/NIH Consensus Development Conference statement. Mood disorders: pharmacologic prevention of recurrences. *Am J Psychiatry.* 1985;142:469-476.

12. Pennix BW, Guralnik JM, Ferrucci L, Simonsick EM, Deeg DJ, Wallace RB. Depressive symptoms and physical decline in community-dwelling older persons. *JAMA.* 1998;276:1720-1726.

References

13. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*. Washington, DC: American Psychiatric Association; 1994.

14. Wells KB, Stewart A, Hays RD, et al. The functioning and well-being of depressed patients. Results from the Medical Outcomes Study. *JAMA.* 1989;262:914-919.

15. Howland R. General health, health care utilization, and medical comorbidity in dysthymia. *Int J Psychiatry Med.* 1993;23:211-238.

16. Wells KB, Burnam MA, Rogers W, Hays R, Camp P. The course of depression in adult outpatients. Results from the Medical Outcomes Study. *Arch Gen Psychiatry.* 1992;49:788-794.

17. Williams JW Jr, Kerber CA, Mulrow CD, Medina A, Aguilar C. Depressive disorders in primary care: prevalence, functional disability, and identification. *J Gen Intern Med.* 1995;10:7-12.

18. Wagner R, Burns BJ, Yarnall K, Sigmon A, Walker R, Gaynes BN. Minor depression in family practice: functional morbidity, comorbidity, service utilization, and outcomes. *Psychol Med*. 2000;30(6):1377-1390.

19. Cooper-Patrick L, Crum RM, Ford DE. Characteristics of patients with major depression who received care in general medical and specialty mental health settings. *Med Care.* 1994;32:15-24.

20. Paykel E, Priest R. Recognition and management of depression in general practice: consensus statement. *BMJ.* 1992;305:1198-1202.

21. Simon GE, Von Korff M, Durham ML. Predictors of outpatient mental health utilization by primary care patients in a health maintenance organization. *Am J Psychiatry.* 1994;151:908-913.

22. Song F, Freemantle N, Sheldon TA, et al. Selective serotonin reuptake inhibitors: meta-analysis of efficacy and acceptability. *BMJ.* 1993;306:683-687.

23. Kessler RC, Nelson CB, McGongale KA, Liu J, Swartz M, Blazer DG. Comorbidity of DSM-IIIR major depressive disorder in the general population: Results from the US National Comorbidity Survey. *Br J Psychiatry.* 1996;168:17-30.

24. Regier DA, Narrow WE, Rae DS, Manderscheid RW, Locke BZ, Goodwin FK. The de facto US mental and addictive disorders service system. Epidemiologic catchment area prospective 1-year prevalence rates of disorders and services. *Arch Gen Psychiatry.* 1993;50:85-94.

25. Pincus HA, Tanielian TL, Marcus SC, et al. Prescribing trends in psychotropic medications: primary care, psychiatry, and other medical specialities. *JAMA.* 1998;279:526-531.

26. Simon GE, Von Korff M. Recognition, management, and outcomes of depression in primary care. *Arch Fam Med.* 1995;4:99-105.

27. Gerber PD, Barrett J, Barrett J, Manheimer E, Whiting R, Smith R. Recognition of depression by internists in primary care: a comparison of internist and 'gold standard' psychiatric assessments. *J Gen Intern Med.* 1989;4:7-13.

28. Klinkman MS. Competing demands in psychosocial care. A model for the identification and treatment of depressive disorders in primary care. *Gen Hosp Psychiatry.* 1997;19:98-111.

29. Borowsky SJ, Rubenstein LV, Meredith LS, Camp P, Jackson-Triche M, Wells KB. Who is at risk of nondetection of mental health problems in primary care? *J Gen Intern Med.* 2000;15:381-388.

30. Linde K, Mulrow CD. St. John's wort for depression (Cochrane Review). *The Cochrane Library.* 1998.

31. US Preventive Services Task Force. *Guide to Clinical Preventive Services*. 2nd ed. Alexandria, Va.: International Medical Publishing; 1996.

32. American College of Physicians. ACP-ASIM Clinical Practice Guidelines: Current and Future Projects. Available at: http://www.acponline.org/sci-policy/guidelines/projects.htm. Accessed November 15, 1999.

33. Feightner JW. Early detection of depression. In: Canadian Task Force on the Periodic Health Examination. *Canadian Guide to Clinical Preventive Health Care.* Ottawa: Health Canada; 1994:450-454.

34. Arthur A, Jagger C, Lindesay J, Graham C, Clarke M. Using an annual over-75 health check to screen for depression: validation of the short Geriatric Depression Scale (GDS15) within general practice. *Int J Geriatr Psychiatry.* 1999;14:431-439.

35. Banerjee S, Shamash K, MacDonald AJ, Mann AH. The use of the SelfCARE(D) as a screening tool for depression in the clients of local authority home care services—a preliminary study. *Int J Geriatr Psychiatry* . 1998;13:695-699.

36. Bashir K, Blizard R, Jenkins R, Mann A. Validation of the 12-item general health questionnaire in British general practice. *Primary Care Psychiatry.* 1996;2:4-7.

37. Beekman AT, Deeg DJ, Van Limbeek J, Braam AW, De Vries MZ, Van Tilburg W. Criterion validity of the Center for Epidemiologic Studies Depression scale (CES-D): results from a community-based sample of older subjects in The Netherlands. *Psychol Med.* 1997;27:231-235.

38. Bird AS, Macdonald AJD, Mann AH, Philpot MP. Preliminary experience with the Selfcare (D): A self-rating depression questionnaire for use in elderly, non-institutionalized subjects. *Int J Geriatr Psychiatry* . 1987;2:31-38.

39. Broadhead WE, Leon AC, Weissman MM, et al. Development and validation of the SDDS-PC screen for multiple mental disorders in primary care. *Arch Fam Med.* 1995;4:211-219.

40. Burnam MA, Wells KB, Leake B, Landsverk J. Development of a brief screening instrument for detecting depressive disorders. *Med Care.* 1988;26:775-789.

41. D'Ath P, Katona P, Mullan E, Evans S, Katona C. Screening, detection and management of depression in elderly primary care attenders. I: The acceptability and performance of the 15 item Geriatric Depression Scale (GDS15) and the development of short versions. *Fam Pract.* 1994;11:260-266.

42. Fechner-Bates S, Coyne JC, Schwenk TL. The relationship of self-reported distress to depressive disorders and other psychopathology. *J Consult Clin Psychol* . 1994;62:550-559.

43. Finlay-Jones RA, Murphy E. Severity of psychiatric disorder and the 30-item general health questionnaire. *Br J Psychiatry.* 1979;134:609-616.

44. Geisser ME, Roth RS, Robinson ME. Assessing depression among persons with chronic pain using the Center for Epidemiological Studies-Depression Scale and the Beck Depression Inventory: a comparative analysis. *Clin J Pain.* 1997;13:163-170.

45. Gerety MB, Williams JW Jr, Mulrow CD, et al. Performance of case-finding tools for depression in the nursing home: influence of clinical and functional characteristics and selection of optimal threshold scores. *J Am Geriatr Soc.* 1994;42:1103-1109.

46. Goldberg DP, Blackwell B. Psychiatric illness in general practice. A detailed study using a new method of case identification. *BMJ.* 1970;1:439-443.

47. Hendrie HC, Callahan CM, Levitt EE, et al. Prevalence rates of major depressive disorders: The effects of varying the diagnostic criteria in an older primary care population. *Am J Geriatr Psychiatry* . 1995;3:119-131.

48. Holcomb WL Jr, Stone LS, Lustman PJ, Gavard JA, Mostello DJ. Screening for depression in pregnancy: characteristics of the Beck Depression Inventory. *Obstet Gynecol.* 1996;88:1021-1025.

49. Hoyl MT, Alessi CA, Harker JO, et al. Development and testing of a five-item version of the Geriatric Depression Scale. *J Am Geriatr Soc.* 1999;47:873-878.

References

50. Irwin M, Artin KH, Oxman MN. Screening for depression in the older adult: criterion validity of the 10-item Center for Epidemiological Studies Depression Scale (CES-D). *Arch Intern Med.* 1999;159:1701-1704.

51. Klinkman MS, Coyne JC, Gallo S, Schwenk TL. Can case-finding instruments be used to improve physician detection of depression in primary care? *Arch Fam Med.* 1997;6:567-573.

52. Leon AC, Olfson M, Weissman MM, et al. Brief screens for mental disorders in primary care. *J Gen Intern Med.* 1996;11:426-430.

53. Leung KK, Lue BH, Lee MB, Tang LY. Screening of depression in patients with chronic medical diseases in a primary care setting. *Fam Pract.* 1998;15:67-75.

54. Lewinsohn PM, Seeley JR, Roberts RE, Allen NB. Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. *Psychol Aging.* 1997;12:277-287.

55. Lustman PJ, Clouse RE, Griffith LS, Carney RM, Freedland KE. Screening for depression in diabetes using the Beck Depression Inventory. *Psychosom Med.* 1997;59:24-31.

56. Lyness JM, Noel TK, Cox C, King DA, Conwell Y, Caine ED. Screening for depression in elderly primary care patients. A comparison of the Center for Epidemiologic Studies-Depression Scale and the Geriatric Depression Scale. *Arch Intern Med.* 1997;157:449-454.

57. Myers JK, Weissman MM. Use of a self-report symptom scale to detect depression in a community sample. *Am J Psychiatry.* 1980;137:1081-1084.

58. Nagel R, Lynch D, Tamburrino M. Validity of the medical outcomes study depression screener in family practice training centers and community settings. *Fam Med.* 1998;30:362-365.

59. Neal RM, Baldwin RC. Screening for anxiety and depression in elderly medical outpatients. *Age Aging.* 1994;23:461-464.

60. Okimoto JT, Barnes RF, Veith RC, Raskind MA, Inui TS, Carter WB. Screening for depression in geriatric medical patients. *Am J Psychiatry.* 1982;139:799-802.

61. Parkerson GR Jr, Broadhead WE. Screening for anxiety and depression in primary care with the Duke Anxiety-Depression Scale. *Fam Med.* 1997;29:177-181.

62. Salokangas RK, Poutanen O, Stengard E. Screening for depression in primary care. Development and validation of the Depression Scale, a screening instrument for depression. *Acta Psychiatr Scand.* 1995;92:10-16.

63. Schulberg HC, Saul M, McClelland M, Ganguli M, Christy W, Frank R. Assessing depression in primary medical and psychiatric practices. *Arch Gen Psychiatry.* 1985;42:1164-1170.

64. Spitzer RL, Kroenke K, Williams JBW. Validation and utility of a self-report version of PRIME-MD. *JAMA.* 1999;282:1737-1744.

65. Steer RA, Cavalieri TA, Leonard DM, Beck AT. Use of the Beck Depression Inventory for Primary Care to screen for major depression disorders. *Gen Hosp Psychiatry.* 1999;21:106-111.

66. Turk DC, Okifuji A. Detecting depression in chronic pain patients: adequacy of self-reports. *Behav Res Ther.* 1994;32:9-16.

67. Upadhyaya AK, Stanley I. Detection of depression in primary care: comparison of two self-administered scales. *Int J Geriatr Psychiatry.* 1997;12:35-37.

68. van Marwijk HW, Wallace P, de Bock GH, Hermans J, Kaptein AA, Mulder JD. Evaluation of the feasibility, reliability and diagnostic value of shortened versions of the geriatric depression scale. *Br J Gen Pract.* 1995;45:195-199.

69. Von Korff M, Shapiro S, Burke JD, et al. Anxiety and depression in a primary care clinic. Comparison of Diagnostic Interview Schedule, General Health Questionnaire, and practitioner assessments. *Arch Gen Psychiatry.* 1987;44:152-156.

70. Weyerer S, Killmann U, Ames D, Allen N. The even briefer assessment scale for depression (EBAS DEP): its suitability for the elderly in geriatric care in english- and german-speaking countries. *Int J Geriatr Psychiatry.* 1999;14:473-480.

71. Whooley MA, Avins AL, Miranda J, Browner WS. Case-Finding Instruments for Depression. *J Gen Intern Med.* 1997;12:439-445.

72. Wickberg B, Hwang CP. The Edinburgh Postnatal Depression Scale: validation on a Swedish community sample. *Acta Psychiatr Scand.* 1996;94:181-184.

73. Zich JM, Attkisson CC, Greenfield TK. Screening for depression in primary care clinics: the CES-D and the BDI. *Int J Psychiatry Med.* 1990;20:259-277.

74. Appleby L, Warner R, Whitton A, Faragher B. A controlled study of fluoxetine and cognitive-behavioural counselling in the treatment of postnatal depression. *BMJ.* 1997;314:932-936.

75. Mynors-Wallis LM, Gath DH, Day A, Baker F. Randomized controlled trial of problem solving treatment, antidepressant medication, and combined treatment for major depression in primary care. *BMJ.* 2000;320:26-30.

76. Malt UF, Robak OH, Madsbu HP, Bakke O, Loeb M. The Norwegian naturalistic treatment study of depression in general practice (NORDEP)-I: randomised double blind study. *BMJ.* 1999;318:1180-1184.

77. Mynors-Wallis LM, Gath DH, Lloyd-Thomas AR, Tomlinson D. Randomised controlled trial comparing problem solving treatment with amitriptyline and placebo for major depression in primary care. *BMJ.* 1995;310:441-445.

78. Schulberg HC, Block MR, Madonia MJ, et al. Treating major depression in primary care practice. Eight-month clinical outcomes. *Arch Gen Psychiatry.* 1996;53:913-919.

79. Brown C, Schulberg HC, Madonia MJ, Shear MK, Houck PR. Treatment outcomes for primary care patients with major depression and lifetime anxiety disorders. *Am J Psychiatry.* 1996;153:1293-1300.

80. Coulehan JL, Schulberg HC, Block MR, Madonia MJ, Rodriguez E. Treating depressed primary care patients improves their physical, mental, and social functioning. *Arch Intern Med.* 1997;157:1113-1120.

81. Scott AI, Freeman CP. Edinburgh primary care depression study: treatment outcome, patient satisfaction, and cost after 16 weeks. *BMJ.* 1992;304:883-887.

82. 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.

83. Blackburn IM, Bishop S, Glen AI, Whalley LJ, Christie JE. The efficacy of cognitive therapy in depression: a treatment trial using cognitive therapy and pharmacotherapy, each alone and in combination. *Br J Psychiatry.* 1981;139:181-189.

84. Holden JM, Sagovsky R, Cox JL. Counselling in a general practice setting: controlled study of health visitor intervention in treatment of postnatal depression. *BMJ.* 1989;298:223-226.

85. Katon W, Robinson P, Von Korff M, et al. A multifaceted intervention to improve treatment of depression in primary care. *Arch Gen Psychiatry.* 1996;53:924-932.

86. Lynch DJ, Tamburrino MB, Nagel R. Telephone counseling for patients with minor depression: preliminary findings in a family practice setting. *J Fam Pract.* 1997;44:293-298.

87. Miranda J, Munoz R. Intervention for minor depression in primary care patients. *Psychosom Med.* 1994;56:136-141.

88. Ross M, Scott M. An evaluation of the effectiveness of individual and group cognitive therapy in the treatment of depressed patients in an inner city health centre. *J R Coll Gen Pract.* 1985;35:239-242.

89. Scott C, Tacchi MJ, Jones R, Scott J. Acute and one-year outcome of a randomised controlled trial of brief cognitive therapy for major depressive disorder in primary care. *Br J Psychiatry.* 1997;171:131-134.

90. Teasdale JD, Fennell MJ, Hibbert GA, Amies PL. Cognitive therapy for major depressive disorder in primary care. *Br J Psychiatry.* 1984;144:400-406.

91. 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;42:839-846.

92. Callahan CM, Dittus RS, Tierney WM. Primary care physicians' medical decision making for late-life depression. *J Gen Intern Med.* 1996;11:218-225.

93. Dowrick C. Does testing for depression influence diagnosis or management by general practitioners? *Fam Pract.* 1995;12:461-465.

94. Johnstone A, Goldberg D. Psychiatric screening in general practice. A controlled trial. *Lancet.* 1976;1:605-608.

95. Linn LS, Yager J. The effect of screening, sensitization, and feedback on notation of depression. *J Med Educ.* 1980;55:942-949.

96. Lewis G, Sharp D, Bartholomew J, Pelosi AJ. Computerized assessment of common mental disorders in primary care: effect on clinical outcome. *Fam Pract.* 1996;13:120-126.

97. 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;28:239-250.

98. Moore JT, Silimperi DR, Bobula JA. Recognition of depression by family medicine residents: the impact of screening. *J Fam Pract.* 1978;7:509-513.

99. Rand EH, Badger LW, Coggins DR. Toward a resolution of contradictions. Utility of feedback from the GHQ. *Gen Hosp Psychiatry.* 1988;10:189-196.

100. 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;156:2593-2599.

101. Wells KB, Sherbourne C, Schoenbaum M, et al. Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. *JAMA.* 2000;283:212-220.

102. Williams JW Jr, Mulrow CD, Kroenke K, et al. Case-finding for depression in primary care: a randomized trial. *Am J Med.* 1999;106:36-43.

103. Zung WW, King RE. Identification and treatment of masked depression in a general medical practice. *J Clin Psychiatry.* 1983;44:365-368.

104. 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;122:913-921.

105. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition: Primary Care Version (DSM-IV-PC).* 1st ed. Washington, DC: American Psychiatric Press; 1995.

106. Furukawa T, Hirai T, Kitamura T, Takahashi K. Application of the Center for Epidemiological Studies Depression Scale among first-visit psychiatric patients: a new approach to improve its performance. *J Affect Disord.* 1997;46:1-13.

107. Conde MV, Williams JW Jr, Mulrow CD. Targeting depression interviewing: an exercise in diagnostic reasoning. *J Gen Intern Med.* 1998;13:262-265.

108. Gutterman EM, O'Brien JD, Young JG. Structured diagnostic interviews for children and adolescents: current status and future directions. *J Am Acad Child Adolesc Psychiatry.* 1987;26:621-630.

109. Kashani JH, Sherman DD, Parker DR, Reid JC. Utility of the Beck Depression Inventory with clinic-referred adolescents. *J Am Acad Child Adolesc Psychiatry.* 1990;29:278-282.

110. Ambrosini PJ, Metz C, Bianchi MD, Rabinovich H, Undie A. Concurrent validity and psychometric properties of the Beck Depression Inventory in outpatient adolescents. *J Am Acad Child Adolesc Psychiatry.* 1991;30:51-57.

111. Roberts RE, Lewinsohn PM, Seeley JR. Screening for adolescent depression: a comparison of depression scales. *J Am Acad Child Adolesc Psychiatry.* 1991;30:58-66.

112. Barrera M Jr, Garrison-Jones CV. Properties of the Beck Depression Inventory as a screening instrument for adolescent depression. *J Abnorm Child Psychol.* 1988;16:263-273.

113. Whitaker A, Johnson J, Shaffer D, et al. Uncommon troubles in young people: prevalence estimates of selected psychiatric disorders in a nonreferred adolescent population. *Arch Gen Psychiatry.* 1990;47:487-496.

114. Winter LB, Steer RA, Jones-Hicks L, Beck AT. Screening for major depression disorders in adolescent medical outpatients with the Beck Depression Inventory for Primary Care. *J Adolesc Health.* 1999;24:389-394.

References

115. Costello EJ, Angold A. Scales to assess child and adolescent depression: checklists, screens, and nets. *J Am Acad Child Adolesc Psychiatry.* 1988;27:726-737.

116. Garrison CZ, Addy CL, Jackson KL, McKeown RE, Waller JL. The CES-D as a screen for depression and other psychiatric disorders in adolescents. *J Am Acad Child Adolesc Psychiatry.* 1991;30:636-641.

117. Fendrich M, Weissman MM, Warner V. Screening for depressive disorder in children and adolescents: validating the Center for Epidemiologic Studies Depression Scale for Children. *Am J Epidemiol.* 1990;131:538-551.

118. Angold A, Costello EJ, Messer SC, Pickles A. Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *Int J Methods Psychiatric Res.* 1995;5:237-249.

119. King CA, Katz SH, Ghaziuddin N, Brand E, Hill E, McGovern L. Diagnosis and assessment of depression and suicidality using the NIMH Diagnostic Interview Schedule for Children (DISC-2.3). *J Abnorm Child Psychol.* 1997;25:173-181.

120. Jensen PS, Salzberg AD, Richters JE, Watanabe HK. Scales, diagnoses, and child psychopathology: I. CBCL and DISC relationships. *J Am Acad Child Adolesc Psychiatry.* 1993;32:397-406.

121. Stancin T, Palermo TM. A review of behavioral screening practices in pediatric settings: do they pass the test? *J Dev Behav Pediatr.* 1997;18:183-194.

122. Marton P, Churchard M, Kutcher S, Korenblum M. Diagnostic utility of the Beck Depression Inventory with adolescent psychiatric outpatients and inpatients. *Can J Psychiatry.* 1991;36:428-431.

123. Canning EH, Kelleher K. Performance of screening tools for mental health problems in chronically ill children. *Arch Pediatr Adolesc Med.* 1994;148:272-278.

124. Mulrow CD, Williams JW Jr, Chiquette E, et al. Efficacy of newer pharmacotherapies for treating depression in primary care patients. *Am J Med.* 2000;108:54-64.

125. Lima MS, Moncrieff J. Drugs versus placebo for dysthymia (Cochrane Review). *The Cochrane Library.* 1999.

126. Gill D, Hatcher S. Antidepressant drugs in depressed patients who also have a physical illness (Cochrane Review). *The Cochrane Library.* 1999.

127. Evans M, Hammond M, Wilson K, Lye M, Copeland J. Placebo-controlled treatment trial of depression in elderly physically ill patients. *Int J Geriatr Psychiatry.* 1997;12:817-824.

128. Williams JW Jr, Mulrow CD, Chiquette E, Noel PH, Aguilar C, Cornell J. A systematic review of newer pharmacotherapies for depression in adults: evidence report summary. *Ann Intern Med.* 2000;132:743-756.

129. Katon W, Von Korff M, Lin E, et al. Collaborative management to achieve treatment guidelines. Impact on depression in primary care. *JAMA.* 1995;273:1026-1031.

130. Simon GE, Von Korff M, Rutter C, Wagner E. Randomised trial of monitoring, feedback, and management of care by telephone to improve treatment of depression in primary care. *BMJ.* 2000;320:550-554.

131. 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;9:345-351.

132. Llewellyn-Jones RH, Baikie KA, Smithers H, Cohen J, Snowdon J, Tennant CC. Multifaceted shared care intervention for late life depression in residential care: randomised controlled trial. *BMJ.* 1999;319:676-682.

133. Peveler R, George C, Kinmonth AL, Campbell M, Thompson C. Effect of antidepressant drug counselling and information leaflets on adherence to drug treatment in primary care: randomised controlled trial. *BMJ.* 1999;319:612-615.

134. Boyer P, Lecrubier Y. Atypical antipsychotic drugs in dysthymia: placebo controlled studies of amisulpride versus imipramine versus amineptine. *Eur Psychiatry.* 1996;11:135S-140S.

135. Mynors-Wallis L. Problem-solving treatment: evidence for effectiveness and feasibility in primary care. *Int J Psychiatry Med.* 1996;26:249-262.

136. Robins L, Helzer J, Croughan J, Ratcliff K. National Institute of Mental Health Diagnostic Interview Schedule. *Arch Gen Psychiatry.* 1981;38:381-391.

137. Kendrick T. Why can't GPs follow guidelines on depression? We must question the basis of the guidelines themselves (letter). *BMJ.* 2000;320:200-201.

138. Spitzer RL, Endicott J, Robins E. Research Diagnostic Criteria: Rationale and reliability. *Arch Gen Psychiatry.* 1978;35:773-782.

139. Simon GE, Von Korff M, Heiligenstein JH, et al. Initial antidepressant choice in primary care. Effectiveness and cost of fluoxetine vs tricyclic antidepressants. *JAMA.* 1996;275:1897-1902.

140. Reynolds CF III, Frank E, Perel JM, et al. Nortriptyline and interpersonal psychotherapy as maintenance therapies for recurrent major depression: a randomized controlled trial in patients older than 59 years. *JAMA.* 1999;281:39-45.

141. Keller MB, McCullough JP, Klein DN, et al. A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. *N Engl J Med.* 2000;342:1462-1470.

142. Zigmond A, Snaith R. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand.* 1983;67:361-370.

143. Jaycox LH, Reivich KJ, Gillham J, Seligman ME. Prevention of depressive symptoms in school children. *Behav Res Ther.* 1994;32:801-816.

144. Clarke GN, Hawkins W, Murphy M, Sheeber LB, Lewinsohn PM, Seeley JR. Targeted prevention of unipolar depressive disorder in an at-risk sample of high school adolescents: a randomized trial of a group cognitive intervention. *J Am Acad Child Adolesc Psychiatry.* 1995;34:312-321.

145. Lamb JM, Puskar KR, Sereika SM, Corcoran M. School-based intervention to promote coping in rural teens. *MCN Am J Matern Child Nurs.* 1998;23:187-194.

146. Reinecke MA, Ryan NE, DuBois DL. Cognitive-behavioral therapy of depression and depressive symptoms during adolescence: a review and meta-analysis. *J Am Acad Child Adolesc Psychiatry.* 1998;37:26-34.

147. Wood A, Harrington R, Moore A. Controlled trial of a brief cognitive-behavioural intervention in adolescent patients with depressive disorders. *J Child Psychol Psychiatry.* 1996;37:737-746.

148. Reynolds WM, Coats KI. A comparison of cognitive-behavioral therapy and relaxation training for the treatment of depression in adolescents. *J Consult Clin Psychol.* 1986;54:653-660.

149. Lewinsohn P, Clarke G, Hops H, Andrews J. Cognitive-behavioural treatment for adolescents. *Behavior Therapy.* 1990;21:385-401.

150. Lerner M, Clum G. Treatment of suicide ideators: a problem-solving approach. *Behavior Therapy.* 1990;21:403-411.

151. Kahn J, Kehle T, Jensen W, Clark E. Comparison of cognitive-behavioral, relaxation, and self-modeling interventions for depression among middle-school students. *School Psychology Review.* 1990;19:196-211.

152. Kazdin A, Bass D, Ayers W, Rodgers A. Empirical and clinical focus of child and adolescent psychotherapy needs. *J Consult Clin Psychol.* 1990;58:729-740.

153. Weisz J, Weiss B, Alicke M, Klotz M. Effectiveness of psychotherapy with children and adolescents: a meta-analysis for clinicians. *Consul and Clin Psychol.* 1987;55:542-549.

154. Berney T, Kolvin I, Bhate SR, et al. School phobia: a therapeutic trial with clomipramine and short-term outcome. *Br J Psychiatry.* 1981;138:110-118.

155. Kramer AD, Feguine RJ. Clinical effects of amitriptyline in adolescent depression. *J Am Acad Child Psychiatry.* 1981;20:636-644.

156. Petti TA, Law W. Imipramine treatment of depressed children: a double-blind pilot study. *J Clin Psychopharmacol.* 1982;2:107-110.

157. Preskorn SH, Weller EB, Weller RA. Depression in children: relationship between plasma imipramine levels and response. *J Clin Psychiatry.* 1982;43:450-453.

158. Kashani JH, Shekim WO, Reid JC. Amitriptyline in children with major depressive disorder: a double-blind cross-over pilot study. *J Am Acad Child Psychiatry.* 1984;23:348-351.

159. Geller B, Cooper TB, McCombs HG, Graham D, Wells J. Double-blind, placebo-controlled study of nortriptyline in depressed children using a "fixed plasma level" design. *Psychopharmacol Bull.* 1989;25:101-108.

160. Puig-Antich J, Perel JM, Lupatkin W, et al. Imipramine in prepubertal major depressive disorders. *Arch Gen Psychiatry.* 1987;44:81-89.

161. Bernstein GA, Garfinkel BD, Borchardt CM. Comparative studies of pharmacotherapy for school refusal. *J Am Acad Child Adolesc Psychiatry.* 1990;29:773-781.

162. Geller B, Cooper TB, Graham DL, Marsteller FA, Bryant DM. Double-blind placebo-controlled study of nortriptyline in depressed adolescents using a "fixed plasma level" design. *Psychopharmacol Bull.* 1990;26:85-90.

163. Hughes CW, Preskorn SH, Weller EB, Weller RA, Hassanein R, Tucker S. The effect of concomitant disorders in childhood depression on predicting treatment response. *Psychopharmacol Bull.* 1990;26:235-238.

164. Boulos C, Kutcher S, Marton P, Simeon J, Ferguson B, Roberts N. Response to desipramine treatment in adolescent major depression. *Psychopharmacol Bull.* 1991;27:59-65.

165. Geller B, Cooper TB, Graham DL, Fetner HH, Marsteller FA, Wells JM. Pharmacokinetically designed double-blind placebo-controlled study of nortriptyline in 6- to 12-year-olds with major depressive disorder. *J Am Acad Child Adolesc Psychiatry.* 1992;31:34-44.

166. Hazell P, O'Connell D, Heathcote D, Robertson J, Henry D. Efficacy of tricyclic drugs in treating child and adolescent depression: a meta-analysis. *BMJ.* 1995;310:897-901.

167. Kutcher S, Boulos C, Ward B, et al. Response to desipramine treatment in adolescent depression: a fixed-dose, placebo-controlled trial. *J Am Acad Child Adolesc Psychiatry.* 1994;33:686-694.

168. Kye CH, Waterman GS, Ryan ND, et al. A randomized, controlled trial of amitriptyline in the acute treatment of adolescent major depression. *J Am Acad Child Adolesc Psychiatry.* 1996;35:1139-1144.

169. Sallee FR, Vrindavanam NS, Deas-Nesmith D, Carson SW, Sethuraman G. Pulse intravenous clomipramine for depressed adolescents: double-blind, controlled trial. *Am J Psychiatry.* 1997;154:668-673.

170. Birmaher B, Waterman GS, Ryan ND, et al. Randomized, controlled trial of amitriptyline versus placebo for adolescents with "treatment-resistant" major depression. *J Am Acad Child Adolesc Psychiatry.* 1998;37:527-535.

171. Klein RG, Mannuzza S, Koplewicz HS, et al. Adolescent depression: controlled desipramine treatment and atypical features. *Depress Anxiety.* 1998;7:15-31.

172. Geller B, Reising D, Leonard HL, Riddle MA, Walsh BT. Critical review of tricyclic antidepressant use in children and adolescents. *J Am Acad Child Adolesc Psychiatry.* 1999;38:513-516.

173. Colle LM, Belair JF, DiFeo M, Weiss J, LaRoche C. Extended open-label fluoxetine treatment of adolescents with major depression. *J Child Adolesc Psychopharmacol.* 1994;4:225-232.

174. Ghaziuddin N, Naylor MW, King CA. Fluoxetine in tricyclic refractory depression in adolescents. *Depression.* 1995;2:287-291.

175. Nguyen N, Biu B, Wettllesey S, DiGiacome D, Bui T. A pilot study: pattern analysis of fluoxetine in childhood depression. *Psychopharmacol Bull.* 1995;31:602.

176. Armitage R, Yonkers K, Cole D, Rush AJ. A multicenter, double-blind comparison of the effects of nefazodone and fluoxetine on sleep architecture and quality of sleep in depressed outpatients. *J Clin Psychopharmacol.* 1997;17:161-168.

177. Riggs PD, Coffman LM, Crowley TJ. Fluoxetine in drug-dependent delinquents with major depression: an open trial. *J Child Adolesc Psychopharmacol.* 1997;7:87-95.

178. Tierney E, Paramjit TJ, Llinas JF, Rosenberg L, Riddle MA. Sertraline for major depression in children and adolescents: preliminary clinical experience. *J Child Adolesc Psychopharmol.* 1995;5:13-27.

179. McConville BJ, Minnery KL, Sorter JT, West SA, Friedman LM, Christian K. An open study of the effects of sertraline on adolescent major depression. *J Child Adolesc Psychopharmacol.* 1996;6:41-51.

180. Ambrosini PJ, Wagner KD, Biederman J, et al. Multicenter open-label sertraline study in adolescent outpatients with major depression. *J Am Acad Child Adolesc Psychiatry.* 1999;38:566-572.

181. Rey-Sanchez F, Gutierrez-Casares JR. Paroxetine in children with major depressive disorder: an open trial. *J Am Acad Child Adolesc Psychiatry.* 1997;36:1443-1447.

182. Masi G, Marcheschi M, Pfanner P. Paroxetine in depressed adolescents with intellectual disability: an open label study. *J Intellect Disabil Res.* 1997;41:268-272.

183. Kiev A, Feiger A. A double-blind comparison of fluvoxamine and paroxetine in the treatment of depressed outpatients. *J Clin Psychiatry.* 1997;58:146-152.

184. Simeon JG, Dinicola VF, Ferguson HB, Copping W. Adolescent depression: a placebo-controlled fluoxetine treatment study and follow-up. *Prog Neuropsychopharmacol Biol Psychiatry.* 1990;14:791-795.

185. Emslie GJ, Rush AJ, Weinberg WA, et al. A double-blind, randomized, placebo-controlled trial of fluoxetine in children and adolescents with depression. *Arch Gen Psychiatry.* 1997;54:1031-1037.

186. Hughes CW, Emslie GJ, Crimson ML, et al. The Texas Children's Medication Algorithm Project: report of the Texas Consensus Conference Panel on Medication Treatment of Childhood Major Depressive Disorder. *J Am Acad Child Adolesc Psychiatry.* 1999;38:1442-1454.

187. Texas Department of Mental Health and Mental Retardation. The Texas Medication Algorithm Project. Available at: http://www.mhmr.state.tx.us/centraloffice/medicaldirector/tmapover.html. Accessed December 9, 1999.

188. American Academy of Child and Adolescent Psychiatrists. Practice parameters for the assessment and treatment of children and adolescents with depressive disorders. *J Am Acad Child Adolesc Psychiatry.* 1998;37:63S-83S.

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

References

190. Robins LN, Wing JK, Wittchen HU, et al. The Composite International Diagnostic Interview: An epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Arch Gen Psychiatry.* 1988;45:1069-1077.

191. Friedman RJ, Butler LF. *Development and Evaluation of a Test Battery to Assess Childhood Depression.* Ottawa, Canada: Health and Welfare Canada; 1979.

192. US Preventive Services Task Force. *Guide to Clinical Preventive Services: An Assessment of the Effectiveness of 169 Interventions: Report of the U.S. Preventive Services Task Force*. Baltimore, Md.: Williams & Wilkins; 1989.

APPENDIX A

Acknowledgments

Acknowledgments

This study was supported by Contract 290-97-0011 from the Agency for Healthcare Research and Quality(AHRQ) (Task No. 3 to support the U.S. Preventive Services Task Force). We acknowledge the ongoing guidance and assistance of David Atkins, MD, MPH, Director of the Preventive Services Program at AHRQ; Dana Best, MD, MPH, AHRQ Task Order Officer; and the continuing support of Jacqueline Besteman, JD, MA, the EPC Program Officer. The investigators deeply appreciate the considerable support and contributions of faculty and staff from the University of North Carolina at Chapel Hill – Timothy S. Carey, MD, M.P.H., Gordon H. DeFriese, PhD, Anne Jackman, MSW, and Carol Krasnov.

They are equally grateful to Linda Lux, MPA, and Sonya Sutton, BSPH, from Research Triangle Institute for substantive project assistance and to Sheila White and Loraine Monroe for superior secretarial support.

We appreciate the efforts of the following external peer reviewers who provided insightful and constructive suggestions for improvements in the systematic evidence review: David Brody, MD, Hahnemann University Hospital, Philadelphia, PA; Elizabeth Costello, PhD, Duke University, Durham, NC; Allen Dietrich, MD, Dartmouth Medical School, Hanover, NH; Wayne Katon, MD, University of Washington, Seattle, WA; Kathryn Magruder, Medical University of South Carolina, Charleston, SC; Mary Mitchell, American College of Obstetricians and Gynecologists, Washington, DC; Herbert Schulberg, University of Pittsburgh, Pittsburgh, PA; H. R. Taylor, MD, MC, USN; Mary Whooley, MD, VA Medical Center, San Francisco, CA.

APPENDIX B

Glossary

Glossary of Acronyms, Abbreviations and Initialisms for Screening for Depression

Abbreviation/Acronym	Phrase, Term, Name of Instrument
AACAP	American Academy of Child and Adolescent Psychiatrists
AAP	American Academy of Pediatrics
ARR	Absolute risk reduction
ACP	American College of Physicians
AHCPR	Agency for Health Care Policy and Research
AHRQ	Agency for Healthcare Research and Quality
BDI	Beck Depression Inventory
BDI-PC	Beck Depression Inventory for Primary Care
CAGE	(Cut down, Annoyed, Guilty, Eye-opener) Screening Questionnaire
CAS	Child Assessment Schedule
CBT	Cognitive-behavioral therapy (sometimes, cognitive therapy)
CBCL	Child Behavior Checklist
CDI	Children's Depression Inventory
CDRS-R	Children's Depression Rating Schedule - Revised
CDS	Child Depression Scale
CES-D	Center for Epidemiology Study Depression scale
CED-DC (or D-C)	Center for Epidemiology Study Depression scale for Children
CGI	Clinical Global Impressions
CI	Confidence interval
CSRS	Children's Self-report Rating Scale
DACL	Depressive Adjective Checklist
DEPS	The Depression Scale
DICA	Diagnostic Interview for Children and Adolescents
DIS	Diagnostic Interview Schedule
DISC	Diagnostic Interview Schedule for Children
DSM	Diagnostic and Statistical Manual of Mental Disorders (III, third edition; IIIR, third edition revised; IV, fourth edition)
DSRS	Depression Self-Rating Scale
Duke AD	Duke Anxiety-Depression Scale
EPC	Evidence-based Practice Center
GAS	Global Assessment Score
GDS	Geriatric Depression Scale
GHQ	General Health Questionnaire
HAM-D	Hamilton Depression Rating Scale
HIV	Human immunodeficiency virus
HSCL	Hopkins Symptomatic Checklist (as in HCSL-20)

ICD	International Classification of Diseases
IDS	Inventory of Depressive Symptomatology
IPT	Interpersonal therapy (interpersonal psychotherapy)
K-SADS	Schedule for Affective Disorders and Schizophrenia for School-age Children
MADRS:	Montgomery-Asberg Depression Rating Scale
MAOI	Monoamine oxidase inhibitors
MDD	Major depression disorder(s)
MHI	Mental Health Index (as in MHI-5)
MOS	Medical Outcomes Study
NNT	Number needed to treat
OR	Odds ratio
PRIME-MD	Primary Care Evaluation of Mental Disorders
PSC	Pediatric Symptom Checklist
RADS	Reynolds Adolescent Depression Scale
RCT	Randomized controlled trial
RDC	Research Diagnostic Criteria
ROC	Receiver operating characteristics (curve)
RR	Risk ratio
SCI	Structured clinical interview
SCID	Structured Clinical Interview for DSM-III-R (or -IV)
SCL-90	Symptom Checklist 90
SDDS-PC	Symptom Drive Diagnostic System – Primary Care
SER	Systematic evidence review
SDS	Zung self-depression screener
SF	Short Form (for MOS SF-36)
SIP	Sickness Impact Profile
SSLR	Stratum-specific likelihood ratio
SSRI	Selective serotonin reuptake inhibitor
TCA	Tricyclic antidepressants
TMAP	Texas Medication Algorithm Project
USPSTF	U.S. Preventive Services Task Force
WHO	World Health Organization

APPENDIX C

Grading System

Criteria for Grading the Internal Validity of Individual Studies

Introduction

The Methods Work Group for the U.S. Preventive Services Task Force (USPSTF) developed a set of criteria by which the quality of individual studies could be evaluated in terms of both internal validity and external validity. The USPSTF accepted the criteria, and the associated definitions of quality categories, that relate to internal validity at its September 1999 quarterly meeting.

This document describes the criteria relating to internal validity and the procedures that topic teams will follow for all updates and new assessments in making these judgments. The overall evaluation for each study is recorded in the Evidence Tables in Appendix D.

All topic teams will use initial "filters" to select studies for review that deal most directly with the question at issue and that are applicable to the population at issue. Thus, studies of any design that use outdated technology or that use technology that is not feasible for primary care practice may be filtered out before the abstraction stage, depending on the topic and the decisions of the topic team. The teams will justify such exclusion decisions if there could be reasonable disagreement about this step. The criteria below are meant for those studies that pass this initial filter.

Design-Specific Criteria and Quality Category Definitions

Presented below are a set of minimal criteria for each study design and then a general definition of 3 categories, "good," "fair," and "poor," based on those criteria. These specifications are not meant to be rigid rules but rather are intended to be general guidelines, and individual exceptions, when explicitly explained and justified, can be made. In general, a "good" study is one that meets all criteria well. A "fair" study is one that does not meet (or it is not clear that it meets) at least 1 criterion but has no known "fatal flaw." "Poor" studies have at least 1 fatal flaw.

Systematic Reviews

Criteria:

X Comprehensiveness of sources considered/search strategy used

X Standard appraisal of included studies

X Validity of conclusions

X Recency and relevance are especially important for systematic reviews

Definition of ratings from above criteria:

Good: Recent, relevant review with comprehensive sources and search strategies; explicit and relevant selection criteria; standard appraisal of included studies; and valid conclusions.

Fair: Recent, relevant review that is not clearly biased but lacks comprehensive sources and search strategies.

Poor: Outdated, irrelevant, or biased review without systematic search for studies, explicit selection criteria, or standard appraisal of studies.

Case-Control Studies

Criteria:

X Accurate ascertainment of cases

X Nonbiased selection of cases/controls with exclusion criteria applied equally to both

X Response rate

X Diagnostic testing procedures applied equally to each group

X Measurement of exposure accurate and applied equally to each group

X Appropriate attention to potential confounding variables

Definition of ratings based on criteria above:

Good: Appropriate ascertainment of cases and nonbiased selection of case and control participants; exclusion criteria applied equally to cases and controls; response rate equal to or greater than 80%; diagnostic procedures and measurements accurate and applied equally to cases and controls; and appropriate attention to confounding variables.

Fair: Recent, relevant, without major apparent selection or diagnostic work-up bias but with response rate less than 80 percent or attention to some but not all important confounding variables.

Poor: Major selection or diagnostic work-up biases, response rates less than 50 percent, or inattention to confounding variables.

Randomized Controlled Trials and Cohort Studies

Criteria:

X Initial assembly of comparable groups

for RCTs: adequate randomization, including first concealment and whether potential confounders were distributed equally among groups

for cohort studies: consideration of potential confounders with either restriction or measurement for adjustment in the analysis; consideration of inception cohorts

X Maintenance of comparable groups (includes attrition, cross-overs, adherence, contamination)

X Important differential loss to follow-up or overall high loss to follow-up

X Measurements: equal, reliable, and valid (includes masking of outcome assessment)

X Clear definition of interventions

X All important outcomes considered

X Analysis: adjustment for potential confounders for cohort studies, or intention to treat analysis for RCTs.

Definition of ratings based on above criteria:

Good: Meets all criteria: Comparable groups are assembled initially and maintained throughout the study (follow-up at least 80 percent); reliable and valid measurement instruments are used and applied equally to the groups; interventions are spelled out clearly; all important outcomes are considered; and appropriate attention to confounders in analysis. In addition, for RCTs, intention to treat analysis is used.

Fair: Studies will be graded "fair" if any or all of the following problems occur, without the fatal flaws noted in the "poor" category below: Generally, comparable groups are assembled initially but some question remains whether some (although not major) differences occurred with follow-up; measurement instruments are acceptable (although not the best) and generally applied equally; some but not all important outcomes are considered; and some but not all potential confounders are accounted for. Intention to treat analysis is done for RCTs.

Poor: Studies will be graded "poor" if any of the following fatal flaws exists: Groups assembled initially are not close to being comparable or maintained throughout the study; unreliable or invalid measurement instruments are used or not applied at all equally among groups (including not masking outcome assessment); and key confounders are given little or no attention. For RCTs, intention to treat analysis is lacking.

Diagnostic Accuracy Studies

Criteria:

X Screening test relevant, available for primary care, adequately described

X Study uses a credible reference standard, performed regardless of test results

X Reference standard interpreted independently of screening test

X Handles indeterminate results in a reasonable manner

X Spectrum of patients included in study

X Sample size

X Administration of reliable screening test

Definition of ratings based on above criteria:

Good: Evaluates relevant available screening test; uses a credible reference standard; interprets reference standard independently of screening test; reliability of test assessed; has few or handles indeterminate results in a reasonable manner; includes large number (more than 100) broad-spectrum patients with and without disease.

Fair: Evaluates relevant available screening test; uses reasonable although not best standard; interprets reference standard independent of screening test; moderate sample size (50 to 100 subjects) and a "medium" spectrum of patients.

Poor: Has fatal flaw such as: Uses inappropriate reference standard; screening test improperly administered; biased ascertainment of reference standard; very small sample size or very narrow selected spectrum of patients.

Criteria for Grading Linkages in the Analytic Framework

Introduction

As noted in the previous document in this Appendix, the Methods Work Group for the U.S. Preventive Services Task Force (USPSTF) developed a set of criteria by which the quality of individual studies could be evaluated in terms of both internal validity. The Methods Work Group also developed definitions and criteria for judging the strength or quality of evidence for key questions—ie, linkages in the analytic frameworks—for the topics of systematic evidence reviews. These quality criteria were discussed at the May 1999 quarterly meeting and were essentially adopted for use by the Evidence-based Practice Centers in developing their first set of systematic evidence reviews. This document describes the criteria relating specifically to linkages in the analytic framework.[1]

[1] **The USPSTF is developing a separate set of criteria for rating its recommendations about an entire preventive service, including policies for appropriate extrapolation to populations or settings not reflected in the reviewed literature, but because the SERs do not contain USPSTF recommendations, those ways of grading recommendations are not dealt with here.**

Linkage Category Definitions

The rating scheme for grading the evidence for a linkage in the analytic framework rests on 3 classes of criteria: aggregate internal validity, aggregate external validity, and consistency or coherence. The Methods Work Group did not establish set formulae for arriving at any linkage score for these criteria sets. As with the criteria for quality of individual articles, they are intended to be applied as general guidelines, and the judgments are made implicitly. Judgments can be made about evidence of benefits and evidence of harms. In addition, a summative grade—ie, an overall rating—combining the evaluations of the 3 categories defined below can be given.

Also, as with the criteria for individual studies, these 3 categories can be labeled as "good," "fair," or "poor." That is, the linkages can be understood to be supported by good evidence, fair evidence, or poor evidence. The summative, overall rating can also range from good to poor.

Aggregate Internal Validity:

This category refers to the overall extent to which data are valid for conditions addressed within studies. It would be rated according to quality grading information about individual studies.

Aggregate External Validity:

This category concerns the generalizability of evidence to questions addressed by the linkage. This would include the concordance between populations, interventions and outcomes in the studies reviewed and those to which the linkage pertains. In short, this category reflects the applicability of the evidence to real-world conditions.

It is expected that differences between conditions examined in studies and those addressed by the linkages should be considered if they could potentially influence outcomes. These might include (but not necessarily be limited to): (a) biologic or pathologic characteristics; (b) incidence and prevalence of clinical conditions; (c) distribution of comorbid conditions that might affect outcomes; and (d) likelihood of acceptability and adherence on the part of patients or providers (or both).

Consistency:

This category relates to the overall "coherence" of the body of evidence relating to the linkage. Specifically, it includes the number of studies, the homogeneity of those studies (in terms of clinical conditions, populations, settings, and the like), the level of precision of findings in the studies, and the direction of results. In addition, it can include dose-response relationships.

APPENDIX D

Evidence Tables and Specialized Glossary

Glossary of Evidence Table Abbreviations Screening for Depression

Abbreviation	Term
AGECAT	Computerized diagnostic screening test
AMI	Amitriptyline
ARR	Absolute risk reduction
BDI	Beck Depression Inventory
CAGE	CAGE questionnaire
CAMDEX	Cambridge Examination for Mental Disorders of the Elderly
CBT	Cognitive behavior therapy
CES-D	Centers for Epidemiologic Studies Depression scale
CGI	Clinical Global Impressions
Cntrl	Control
COG	Cognitive therapy
CT	Cognitive treatment
DI	Diagnostic interview
DIS	Diagnostic interview schedule
DK	Don't know
DSM-IIIR	Diagnostic and Statistical Manual, 3rd Edition (Revised)
DUSOI	Duke University Severity of Illness Scale
dx	Diagnosis
EPDS	Edinburgh Postnatal Depression Scale
FP	Family practice
GAS	Global Assessment Score
GDS	Geriatric Depression Scale
GDS-S	Short form Geriatric Depression Scale
GHQ	General Health Questionnaire
GMSS	Geriatric Mental Status Schedule
GP	General Practice Patients
HAM-D	Hamilton Depression Rating Scale
HS	High school
HSCL-20	Hopkins Symptomatic Checklist-20
HV	Health Visitor Counseling
ICD 10	International Classification of Disease - 10th Revision
IDS	Inventory of Depressive Symptomatology
IPT	Interpersonal psychotherapy
ITT	Intention-to-treat
MADS	Montgomery-Asberg Depression Scale
MD	Medical Doctor (when under gold standard used)
MD-UC	Depression program for major depression
MD_UC	Usual care for major depression
mD-CT	Depression program for minor depression
mD-UC	Usual care for minor depression
MMSE	Mini-Mental State Exam
MOS	Medical Outcomes Study
NA	Not available
NNT	Numbers needed to treat
NOR	Nortriptyline
NOS	Not otherwise specified
NR	Not reported
NS	Not significant
OSI	Other Structured Interview
OTH	Other

PC	Private clinic
PD	Physician diagnosis
PHQ	Patient health questionnaire
Pl	Placebo
PROQSY	Self-administered computerized assessment
PS	Problem solving
PSE	Present State Exam
PS-GP	Problem solving by general practitioner
PS-N	Problem solving by nurse
QDIS	Quick Diagnostic Interview Schedule
RDC	Research Diagnostic Criteria
RRR	Relative risk reduction
SADS-L	Schedule for Affective Disorders and Schizophrenia, Lifetime Version
SCI	Structured clinical interview
SCID	Structure clinical interview for DSM-III-R (or –IV)
SD	Standard Deviation
SDS	Zung Self-Rating Depression Scale
SDDS	Symptom Drive Diagnostic System
SF-36	Short Form 36
SI	Structured Interview
SIP	Sickness Impact Profile
SSD	Subsyndromal Symptom Depression
SW	Social worker
TC	Telephone counseling
UC	Usual Care
VA	Veterans Administration / Department of Veterans Affairs

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies

Author, Year	Screening Setting	Age Range (years)	Other Inclusion Criteria	Conditions of Interest	Exclusions
Williams et al, 1995[17]	Primary Care: 3 different clinics	>30		Major Depession, Sub-syndromal depression	Current substance abuse, major psychiatric illness, known depression or anti-depressants, chronic pain, dementia
Arthur et al, 1999[34]	Primary care	>75		Depression	
Banerjee et al, 1998[35]	Community	>65	Home care patients from Lewishon East in UK	Depression	Current psychiatric care
Bashir et al, 1996[36]	Primary care	18-74	Attending an appointment at a study practice	Depression	
Beekman et al, 1997[37]	Community	55-85		Major Depression	

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Definition of Groups	Number of Subjects	Mean Age (years)	% Female	Ethnicity	Education
All	221	60 (SD=12.7)	32%	38% White 9% Black 53% Hispanic	9.5 mean years of education
All	201	79	57%	NR	NR
All	214	NR	NR	NR	NR
Grp 1: All Grp 2: Men Grp 3: Women	129	34	62%	NR	NR
All	487	NR	58%	NR	55% HS grads

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Author, Year	Test(s) Defined with Cut-Offs: (1): Test 1 (2): Test 2 (3): Test 3 (4): Test 4	Primary Screening Test Evaluated	Gold Standard Used	Sensitivity	Specificity	LR+/LR-
Williams et al, 1995[17]	(1): SDS for MDD: >.043 (2): SDS for SSD>.043	SDS	DI: HAM-D, DIS: OSI: SCID, OTH: SF-36	Test 1: 100% Test 2: 66%	Test 1: 72% Test 2: 79%	Test 1: 3.6/0 Test 2: 3.1/0.4
Arthur et al, 1999[34]	(1):GDS 15: ≥3	GDS 15	968	100%	72%	3.6 / NA
Banerjee et al, 1998[35]	(1): Self-care (D): ≥8	Self-care (D)	GMS/ AGECAT	90%	53%	1.9/0.19
Bashir et al, 1996[36]	(1): GHQ: >3 (all) (2): GHQ: >3 (men) (3): GHQ: >3 (women)	GHQ	PROQSY (Computerized self-administered diagnostic review)	Test 1: 76% Test 2: 75% Test 3: 77%	Test 1: 74% Test 2: 68% Test 3: 78%	Test 1: 2.9/0.32 Test 2: 2.3/0.37 Test 3: 3.5/0.29
Beekman et al, 1997[37]	(1): CES-D: ≥16 (2): CES-D: ≥18 (3): CES-D: ≥20	CES-D	DIS	Test 1: 100% Test 2: 93.5% Test 3: 93.5%	Test 1: 53.3% Test 2: 65.6% Test 3: 73.5%	Test 1: 2.12/0 Test 2: 2.7/0.10 Test 3: 3.5/0.09

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Adequate Inclusion Criteria?	Appropriate Spectrum of Patients?	Gold Standard Applied to All Subjects?	Masked Assessors?	Aggregate Internal Validity	External Validity	Overall Quality
Yes	Yes	Yes	NR	Good	Good	Good
Yes	Yes	Yes	Yes	Good	Good	Good
Yes	Yes	Yes	DK	Good	Good	Good
Yes	Yes	Yes	No	Good	Good	Good
Yes	Yes	Yes	DK	Good	Good	Good

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies

Author, Year	Screening Setting	Age Range (years)	Other Inclusion Criteria	Conditions of Interest	Exclusions
Bird et al, 1987[38]	Primary care	>65	NR	Significantly depressed state	NR
Broadhead et al, 1995[39]	Primary care	18-70	Able to read and write in English and scheduled to have a physician visit	Major Depression	
Burnam et al, 1988[40]	Primary care	NR		Major Depression, Depression	
D'Ath et al, 1994[41]	Primary care	65-90		Depression	Substance Abuse: DK, Suicidality: DK, Pregnancy: NA
Fechner-Bates et al, 1994[42]	Primary Care	17-80		Major Depression	NR

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Definition of Groups	Number of Subjects	Mean Age (years)	% Female	Ethnicity	Education
Grp 1: Depressed Grp 2: Non-depressed	75	Grp 1: 72.4 (SD=4.7) Grp 2: 73.3 (SD=4.7)	Grp 1: 68% Grp 2: 65%	NR	NR
All	388	39.4 (SD=12.4)	73%	98% White 0.3% Black 1.7% Other	NR
Primary care sample	1,450	43	60%	51% White 39% Hispanic	NR
Grp 1: Validation	198	74.1	68%	NR	NR
Grp 1: All	1,928 screened, 425 in study	39.6	76.70%	93% White	62.4% HS grads

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Author, Year	Test(s) Defined with Cut-Offs: (1): Test 1 (2): Test 2 (3): Test 3 (4): Test 4	Primary Screening Test Evaluated	Gold Standard Used	Sensitivity	Specificity	LR+/LR-
Bird et al, 1987[38]	(1): Self-care (D):≥5 (2): Self-care (D):≥6 (3): Self-care (D):≥7 (4): Zung: ≥38	Self-care (D)	MD	Test 1: 77% Test 2: 77% Test 3: 61% Test 4: ~87%	Test 1: 91% Test 2: 98% Test 3: 98% Test 4: ~76%	Test 1: 8.5/0.25 Test 2: 38.5/0.23 Test 3: 30.5/0.40 Test 4: 3.6/0.17
Broadhead et al, 1995[39]	(1): SDDS-PC: NR	SDDS-PC	SCID-P	90.40%	77.20%	4.0/0.12
Burnam et al, 1988[40]	(1): MOS screener: 0.060	MOS screener	DIS	86%	95%	17.2 / 0.15
D'Ath et al, 1994[41]	(1):GDS 15: ≥5/≥6 (2):GDS 10: ≥3/≥4 (3):GDS 4: ≥2/≥1 (4):GDS 1:	GDS	GMS	Test 1: 91%/ 78% Test 2: 93%/ 89% Test 3: 61%/ 93% Test 4: 59%	Test 1: 72%/ 82% Test 2: 63%/ 77% Test 3: 88%/ 63% Test 4: 75%	Test 1: 3.25 /0.125
Fechner-Bates et al, 1994[42]	(1): CES-D ≥16: for major depression (2): CES-D ≥16: depression broadly defined (3): CES-D ≥16: for dysthymia	CES-D	SCID within 14 days of CES-D	Test 1: 79.5% Test 2: 72% Test 3: 100% (9/9)	Test 1: 71.1% Test 2: 75.6% Test 3: NR	Test 1: 2.7 / 0.30

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Adequate Inclusion Criteria?	Appropriate Spectrum of Patients?	Gold Standard Applied to All Subjects?	Masked Assessors?	Aggregate Internal Validity	External Validity	Overall Quality
Yes	Yes	Yes	DK	Fair	Fair	Good
Yes	No	Yes, All patients who consented had gold standard exam, but only 41% of those screened had gold standard.	Yes	Fair	Good	Fair
Yes	Yes	Yes	Yes	Good	Good	Good
Yes	Yes	Yes	Yes	Good	Good	Good
Yes	No	Yes, Gold standard administered only to group selected to have confirmatory testing	NR	Fair	Good	Fair

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies

Author, Year	Screening Setting	Age Range (years)	Other Inclusion Criteria	Conditions of Interest	Exclusions
Finlay-Jones et al, 1979[43]	Primary care	Grp 1: 18-40 Grp 2: 18-65	Women only	Major Depression	
Geisser et al, 1997[44]	Pain clinic	NR	Duration of pain >3 months	Major Depression	
Gerety et al, 1994[45]	Nursing home	>60		Major Depression	English-speaking, >50% on MMSE (pro-rated)
Goldberg et al, 1970[46]	Primary care	NR		Major Depression, Depression, any depressive symptoms	
Hendrie et al, 1995[47]	Primary care	>60 yr		Major Depression, Depression Minor depression	Prisoners, non-English speaking, hearing-impaired
Holcomb et al, 1996[48]	OB clinic	NR	Women already attending OB-GYN clinics	Current, remittent, no depression	

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Definition of Groups	Number of Subjects	Mean Age (years)	% Female	Ethnicity	Education
Grp 1: General practice patients Grp 2: Medically ill patients	197	Grp 1: 28 (SD=5) Grp 2: 42 (SD=14)	Grp 1: 100% Grp 2: 100%	NR	NR
All	132	40.7 (SD=11.2)	71%	92% White 5% Black 2% Hispanic	82% HS grads
All	123	78.9 (SD=9.6)	56%	74% White NR Black 15% Hispanic	NR
All	200	NR	64%	NR	NR
All	125	68	68.80%	NR White 63.4% Black NR Other	NR
All women	105	24.2 (SD=6.3)	100%	33% White 63% Black 4% Other	66% HS grads

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Author, Year	Test(s) Defined with Cut-Offs: (1): Test 1 (2): Test 2 (3): Test 3 (4): Test 4	Primary Screening Test Evaluated	Gold Standard Used	Sensitivity	Specificity	LR+/LR-
Finlay-Jones et al, 1979[43]	(1): GHQ: ≥5	GHQ	PSE	Grp 1: 75% (GP patients) Grp 2: 76% (medically ill patients)	Grp 1: 73% (GP patients) Grp 2: 71% (medically ill patients)	Grp 1: 2.8/0.34 (GP patients) Grp 2: 2.6/0.34 (medically ill patients)
Geisser et al, 1997[44]	(1): CES-D: ≥27 (2): BDI ≥16	CES-D, BDI	Interview by trained clinical psychologist using DSM-IV criteria	Test 1: 81.8% Test 2: 68.2%	Test 1: 72.7% Test 2: 78.4%	Test 1: 3.0/0.25 Test 2: 3.1/0.41
Gerety et al, 1994[45]	(1): GDS: >11 (2): GDS-S >6 (3): Brief Carroll Dep. Rating Scale: >5 (4):CES-D: >16	GDS CES-D Other	SCID	Test 1: 89% Test 2: 88% Test 3: 85% Test 4: 74%	Test 1: 68% Test 2: 62% Test 3: 77% Test 4: 70%	Test 1: 2.8/0.16 Test 2: 2.3/0.19 Test 3: 3.7/0.19 Test 4: 2.5
Goldberg et al, 1970[46]	(1): GHQ: ≥12	GHQ	MD	95.80%	87.80%	7.9/0.05
Hendrie et al, 1995[47]	(1): CES-D ≥16 (2): CAMDEX (inclusive method) (3): CAMDEX (substitution)	CES-D	DSM-IIIR, SI: CAMDEX inclusive and substitutive methods	Test 1: 80% Test 2: 65% Test 3: 72%	Test 1: 91% Test 2: 87% Test 3: 86%	Test 1: 8.9 / 0.22
Holcomb et al, 1996[48]	(1): BDI: ≥16	BDI	DIS	83% (.56, .98)	89% (.81, .95)	Test 1: 7.75/0.19

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Adequate Inclusion Criteria?	Appropriate Spectrum of Patients?	Gold Standard Applied to All Subjects?	Masked Assessors?	Aggregate Internal Validity	External Validity	Overall Quality
Yes	Yes	Yes	No, team assessment of outcome, with one member aware of screening status	Fair	Good	Fair
Yes	Yes	Yes	DK	Good	Fair	Fair
Yes	Yes	Yes	NR	Good	Good	Good
Yes	No	Yes	Yes	Fair	Good	Fair
Yes	Yes	No, only 45% of patients with CES-D >16 completed gold standard, but 81 subjects with CES-D <16 also completed the gold standard	Yes	Fair	Good	Fair
Yes	Yes	Yes	Yes	Good	Good	Good

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies

Author, Year	Screening Setting	Age Range (years)	Other Inclusion Criteria	Conditions of Interest	Exclusions
Hoyl et al, 1999[49]	Geriatric clinic at VA hospital	65-87	Attending geriatric assessment clinic	Major Depression, Depressive NOS	
Irwin et al, 1999[50]	Community	NR		Major Depression	Physical disorder leading to secondary depression
Klinkman et al, 1997[51]	Primary care	NR		Major Depression	
Leon et al, 1996[52]	Primary care	18-70	English speaking/writing	Major Depression	
Leung et al, 1998[53]	Primary care	NR	Taiwan Primary care patients with chronic medical problems and no known previous history of depression	Any depressive symptom	

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Definition of Groups	Number of Subjects	Mean Age (years)	% Female	Ethnicity	Education
All	74	74.6	1%	NR	75.7% HS grads
Grp 1: Patients from mental health research setting Grp 2: Community dwelling older adults from primary care practices	Grp 1: 83 Grp 2: 68	Grp 1: 44.9 (SD=10.3) int. 40.0 (SD=12.8) control Grp 2: 72.0 (SD=7.0)	Grp 1: 50% int./ 42% control Grp 2: 48%	Grp 1: NR Grp 2: NR	Grp 1: NR Grp 2: NR
All	425	39.6	77%	93%	62.4% HS grads
All	501	49.4 (SD=12.8)	66%	35% White 46% Black 9% Hispanic 9% Asian	NR
All	268	59	60%	100% Asian	31% HS grads

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Author, Year	Test(s) Defined with Cut-Offs: (1): Test 1 (2): Test 2 (3): Test 3 (4): Test 4	Primary Screening Test Evaluated	Gold Standard Used	Sensitivity	Specificity	LR+/LR-
Hoyl et al, 1999[49]	(1): GDS 15: ≥5 (2): GDS 5: ≥2 (3): Single item: ≥1	GDS 15 and GDS 5 and single item	Prime MD mood module, MD	Test 1: 94% Test 2: 97% Test 3: 85%	Test 1: 82% Test 2: 85% Test 3: 65%	Test 1: 5.2/0.07 Test 2: 6.5/0.04 Test 3: 2.4/0.23
Irwin et al, 1999[50]	(1):CES-D: ≥4	CES-D	DSM-IV, SCID	Test 1: 99% Test 2: 100%	Test 1: 84% Test 2: 92%	Test 1: 6.2 / 0.01 Test 2: N/A
Klinkman et al, 1997[51]	(1): CES-D ≥16 (2): CES-D ≥22 (3): MD ID (4): MD ID and CES-D ≥16	CES-D	OSI: SCID	Test 1: 81% Test 2: 61% Test 3: 35% Test 4: 31%	Test 1: 72% Test 2: 85% Test 3: 93% Test 4: 95%	Test 1: 2.9/.26 Test 2: 4.0/.46 Test 3: 4.9/.70 Test 4: 5.7/.73
Leon et al, 1996[52]	SDDS-PC: NA	SDDS-PC	Specific diagnosis manuals	71%	87%	5.5/0.33
Leung et al, 1998[53]	(1): Zung SDS: ≥50 (2): Zung SDS: ≥55 (3): Zung SDS: ≥60	Zung SDS	MD: Following DSM-IV criteria	Test 1: 100% Test 2: 66.7% Test 3: 44.4%	Test 1: 70.7% Test 2: 90.2% Test 3: 90.2%	6.8/0.37

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Adequate Inclusion Criteria?	Appropriate Spectrum of Patients?	Gold Standard Applied to All Subjects?	Masked Assessors?	Aggregate Internal Validity	External Validity	Overall Quality
Yes	Yes	Yes, The gold standard was the PRIME-MD based assessment (unblinded)	No	Fair	Good	Fair
Yes	No	Yes	Yes	Fair	Fair	Fair
DK	No	Yes	DK	Fair	Good	Fair
Yes	Yes	Yes	Yes	Good	Good	Good
Yes	Yes	Yes	Yes	Good	Good	Good

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies

Author, Year	Screening Setting	Age Range (years)	Other Inclusion Criteria	Conditions of Interest	Exclusions
Lewinsohn et al, 1997[54]	Community	>50	Licensed drivers	Major Depression, Depression	
Lustman et al, 1997[55]	Special diabetes clinic with patients recruited from community	Blank	Patients with poorly controlled diabetes	Major Depression	
Lyness et al, 1997[56]	Primary care	>60		Major Depression	
Myers and Weissman, 1980[57]	Community clinic	≥18		Major Depression OTH: Minor depression	
Nagel et al, 1998[58]	Primary care	≥18		Major Depression; Dysthymia	

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Definition of Groups	Number of Subjects	Mean Age (years)	% Female	Ethnicity	Education
All	1,005	63.9 (SD=7.9)	58%	Blank	94% HS grads
Grp 1: All Grp 2: Depressed Grp 3: Non-depressed	172	Grp 1: 48 (SD=13.6) (all) Grp 2: 43.2 (SD=13.0) (depressed) Grp 3: 50.7 (SD=13.1) (non-depressed)	Grp 1: 48% Grp 2: 56% Grp 3: 43%	Grp 1: 84% White/ 15% Black/ 1% Other Grp 2: 89% White/ 11% Black Grp 3: 82% White/ 16% Black/ 2% Other	Grp 1: 13.7±2.6 years Grp 2: 12.9±2.0 years Grp 3: 14.1±2.7 years
All	130	71.0 (SD=6.8)	58.50%	97.7% White 2.3% Black	13.7 mean years
All	515	NR	57%	89.6% White 10.4% Black	NR
Grp 1: Training center Grp 2: Community	Grp 1: 566 Grp 2: 457	Grp 1: 47.2 Grp 2: 46.1	Grp 1: 66% Grp 2: 70%	Grp 1: 78.6% White 19.2% Black Grp 2: 93.4% White 4.6% Black	Grp 1: 89% HS graduates Grp 2: 87% HS graduates

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Author, Year	Test(s) Defined with Cut-Offs: (1): Test 1 (2): Test 2 (3): Test 3 (4): Test 4	Primary Screening Test Evaluated	Gold Standard Used	Sensitivity	Specificity	LR+/LR-
Lewinsohn et al, 1997[54]	(1): CES-D: ≥12 (2): brief CES-D: ≥4	CES-D, brief CES-D	DI: RDC, DSM-IIIR OTH: SADS-L	Test 1: 76% Test 2: 80%	Test 1: 77% Test 2: 80%	Test 1: 3.3/.31 Test 2: 4/.25
Lustman et al, 1997[55]	(1): BDI: ≥13	BDI	DIS	85%	88%	Test 1: 7.1/0.17
Lyness et al, 1997[56]	(1): CES-D: ≥21 (2): GDS: ≥10 (3): GDS-S: ≥5	CES-D, GDS, GDS-S	DI: SCID	Test 1: 92/40 (major/minor) Test 2:100/70 Test 3: 92/80	Test 1: 87/82 (major/minor) Test 2: 84/80 Test 3: 81/78	Test 1: 7.1 / 0.09
Myers and Weissman, 1980[57]	(1): CES-D: ≥16 (2): CES-D: ≥17 (3): CES-D: ≥21 (4): CES-D: ≥16 (minor depression)	CES-D for MDD CES-D for minor depression	DI: SADS-RDC	Test 1: 63.6% Test 2: 63.6% Test 3: 54.5% Test 4: 23%	Test 1: 93.9% Test 2: 94.4% Test 3: 95.9% Test 4: 92%	10.4/0.39
Nagel et al, 1998[58]	(1): MOS (2): MOS	MOS	DIS	Test 1: 88% Test 2: 100%	Test 1: 72% Test 2: 77%	Test 1: 3.1/0.17 Test 2: 4.3/0

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Adequate Inclusion Criteria?	Appropriate Spectrum of Patients?	Gold Standard Applied to All Subjects?	Masked Assessors?	Aggregate Internal Validity	External Validity	Overall Quality
Yes	No	Yes, analyses based on 1,005/1,554 who completed the diagnostic interview	DK	Fair	Good	Fair
Yes	Yes, sample drawn from a population of trial enrolees	Yes	Yes	Good	Fair	Fair
Yes	No	No	DK	Fair	Good	Fair
Yes	Yes	Yes	DK	Good	Good	Good
Yes	Yes	Yes	Yes	Good	Good	Good

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies

Author, Year	Screening Setting	Age Range (years)	Other Inclusion Criteria	Conditions of Interest	Exclusions
Neal and Baldwin, 1994[59]	Geriatric outpatient clinic	>65	Consecutive new patients	Depression	Severe cognitive impairment, unable to give consent
Okimoto et al, 1982[60]	VA Primary care	>60	VA patients	NR	NR
Parkerson and Broadhead, 1997[61]	Primary care	18-64	Clinic attendees	Major Depression & Dysthymia	NR
Salokangas et al, 1995[62]	Primary care	18-64		Severe depression	
Schulberg et al, 1985[63]	Primary care	NR	New patients at 3 primary care practices and 3 community MH sites	Multiple categories	
Spitzer et al, 1999[64]	Primary care	>18		Major Depression OTH: Any mood disorder	
Steer et al, 1999[65]	Primary care	NR	NR	Major Depression	

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Definition of Groups	Number of Subjects	Mean Age (years)	% Female	Ethnicity	Education
NR	45	77	62%	NR	NR
NR	55	Grp 1: 69.4 (SD=8.1) Grp 2: same	Grp 1: 2% Grp 2: same	Grp 1: NR	Grp 1: NR
Duke AD	481	36.9 (SD=10.9)	72%	73% White NR Black NR Other	93% HS grads
All subjects	436	NR	NR	NR	NR
Primary care patients	294	NR	NR	NR	NR
All	585	46 (SD=17.2)	66%	79% White 13% Black 4% Hispanic	25% College graduates
All patients	120	58.4 (SD=15.5)	50%	86% White 9% Black 2.5% Hispanic 2.5% Asian	NR

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Author, Year	Test(s) Defined with Cut-Offs: (1): Test 1 (2): Test 2 (3): Test 3 (4): Test 4	Primary Screening Test Evaluated	Gold Standard Used	Sensitivity	Specificity	LR+/LR-
Neal and Baldwin, 1994[59]	(1): GDS >11 (2): GDS >4 (3): MD detection (4): short GDS >5	GDS	SI: GMSS	Test 1: 83% Test 2: 67% Test 3: 40% Test 4: 71%	Test 1: 80% Test 2: 80% Test 3: Blank Test 4: 80%	Test 1: 4.15/.21 Test 2: 3.35/.41 Test 3: Blank Test 4: 3.55/.36
Okimoto et al, 1982[60]	(1): Zung: 60 (2): Popoff: not given	SDS	Physician diagnosis based on DSM-IIIR	Test 1: 76% Test 2: 88%	Test 1: 82% Test 2: 61%	Test 1: 4.2/.29 Test 2: 2.3/.20
Parkerson and Broadhead, 1997[61]	(1): Duke-AD >30	Duke AD	DIS (DSM-IIIR based)	Test 1: 81%	Test 1: 64%	Test 1: 2.25/.30
Salokangas et al, 1995[62]	(1): DEPS: ≥9	DEPS	PSE (Present state exam)	74%	85%	4.9 / 0.31
Schulberg et al, 1985[63]	(1): CES-D ≥16 (2): CES-D ≥27 (3): MD dx	CES-D	DIS	Test 1: 96.3% Test 2: 89% Test 3: 26%	Test 1: 38.6% Test 2: 70% Test 3: 98%	Test 1: 1.6/.10 Test 2: 2.97/.16 Test 3: 13/.75
Spitzer et al, 1999[64]	(1): PHQ: NR (2): PHQ: NR	PHQ	Mental health professional telephone interview	Test 1: 73% Test 2: 61%	Test 1: 98% Test 2: 94%	Test 1: 36/0.28 Test 2: 10.2/0.41
Steer et al, 1999[65]	(1): BDI-PC: ≥4	BDI-PC	Mood module of prime-MD	97% (82-99)	99% (94-99)	Test 1: 97 / 0.03

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Adequate Inclusion Criteria?	Appropriate Spectrum of Patients?	Gold Standard Applied to All Subjects?	Masked Assessors?	Aggregate Internal Validity	External Validity	Overall Quality
Yes	Yes	Yes	DK	Good	Good	Good
DK	Yes	No	Yes	Fair	Good	Fair
Yes	Yes	Yes	NR	Good	Good	Good
Yes	Yes	Yes	NR	Good	Good	Good
Yes	No	Yes, only those who agreed to participate in the diagnostic interview are included in the final sample	Yes	Fair	Good	Fair
Yes	Yes	Yes	Yes	Good	Good	Good
Yes	Yes	Yes, mood module of PRIME-MD used as criteria standard	Yes	Fair	Good	Fair

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies

Author, Year	Screening Setting	Age Range (years)	Other Inclusion Criteria	Conditions of Interest	Exclusions
Turk and Okifuji, 1994[66]	Pain clinic	NR		Major Depression; Dysthymia	
Upadhyaya and Stanley, 1997[67]	Primary care	65-90		Depression	NR
van Marwijk et al, 1995[68]	Primary care	65-94	Consecutive patients invited to participate	Major Depression; Dysthymia	MD assessment of fitness to participate
Von Korff et al, 1987[69]	Primary care	>18	Visit between 12/81-3/21/82	Major Depression OTH: Depression and anxiety	Living far from clinic; participation in another study
Weyerer et al, 1999[70]	Old age home and a psychogeriatric clinic	NR		OTH: Depression	
Whooley et al, 1997[71]	VA urgent care clinic	21-89	Consecutive patients attending urgent care clinic	Major Depression	Bipolar Disorder, too delusional/ intoxicated, schizophrenia

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Definition of Groups	Number of Subjects	Mean Age (years)	% Female	Ethnicity	Education
Grp 1: Depressed Grp 2: Non-depressed	Grp 1: 50 Grp 2: 50	Grp 1: 40.2 Grp 2: 43.7	Grp 1: 70% Grp 2: 62%	Grp 1: 88% White Grp 2: 90% White	Grp 1: 88% HS graduates Grp 2: 88% HS graduates
Patients attending health center	72	71 (SD=4.6)	49%	NR	NR
FP patients	586	73.5	60%	NR	NR
All	809	NR	68%	16.6% White 83.4% Non-White	23.7% HS grads
All	71	NR	NR	NR	NR
All	590 to 536	53 (SD=14.0)	3%	55% White 29% Black 8% Hispanic 6% Asian 2% Other	86% HS grads

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Author, Year	Test(s) Defined with Cut-Offs: (1): Test 1 (2): Test 2 (3): Test 3 (4): Test 4	Primary Screening Test Evaluated	Gold Standard Used	Sensitivity	Specificity	LR+/LR-
Turk and Okifuji, 1994[66]	CES-D: ≥16	CES-D	Psychologist using DSM-IIIR	86%	50%	1.72/0.28
Upadhyaya and Stanley, 1997[67]	(1): HAD-D: ≥9 (2): Self-care: ≥6	Other	GMS-AGECAT	Test 1: 70% Test 2: 95%	Test 1: 87% Test 2: 86%	Test 1: 5.38/.34 Test 2: 6.8/.06
van Marwijk et al, 1995[68]	(1):GDS 30: <11/≥11 (2):GDS 15: <3/≥3 (3):GDS 10: <3/≥3 (4):GDS 4: <2/≥2	GDS	DIS	Test 1: 55% Test 2: 67% Test 3: 52% Test 4: 67%	Test 1: 86% Test 2: 73% Test 3: 83% Test 4: 66%	Test 1: 3.9/0.52 Test 2: 2.48/1.22 Test 3: 3.06/0.7 Test 4: 1.97/0.5
Von Korff et al, 1987[69]	(1): GHQ: ≥5 (2): MD assessment of "mental disorder"	GHQ	DIS	Test 1: 78.5% Test 2: 74.8%	Test 1: 66% Test 2: 72%	Test 1: 2.3/.33 Test 2: 2.7/.35
Weyerer et al, 1999[70]	(1): EBAS DEP: >2 (2): EBAS DEP: >3 (3): EBAS DEP: <4	EBAS	Physician or psychiatric nurse diagnosis	Test 1: 100% Test 2: 100% Test 3: 93.3%	Test 1: 47% Test 2: 67.7% Test 3: 85.3%	Test 1: 2.1/0 Test 2: 3.0/0 Test 3: 6.2/0.08
Whooley et al, 1997[71]	(1): New 2-item (2): CES-D:>16 (3): Short CES-D: 10 (4): BDI: 10 (5): Short BDI: 5 (6): MOS: 0.060 (7): SDDS-PC: 2	CES-D BDI MOS Other	DI: QDIS DSM-IIIR	Test 1: 96% Test 2: 93% Test 3: 90% Test 4: 89% Test 5: 92% Test 6: 93% Test 7: 96%	Test 1: 57% Test 2: 69% Test 3: 72% Test 4: 64% Test 5: 61% Test 6: 72% Test 7: 51%	Test 1: 2.2/0.07 Test 2: 3/0.10 Test 3: 3.2/0.14 Test 4: 2.5/0.17 Test 5: 2.4/0.13

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Adequate Inclusion Criteria?	Appropriate Spectrum of Patients?	Gold Standard Applied to All Subjects?	Masked Assessors?	Aggregate Internal Validity	External Validity	Overall Quality
Yes	No	Yes	Yes	Fair	Fair	Fair
Yes	Yes	Yes	Yes	Good	Good	Good
Yes	Yes	Yes	NR	Good	Good	Good
Yes	Yes	Yes	DK	Good	Good	Good
No	No	Yes	Yes	Fair	Fair	Fair
Yes	Yes	Yes	NR	Good	Good	Good

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies

Author, Year	Screening Setting	Age Range (years)	Other Inclusion Criteria	Conditions of Interest	Exclusions
Wickberg and Hwang, 1996[72]	Post-partum women	≥18	Swedish-speaking women	Major Depression	Already in treatment
Zich et al, 1990[73]	Primary Care	≥18	Random sample of pateints from general internal medicine clinic	Major Depression	DK

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Definition of Groups	Number of Subjects	Mean Age (years)	% Female	Ethnicity	Education
All	128	NR	100%	NR	NR
Grp 1: CES-D (34) Grp 2: BDI (31) Grp 3: All screening participants (475)	475	All: 57	All: 63%	All: 50% White 20% Black 15% Hispanic 12% Asian 3% Other	Grp 1: NR

Appendix D. Evidence Tables

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Author, Year	Test(s) Defined with Cut-Offs: (1): Test 1 (2): Test 2 (3): Test 3 (4): Test 4	Primary Screening Test Evaluated	Gold Standard Used	Sensitivity	Specificity	LR+/LR-
Wickberg and Hwang, 1996[72]	(1): EPDS: ≥11.5 (2): EPDS: ≥12.5	EPDS	Clinical psychologist using DSM-IIIR	Test 1: 96% Test 2: 85%	Test 1: 49% Test 2: 63%	Test 1: 2.34/2.30 Test 2: 0.08/0.24
Zich et al, 1990[73]	(1): CES-D ≥16 (2): CES-D ≥27 (3): BDI ≥10 (4): BDI ≥16	CES-D	DIS	Test 1: 100% Test 2: 100% Test 3: 100% Test 4: 100%	Test 1: 53% Test 2: 81% Test 3: 75% Test 4: 89%	Test 1: 2.13/-infinity Test 2: 5.3/infinity Test 3: 4.0/infinity Test 4: 9.1/infinity

Evidence Table 1. Screening Accuracy for Depression Studies (continued)

Adequate Inclusion Criteria?	Appropriate Spectrum of Patients?	Gold Standard Applied to All Subjects?	Masked Assessors?	Aggregate Internal Validity	External Validity	Overall Quality
Yes	Yes	Yes	Yes	Good	Good	Good
Yes	No	No	DK	Fair	Fair	Fair

Appendix D. Evidence Tables

Evidence Table 2. Pharmacologic Treatment in Depression Studies

Author, Year	Treatment Setting	Recruitment Setting	Conditions of Interest	How Diagnosis Made	Exclusions
Appleby et al, 1997[74]	Primary Care	OB wards of 2 hospitals in UK	Major depression or minor depression (post-partum)	SI: Revised clinical interview schedule, MD: RDC, OTH: Edinburgh postnatal depression scale	Substance abuse, suicidal, serious medical illness, breastfeeding, dysthymia, tx-resistant depression, psychosis
Mynors-Wallis et al, 2000[75]	Primary Care or patient's home	Primary Care	Major depression	PD DI: RDC OTH: HAM-D ≥13	Current substance abuse, suicidal, serious medical illness, current treatment for depression
Malt et al, 1999[76]	Primary Care	Primary Care	2 weeks symptons of "severe enough to require treatment" >3 CGI >20 MADS	DI: MADS OTH: CGI severity OTH: DSM-III R checklist	Bipolar, >40 MADS, current depression >1 yearr current alcohol abuse, severe suicidality, non-responding to adequate drug tx

Evidence Table 2. Pharmacologic Treatment in Depression Studies (continued)

Definition of Intervention	Study Duration	No. of potential subjects screened, randomized, analyzed	Definition of Population (inclusion):	Definition of Study Groups
Pharmacotherapy (fluoxetine) with 1 or 6 sessions CBT vs placebo with 1 or 6 sessions CBT	12 weeks	Eligible: 2,978 Screened: 2,395 Randomized: 87 Analyzed:87	Female, childbearing, duration of symptoms: 2 weeks, women 6-8 weeks Postpartum	Grp1: Fluoxetine+1session CBT Grp 2: Fluoxetine+ 6 session CBT Grp 3: Pl+1 session CBT Grp 4: Pl+ 6 session CBT
Psychotherapy (6 tx sesions of problem-solving tx) or drug tx (6 sessions with Fluoxetine or Paroxetine) vs combination of psychotherapy and medication (12 sessions)	52 weeks	Screened: 241 Randomized: 151 Analyzed: 151	Both female and male, 18-65, duration of symptoms: >4 weeks, off medications >4 weeks	Grp 1: B: Problem-solving by GP (PS-GP) Grp 2: B: Problem-solving by Nurse (PS-N) Grp 3: D:Fluoxetine or Paroxetine (Drug) Grp 4: OTH: Combination of drug and PS-N (Combo)
Drug treatment: sertraline (S) vs mianserin (M) vs placebo for 24 weeks	24 weeks	Screened: NR Randomized: 372 Analyzed: 372 S=122, M=121 P=129	Consecutive pts. 18-79 years in Norway	Grp 1: Placebo (Pl)(129) Grp 2: Sertraline (S) (122) Grp 3: Mianserin (M) (121)

Evidence Table 2. Pharmacologic Treatment in Depression Studies (continued)

Author, Year	Mean Age (years)	% Female	Ethnicity	Baseline Severity of Depression
Appleby et al, 1997[74]	Grp 1: 25.7 Grp 2: 26.6 Grp 3: 23.1 Grp 4: 26.0	100%	NR	Scale: HAM-D Grp 1: 14.4 Grp 2: 14.0 Grp 3: 14.0 Grp 4: 13.8
Mynors-Wallis et al, 2000[75]	Grp 1: 36 Grp 2: 33 Grp 3: 34 Grp 4: 35	Grp 1: 85% Grp 2: 68% Grp 3: 86% Grp 4: 69%	NR	Scale: HAM-D Grp 1: 20.5 Grp 2:20.5 Grp 3: 20.2 Grp 4: 19.8
Malt et al, 1999[76]	Grp 1: 47.8 Grp 2: 48.6 Grp 3: 48.2	Grp 1: 71.0% Grp 2: 78.0% Grp 3: 70.0%	NR	Scale: MADS S = 26.8 M = 26.8 P = 26.5

Appendix D. Evidence Tables

Evidence Table 2. Pharmacologic Treatment in Depression Studies (continued)

Baseline QOL/functional status measure(s)	Adherence Measured: Yes/No If yes, how/results?	Outcome Data Defined: (1) Main 1 (2) Sec. 2	Outcome Data Main 1:	Outcome Data Sec 2:
NR	No	(1): % change in Hamilton Scores at 12 weeks (intention-to-treat)	Grp 1 and Grp 2 (Fluoxetine): 66.9% decrease in HAM-D Grp 3 and Grp 4 (Placebo): 54.0% decrease in HAM-D	NR
Scale: Social adjustment scale Grp 1: 29.6 Grp 2: 28.6 Grp 3: 29.3 Grp 4: 29.0	No	(1) % recovered at 12 weeks (HAM-D <7) (2) % recovered at 52 weeks (HAM-D <7)	Grp 3 (Drug) : 67% recovered at 12 weeks Group 1 (PS-GP) 51% recovered Grp 2 (PS-N): 54% recovered Grp 4 (Combo): 60% recovered No significant difference between groups	Grp 1 (PS-GP): 62% recovered Grp 2 (PS-N): 56% recovered Grp 3 (Drug): 56%% recovered at 52 weeks Grp 4 (Combo): 66% recovered No significant difference between groups
Scale: CGIS S = 4.1 M = 4.0 P = 4.0	Yes, pill count, drug level	(1): Treatment Response (2): Change MADS (3): Complete response for 1st time depressed (4): Complete response pts. with reoccurrence	Clinical remission: (1) 50% reduction in MADS, (2) CGI global of 1,2, or 3, and (3) CGI improved 1 or 2 Grp 1 (Pl): 47% remission, OR= 1.0 Grp 2 (S): 61% remission, 0.56 (.33, .96) Grp 3 (M): 54% remission, 0.75 (.44, 1.27)	Grp 1: -12.5 Grp 2: -14.9 Grp 3: -15.5

Appendix D. Evidence Tables

Evidence Table 2. Pharmacologic Treatment in Depression Studies (continued)

Author, Year	Summary Measures: RRR Main Outcome (95% CI or *P* value)	Summary Measures: ARR Main Outcome (95% CI or *P* value)	Summary Measures: NNT (over time)	Adequate Inclusion Criteria?
Appleby et al, 1997[74]	NA	NA	NA	Yes
Mynors-Wallis et al, 2000[75]	NA: difference among treatments insignificant	NA	NA	Yes
Malt et al, 1999[76]	Sertraline 44% (4%, 67%), Mianserin 25% (+27%, 56%)	14%	7 (for sertraline)	Yes

Evidence Table 2. Pharmacologic Treatment in Depression Studies (continued)

Similar attrition between groups?	Adequate Randomization & Concealment?	Intention to Treat Analysis?	Internal Validity?	External Validity?	Overall Quality?
Yes	Yes	Yes	Fair	Fair	Fair
No: 36% PS-GP, 22% PS-N, 17% Drug, 17% Combo	Yes	Yes	Good	Good	Good
No 29% placebo, 16% S, 14% M	Yes	Yes	Good	Good	Good

Appendix D. Evidence Tables

Evidence Table 2. Pharmacologic Treatment in Depression Studies

Author, Year	Treatment Setting	Recruitment Setting	Conditions of Interest	How Diagnosis Made	Exclusions
Mynors-Wallis et al, 1995[77]	Primary Care or patient's home	Primary Care	Major depression	DI: HAM-D OTH: RDC	Bipolar disorder, substance abuse, suicidal, serious medical illness, current tx for depression
Schulberg et al, 1996[78]; Brown et al, 1996[79]; Coulehan et al, 1997[80]	Primary Care	Primary Care	Major depression	DI: CES-D, SI: DIS depression section, OTH: HAM-D	Bipolar disorder, current substance abuse, suicidal, pregnancy, serious medical illness, current tx for depression
Scott et al, 1992[81]	Primary Care	Primary Care	Major Depression	Physician diagnosis confirmed by SCI (DSM-III-R)	Current substance abuse, suicidal, OTH: Psychotic illness

Evidence Table 2. Pharmacologic Treatment in Depression Studies (continued)

Definition of Intervention	Study Duration	No. of potential subjects screened, randomized, analyzed	Definition of Population (inclusion):	Definition of Study Groups
Psychotherapy (6 tx sesions of problem-solving tx) or drug tx (AMI) vs placebo	12 weeks	Screened: 107 Randomized: 91 Analyzed: 82	Both female and male, 18-65, duration of symptoms: >2 weeks	Grp 1: B: Problem-solving (PS) Grp 2: D: Amitriptyline (AMI) Grp 3: OTH: Placebo (PI)
Drug tx (NOR, weekly/monthly visits) or psychotherapy (16 weeks of IPT) vs UC	32 weeks	Screened: 7,652 Randomized: 276 Analyzed: 276	Both female and male, 18-64, duration of symptoms: >2 weeks	Grp 1: D: NOTHR Grp 2: B: IPT Grp 3: OTH: UC
Drug tx (Amitriptyline), CT by psychologist (weekly up to 16 weeks), or counseling by SW (weekly up to 16 weeks) vs UC	16 weeks	Screened: 143 Randomized: 121 Analyzed:113	Both female and male, 18-65, duration of symptoms: 2 weeks	Grp 1: AMI Grp 2: COG Grp 3: SW counseling Grp 4: UC

Evidence Table 2. Pharmacologic Treatment in Depression Studies (continued)

Author, Year	Mean Age (years)	% Female	Ethnicity	Baseline Severity of Depression
Mynors-Wallis et al, 1995[77]	Grp 1: 37.3 Grp 2: 37.2 Grp 3: 37.0	Grp 1: 83.3% Grp 2: 77.4% Grp 3: 70.0%	NR	Scale: HAM-D Grp 1: 19.4 Grp 2: 19.1 Grp 3: 18.4
Schulberg et al, 1996[78]; Brown et al, 1996[79]; Coulehan et al, 1997[80]	Grp 1: 38.6 Grp 2: 37.1 Grp 3: 38.6	Grp 1: 80.2% Grp 2: 82.8% Grp 3: 87.0%	Grp 1: 59.3% white Grp 2: 53.8% white Grp 3: 53.3% white	Scale: CES-D Grp 1: 37.3 Grp 2: 36.7 Grp 3: 48.8
Scott et al, 1992[81]	Grp 1: 30.6 Grp 2: 28.8 Grp 3: 36.2 Grp 4: 31.6	Grp 1: 61.3% Grp 2: 83.3% Grp 3: 83.3% Grp 4: 63.7%	NR	Scale: HAM-D Grp 1: 18.2 Grp 2: 18.3 Grp 3: 15.7 Grp 4: 19.7

Evidence Table 2. Pharmacologic Treatment in Depression Studies (continued)

Baseline QOL/functional status measure(s)	Adherence Measured: Yes/No If yes, how/results?	Outcome Data Defined: (1) Main 1 (2) Sec. 2	Outcome Data Main 1:	Outcome Data Sec 2:
Scale: Social adjustment scale Grp 1: 22.3 Grp 2: 23.4 Grp 3: 21.3	Yes, counting capsules	(1) Recovery at 12 weeks vs placebo (2): Change in depression severity at 12 weeks vs placebo (HAM-D) (3): Change in depression severity at 12 weeks vs drug (HAM-D) (4): Change in soc adjust scale vs placebo	Recovery at 12 weeks: Grp 2 (AMI): 52% Grp 3 (PI): 27%	Grp 2 vs placebo: 3.75 (95%CI -0.59-8.09)
Scale: GAS Grp 1: 50.6 Grp 2: 50.7 Grp 3: 48.8	Yes, NOR blood level	(1): ITT repeated measures anova (HAM-D) over 8 months (2): % recovered at 8 months, ITT (3): % recovered at 8 months, tx completers	Grp 1 (NOR) vs Grp 3 (UC): NOR > UC Grp 1 (NOR) vs Grp 2 (IPT): NOR not different from IPT	Grp 1 (NOR): 48% Grp 3 (UC): 18%
NR	No	(1): Change in HAM-D(depressive severity) at 16 weeks vs UC (2): Percent recovered at 16 weeks (HAM-D)	Grp 1 (AMI) vs Grp 4 (UC): -0.4	Grp 1 (AMI): 58% Grp 4 (UC): 48%

Appendix D. Evidence Tables

Evidence Table 2. Pharmacologic Treatment in Depression Studies (continued)

Author, Year	Summary Measures: RRR Main Outcome (95% CI or *P* value)	Summary Measures: ARR Main Outcome (95% CI or *P* value)	Summary Measures: NNT (over time)	Adequate Inclusion Criteria?
Mynors-Wallis et al, 1995[77]	Recovery at 12 weeks (HAM-D) AMI vs placebo: 0.93	25%	4.00	Yes
Schulberg et al, 1996[78]; Brown et al, 1996[79]; Coulehan et al, 1997[80]	NOR vs UC (intention to treat): 1.67 NOR vs UC (tx completers): 2.35	NOR vs UC (intention to treat): 30% NOR vs UC (tx completers): 47%	NOR vs UC (intention to treat): 3.3 NOR vs UC (tx completers): 2.13	Yes
Scott et al, 1992[81]	AMI vs UC: % recovered at 16 weeks: 20.8%	10%	10.00	Yes

Evidence Table 2. Pharmacologic Treatment in Depression Studies (continued)

Similar attrition between groups?	Adequate Randomization & Concealment?	Intention to Treat Analysis?	Internal Validity?	External Validity?	Overall Quality?
Yes	Yes	No, analyzed only those completing at least 4 sessions	Good	Good	Good
Yes	Yes	Yes	Good	Good (but high intensity)	Good
Yes	Yes	No, analyzed only those beginning treatment	Fair	Fair	Fair

Appendix D. Evidence Tables

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies

Author, Year	Treatment Setting	Recruitment Setting	Conditions of Interest	How Diagnosis Made	Exclusions
Appleby et al, 1997[74]	Primary Care	OTH: OB wards of 2 hospitals in UK	Major depression or minor depression (Post-partum)	SI: Revised clinical interview schedule, MD: RDC, OTH: Edinburgh postnatal depression scale	Substance abuse, suicidal, serious medical illness, breastfeeding, dysthymia, tx-resistant depression, psychosis
Mynors-Wallis et al, 2000[75]	Primary Care or patient's home	Primary Care	Major depression	PD DI: RDC OTH: HAM-D ≥13	Current substance abuse, suicidal, serious medical illness, current treatment for depression
Mynors-Wallis et al, 1995[77]	Primary Care OR patient's home	Primary Care	Major Depression	HAM-D, MD, RDC	Bipolar disorder, substance abuse, suicidal, serious medical illness, current tx for depression

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Definition of Intervention:	Study Duration	No. of potential subjects screened, randomized, analyzed	Definition of Population (inclusion)	Definition of Study Groups
Pharmacotherapy with 1 or 6 sessions CBT; placebo with 1 or 6 sessions CBT	12 weeks	Screened: 2,978 Randomized: 87 Analyzed:87	Female, Childbearing, Duration of symptoms: 2 weeks, Women 6-8 weeks postpartum	Grp 1: Fluoxetine, T: 1 session CBT Grp 2: Fluoxetine, T: 6 session CBT Grp 3: Placebo, T: 1 session CBT Grp 4: Placebo, T: 6 session CBT
Psychotherapy (6 tx sesions of problem-solving tx) or drug tx (6 sessions with Fluoxetine or Paroxetine) vs. combination of psychotherapy and medication (12 sessions)	52 weeks	Screened: 241 Randomized: 151 Analyzed: 151	Both female and male, 18-65, duration of symptoms: >4 weeks, off medications >4 weeks	Grp 1: B: Problem-solving by GP (PS-GP) Grp 2: B: Problem-solving by Nurse (PS-N) Grp 3: D:Fluoxetine or Paroxetine (Drug) Grp 4: OTH: Combination of drug and PS-N (Combo)
Psychotherapy (6 tx sessions of problem-solving tx) or pharmacothx vs. UC	12 weeks	Screened: 107 Randomized: 91 Analyzed: 82	Both female and male, 18-65 years, duration of symptoms: > 2 weeks	Grp 1: B: PS Grp 2: D: AMI Grp 3: OTH: Pl

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Author, Year	Mean Age (years)	% Female	Ethnicity	Baseline Severity of Depression
Appleby et al, 1997[74]	Grp 1: 25.7 Grp 2: 26.6 Grp 3: 23.1 Grp 4: 26.0	100%	NR	Scale: HAM-D Grp 1: 14.4 Grp 2: 14.0 Grp 3: 14.0 Grp 4: 13.8
Mynors-Wallis et al, 2000[75]	Grp 1: 36 Grp 2: 33 Grp 3: 34 Grp 4: 35	Grp 1: 85% Grp 2: 68% Grp 3: 86% Grp 4: 69%	NR	Scale: HAM-D Grp 1: 20.5 Grp 2: 20.5 Grp 3: 20.2 Grp 4: 19.8
Mynors-Wallis et al, 1995[77]	Grp 1: 37.3 Grp 2: 37.2 Grp 3: 37.0	Grp 1: 83.3% Grp 2: 77.4% Grp 3: 70.0%	NR	Scale: HAM-D Cut-off: > 13 Grp 1: 19.4 Grp 2: 19.1 Grp 3: 18.4

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Baseline QOL/functional Status Measure(s)	Adherence Measured: Yes/No If yes, how/results?	Outcome Data Defined: (1) Main 1 (2) Sec. 2	Outcome Data Main 1:	Outcome Data Sec 2:
NR	No	(1): % change in Hamilton Scores at 12 weeks (intention-to-treat), immediately post-treatment	Grp 2 and Grp 4 (6 sessions): 64.0% decrease in HAM-D Grp 1 and Grp 3 (1 session): 57.7% decrease in HAM-D	NR
Scale: Social adjustment scale Grp 1: 29.6 Grp 2: 28.6 Grp 3: 29.3 Grp 4: 29.0	No	(1) % recovered at 12 weeks (HAM-D <7) (2) % recovered at 52 weeks (HAM-D <7)	Grp 1 (PS-GP): 51% recovered at 12 weeks Grp 2 (PS-N): 54% recovered Grp 3 (Drug): 67% recovered Grp 4 (Combo): 60% No significant difference between groups	Grp 1 (PS-GP): 62% recovered at 52 weeks Grp 2 (PS-N): 56% recovered Grp 3 (Drug): 56%% recovered Grp 4 (Combo): 66% recovered No significant difference between groups
Scale: Social adjustment scale Grp 1: 22.3 Grp 2: 23.4 Grp 3: 21.3	Yes, counting capsules	(1) Recovery at 12 weeks (immediately post-tx) (2): Change in depr severity at 12 weeks vs placebo (HAM-D)	Recovery at 12 weeks Grp 1 (PS): 60.0% Grp 3 (Pl): 27.0%	Grp 1 (PS) vs Grp 3 (Pl): 4.69 (95% CI 0.41-8.96)

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Author, Year	Summary Measures: RRR Main Outcome (95% CI or *P* value)	Summary Measures: ARR Main Outcome (95% CI or *P* value)	Summary Measures: NNT (over time)	Adequate Inclusion Criteria?
Appleby et al, 1997[74]	NA	NA	NA	Yes
Mynors-Wallis et al, 2000[75]	NA: difference among treatments insignificant	NA	NA	Yes
Mynors-Wallis et al, 1995[77]	Recovery at 12 weeks (HAM-D) (1): PS vs. placebo: 1.22	33%	3	Yes

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Similar Attrition between groups?	Adequate Randomization & Concealment?	Intention to Treat Analysis?	Internal Validity?	External Validity?	Overall Quality?
Yes	Yes	Yes	Fair	Fair	Fair
No: 36% PS-GP, 22% PS-N, 17% Drug, 17% Combo	Yes	Yes	Good	Good	Good
Yes	Yes	No; analyzed only those completing at least 4 sessions	Good	Good	Good

Appendix D. Evidence Tables

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies

Author, Year	Treatment Setting	Recruitment Setting	Conditions of Interest	How Diagnosis Made	Exclusions
Schulberg et al, 1996[78]; Brown et al, 1996[79]; Coulehan et al, 1997[80]	Primary Care	Primary Care	Major depression	DI: CES-D, SI: DIS depression section, OTH: HAM-D	Bipolar disorder, substance abuse, suicidal, pregnancy, serious medical illness, prior tx for depression, current tx for depression
Scott et al, 1992[81]	Primary Care	Primary Care	Major depression	SI, MD	Substance abuse, suicidal, OTH: Psychotic illness
Blackburn et al, 1981[83]	Primary Care	Primary Care	Major Depression	DI: RDC OTH: Present state exam, BDI: >14	Bipolar disorder, Psychotic illness

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Definition of Intervention:	Study Duration	No. of potential subjects screened, randomized, analyzed	Definition of Population (inclusion)	Definition of Study Groups
Drug tx (NOR, weekly-monthly visits) or psychotherapy (12-16 sessions of IPT) vs. UC	8 months	Screened: 7,652 Randomized: 276 Analyzed: 276	Both female and male, 18-64 years, duration of symptoms: > 2 weeks, no correct tx for mood disorder	Grp 1: NOR Grp 2: IPT Grp 3: UC
Pharmacology, CT by psychologists, or supportive counselling by "qualified" SW	16 weeks	Screened: 143 Randomized: 121 Analyzed: 113	Both female and male, 18-65 years, duration of symptoms: 2 weeks	Grp 1: AMI Grp 2: CT Grp 3: SW Grp 4: UC
Pharmacology or cognitive psychotherapy	20 weeks	Screened: 140 Randomized: 88 Analyzed: 64	Both female and male, 18-65 years, duration of symptoms: 2 weeks	Grp 1: CT Grp 2: Anti-depressant drug of choice (TCA) Grp 3: CT and anti-depressant combined

Appendix D. Evidence Tables

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Author, Year	Mean Age (years)	% Female	Ethnicity	Baseline Severity of Depression
Schulberg et al, 1996[78]; Brown et al, 1996[79]; Coulehan et al, 1997[80]	Grp 1: 38.6 Grp 2: 37.1 Grp 3: 38.6	Grp 1: 80.2% Grp 2: 82.8% Grp 3: 87.0%	Grp 1: 59.3% White Grp 2: 53.8% White Grp 3: 53.3% White	Scale: CES-D Grp 1: 37.3 Grp 2: 36.7 Grp 3: 38.0
Scott et al, 1992[81]	Grp 1: 30.6 Grp 2: 28.8 Grp 3: 36.2 Grp 4: 31.6	Grp 1: 61.3% Grp 2: 83.3% Grp 3: 83.3% Grp 4: 63.7%	NR	Scale: HAM-D Grp 1: 18.2 Grp 2: 18.3 Grp 3: 15.7 Grp 4: 19.7
Blackburn et al, 1981[83]	Grp 1: NR	Grp 1: NR	Grp 1: NR	Grp 1: NR

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Baseline QOL/functional Status Measure(s)	Adherence Measured: Yes/No If yes, how/results?	Outcome Data Defined: (1) Main 1 (2) Sec. 2	Outcome Data Main 1:	Outcome Data Sec 2:
Scale: Global assessment score Grp 1: 50.6 Grp 2: 50.7 Grp 3: 48.8	Yes, nortriptyline blood level	(1): Intent to treat, repeated measures anova of HAM-D over 8 mos (2): % recovered at 8 months for intent to treat (4 mos post-tx) (3): % recovered at 8 months tx completers	Grp 2 (IPT) vs Grp 3 (pl): IPT> UC; Grp 2 (IPT) vs Grp 1 (NOR): NOR not different from IPT	Grp 1 (IPT): 46% Grp 3 (Pl): 18%
NR	No	(1): Change in HAM-D (depressive severity) at 16 weeks vs. UC (immediately post-tx) (2): % recovered at 16weeks (HAM-D) (immediately post-tx)	Grp 2 (CT) vs. Grp 4 (UC): -1.7 Grp 3 (SW) vs. Grp 4 (UC): -3.5 (p=0.05 vs. UC)	Grp 2 (CT): 41% Grp 3 (SW): 72% (p=0.05 vs. UC) Grp 4 (UC): 48%
Grp 1: NR	No	% responding (>=50% decrease in HAM-D or BDI score)	Grp 1 (CT): 72.7% Grp 2 (TCA): 55.0% Grp 3 (CT+TCA): 81.8% No significant difference between groups	NR

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Author, Year	Summary Measures: RRR Main Outcome (95% CI or *P* value)	Summary Measures: ARR Main Outcome (95% CI or *P* value)	Summary Measures: NNT (over time)	Adequate Inclusion Criteria?
Schulberg et al, 1996[78]; Brown et al, 1996[79]; Coulehan et al, 1997[80]	IPT vs. UC (intention to treat): 1.56 IPT vs. UC (tx completers): 2.6	IPT vs. UC (intention to treat): 28% IPT vs. UC (tx completers): 52%	IPT vs. UC (intention to treat): 3.57 IPT vs. UC (tx completers): 1.92	Yes
Scott et al, 1992[81]	% recovered at 16 weeks: SW vs UC: 50%	24%	UC= 4.2	Yes
Blackburn et al, 1981[83]	NA	NA	NA	Yes

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Similar Attrition between groups?	Adequate Randomization & Concealment?	Intention to Treat Analysis?	Internal Validity?	External Validity?	Overall Quality?
Yes	Yes	Yes	Good	Good	Good
Yes	DK	No; analyzed only those beginning treatment	Fair	Fair	Fair
Yes	DK	No	Fair	Fair	Fair

Appendix D. Evidence Tables

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies

Author, Year	Treatment Setting	Recruitment Setting	Conditions of Interest	How Diagnosis Made	Exclusions
Holden et al, 1989[84]	Primary Care	Primary Care	Major depression or minor depression (Post-partum)	DI: Edinburgh post-natal depression scale at 6 weeks post-partum, SI: Goldberg's standardized psychiatric interview at 12 weeks to include RDC (Spitzer)	DK
Katon et al, 1996[85]	Primary Care	Primary Care	Major or minor depression	PD, DI: HSCL-20: > .075, IDS used to characterize minor depression vs. major depression	Substance abuse, suicidal, pregnancy, Serious medical illness, Psychosis, dementia, plan to withdraw from HMO
Lynch et al, 1997[86]	Primary Care	Primary Care	Minor depression	MOS-D, DIS	Already receiving counseling
Miranda et al, 1994[87]	Primary Care	Primary Care	Minor depression (BDI > 18 without diagnosis of major depression or dysthymia)	DIS	Substance abuse, major depression, dysthymia, psychotic disorder, not English or Spanish speaking

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Definition of Intervention:	Study Duration	No. of potential subjects screened, randomized, analyzed	Definition of Population (inclusion)	Definition of Study Groups
Supportive counseling by health visitors	13 weeks	Screened: 734 Randomized: 55 Analyzed: 50	Female, Childbearing, duration of symptoms: 2 weeks, women 12 weeks postpartum	Grp 1: Health visitor counselling (HV) Grp 2: UC
Structured Depression Program with cognitive-behavioral treatments and counseling re medication adherence	7 months	Screened: 217 Randomized: 153 Analyzed: 153	18-80 years, Both male & female, off meds, Duration of symptoms > 2 weeks	Grp 1: Depression program for major depression (MD-UC) Grp 2: Usual care for major depression (MD_UC) Grp 3: Depression program for minor depression (mD-CT) Grp 4: UC for minor depression (mD-UC)
Telephone counseling (TC) using problem-solving therapy	6 sessions	Screened: 239 Randomized: 29 Analyzed: 24	Female and male, > 18 years	Grp 1:TC Grp 2: Placebo (Pl)
Cognitive behavioral therapy (CBT)	12 months	Screened: 708 Randomized: 150 Analyzed:150	Both male and female, 18-69 years	Grp 1: CBT Grp 2: Usual care (UC)

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Author, Year	Mean Age (years)	% Female	Ethnicity	Baseline Severity of Depression
Holden et al, 1989[84]	Grp 1: 27.6 Grp 2: 24.6	100%	NR	Scale: Edinburgh post-natal depression scale Grp 1: 16.0 Grp 2: 15.5
Katon et al, 1996[85]	Grp 1: 43.1 Grp 2: 44.8 Grp 3: 49.2 Grp 4: 47.2	Grp 1: 77.4% Grp 2: 73.5% Grp 3: 71.7% Grp 4: 73.8%	Grp 1: 77.4% white Grp 2: 91.2% white Grp 3: 91.3% white Grp 4: 85.7% white	Scale: SCL-20 Cut-off: 0.75 Grp 1: 2.46 Grp 2: 1.77 Grp 3: 2.35 Grp 4: 1.62
Lynch et al, 1997[86]	Grp 1: 46.8 Grp 2: 49.9	Grp 1: 87.5% Grp 2: 85.7%	NR	Scale: HAM-D Grp 1: 14.4 Grp 2: 12.4
Miranda et al, 1994[87]	Grp 1: 52.5 (NS difference in groups)	Grp 1: 62.0% (NS difference between groups)	Grp 1: 35.1% White 23.7% Black 24.3% Hispanic 10.1% Other	NR

Appendix D. Evidence Tables

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Baseline QOL/functional Status Measure(s)	Adherence Measured: Yes/No If yes, how/results?	Outcome Data Defined: (1) Main 1 (2) Sec. 2	Outcome Data Main 1:	Outcome Data Sec 2:
NR	N/A	(1) Recovery from depression at 13 weeks (~ 1 mon. post-tx)	Grp 1 (HV): 69.0% Grp 2 (UC): 38.0%	NR
Scale: NR	Automated pharmacy data, phone interview, # sessions attended, 74% with 6 sessions and 9% with 5 sessions	(1) % > 50% improvement in SCL-20 at 4 mos(1.5-2 mos post-tx); (2) Medication Adherence at 7 mos (4.5-5 mos post-tx); (3)	Grp 1 (MD-CT): 70.4% Grp 2 (MD-UC): 42.3% Grp 3 (mD-CT): 66.7% (p = 0.22) Grp 4 (mD-UC): 52.8% (p = 0.22)	Grp 1 (MD-CT): 79.0% Grp 2 (MD-UC): 54.0% Grp 3: (mD-CT): 65.0% Grp 4 (mD-UC): 41.0%
Scale: NR	No	(1): HAM-D mean score (2): BDI mean	Grp 1 (TC): 10.9 Grp 2 (PI): 13.3 Nonsignificant difference between groups	Grp 1: 12.9 Grp 2: 22.4 Significant difference between groups p=0.00
NR	NR	(1) Mean depressive severity over 1 yr. (including 10 mos post-tx) (2) No. of missed medical appointments over 1yr.	Grp 1 (CBT) had greater reduction in depressive severity than Grp 2 (UC)	Grp 1 (CBT) (0.57) had fewer missed medical appointments than Grp 2 (UC) (1.22) (p=0.05)

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Author, Year	Summary Measures: RRR Main Outcome (95% CI or *P* value)	Summary Measures: ARR Main Outcome (95% CI or *P* value)	Summary Measures: NNT (over time)	Adequate Inclusion Criteria?
Holden et al, 1989[84]	HV vs. no intervention: % recovered at 13 weeks: 0.83	31.70%	3.15	Yes
Katon et al, 1996[85]	Outcome 1: Grp1(MD-CT) vs Grp2 (MC-UC), RRR=66.4%; Outcome 2: Grp 1 vs Grp 2 (p = 0.07), RRR:46.3%	Outcome 1: ARR: 25% Outcome 2: ARR: 32.5%	Outcome 1: NNT: 3.56 Outcome 2: NNT: 4 Outcome 3: NNT: 3.07	Yes
Lynch et al, 1997[86]	NA	NA	NA	Yes
Miranda et al, 1994[87]	NA	NA	NA	Yes

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Similar Attrition between groups?	Adequate Randomization & Concealment?	Intention to Treat Analysis?	Internal Validity?	External Validity?	Overall Quality?
Yes	Yes	No	Fair (not ITT)	Fair	Fair
Yes	Yes	Yes	Good	Good	Good
No: TC lost 4/15, UC 1/14	Yes	No	Fair	Fair	Fair
Yes	DK	Yes	Fair	Fair	Fair

Appendix D. Evidence Tables

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies

Author, Year	Treatment Setting	Recruitment Setting	Conditions of Interest	How Diagnosis Made	Exclusions
Ross et al, 1985[88]	Primary Care	Primary Care	Major depression	DI: RDC, PD, OTH: BDI: >14	DK
Scott et al, 1997[89]	Primary Care	Primary Care	Major depression	DI: BDI, MD, OTH: Interview by psychiatrist	Bipolar disorder, OTH: Dysthymia, psychosis, previous exposure to cognitive therapy, organic brain damage, illiterate
Teasdale et al, 1984[90]	Primary Care	Primary Care	Major depression	DI: Research criteria	OTH: psychotic

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Definition of Intervention:	Study Duration	No. of potential subjects screened, randomized, analyzed	Definition of Population (inclusion)	Definition of Study Groups
Cognitive therapy	12 weeks	Screened: Blank Randomized: 51 Analyzed: 51	Male and female, Duration of symptoms: 2 weeks	Grp 1: B: Individual cognitive (I-CT) Grp 2: B: Group cognitive therapy (G-CT) Grp 3: Usual Care (UC)
Cognitive therapy (CT)	12 months	Screened: NR Randomized: 48 Analyzed: 48	Both male and female, 18-65 years, duration of symptoms: > 2 weeks And < 2 years	Grp 1: CT Grp 2: UC
Cognitive therapy (CT)	6 months (3 months of tx, follow-up assessment 3 months later)	Screened: NR Randomized: 44 Analyzed: 33	Both female and male, 18-60 years, duration of symptoms: 2 weeks	Grp 1: B: CT Grp 2: UC

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Author, Year	Mean Age (years)	% Female	Ethnicity	Baseline Severity of Depression
Ross et al, 1985[88]	Grp 1 vs Grp 2: "similar"	Grp 1 vs Grp 3: "similar"	Grp 1 vs Grp 3: "similar"	Scale: NR
Scott et al, 1997[89]	Grp 1: 41 (NS between groups)	Grp 1: 67% (NS between groups)	NR	Scale: HAM-D Cut-off: > 20 Grp 1: 21.4 Grp 2: 22.5
Teasdale et al, 1984[90]	Grp 1: 38.0 Grp 2: 37.0	Grp 1: 94% Grp 2: 94%	NR	Scale: BDI Grp 1 (CT): 30 Grp 2 (UC): 29

Appendix D. Evidence Tables

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Baseline QOL/functional Status Measure(s)	Adherence Measured: Yes/No If yes, how/results?	Outcome Data Defined: (1) Main 1 (2) Sec. 2	Outcome Data Main 1:	Outcome Data Sec 2:
Scale: NR	No	(1): Change in MADS score (immediately post-tx)	Grp 1 (I-CT): Significant vs. UC:p < 0.05 Grp 2 (G-CT): Significant vs. UC: p < 0.05	NR
NR	N/A	(1): Recovery at 7 weeks (immediately post-tx) (2): Mean depressive severity at 7 weeks (BDI), immediately post-tx	Grp 1 (CT): 62.5% Grp 2 (UC): 33.3% (p<0.05, Grp1 vs. 2)	Grp 1 (CT): 17.7 Grp 2 (UC): 22.7 (p<0.05, Grp 1 vs. 2)
NR	N/A	(1): BDI <14 immediately post-tx (2): BDI < 14 at 3 mos post-tx	Grp 1 (CT): 14/17 (82%) Grp 2 (UC): 4/17 (24%)	Grp 1 (CT): 10/17 (59%) Grp 2 (UC): 9/17 (53%)

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Author, Year	Summary Measures: RRR Main Outcome (95% CI or *P* value)	Summary Measures: ARR Main Outcome (95% CI or *P* value)	Summary Measures: NNT (over time)	Adequate Inclusion Criteria?
Ross et al, 1985[88]	NA	NA	NA	Yes
Scott et al, 1997[89]	CT vs. UC: % recovery at 7 weeks: 0.87	29%	3.40	Yes
Teasdale et al, 1984[90]	NA	NA	NA	Yes

Evidence Table 3. Psychotherapeutic Treatment for Depression Studies (continued)

Similar Attrition between groups?	Adequate Randomization & Concealment?	Intention to Treat Analysis?	Internal Validity?	External Validity?	Overall Quality?
DK	Yes	Yes	Good	Fair	Fair
No	Yes	Yes (for 7 weeks recovery only)	Fair	Fair (unclear # screened)	Fair
Yes	DK	No	Fair	Fair	Fair

Appendix D. Evidence Tables

Table 4. Screening Outcomes for Depression Studies

Author, Year	Screening Setting	Definition of Population	Exclusions	Screening Instrument	Other Rating	Confirmatory Exam
Callahan et al, 1994[91]	Primary Care (academic)	Age >60 with elevated CES-D and HAM-D scores	Prisoners, Nursing home residents, Non-English speakers, Hearing impaired	CES-D	HAM-D	SIP
Callahan et al, 1996[92]	Primary Care (academic)	First patient of providers randomized to intervention or control	NR	CES-D	CAGE Short, portable mental status questionnaire	HAM-D, SIP
Dowrick, 1995[93]	Primary Care (community)	Patients who were not diagnosed as depressed on a visit	Beck >35 Clearly suicidal	Beck	Hamilton Scale RDC criteria for probable depression	ICD 10 based structured interview
Johnstone and Goldberg, 1976[94]	Primary Care	Primary care attenders	NR	GHQ	Physician recognition	NR
Linn and Yager, 1980[95]	Primary Care	New patients attending an academic Medicine clinic	Non-English speaking	Zung SDS	Physician rating	None

Table 4. Screening Outcomes for Depression Studies (continued)

Study Groups & Intervention Defined	Number of Subjects (intervention/ control)	Study Duration	Mean Age (years)	% Female	Ethnicity	Education
Int.: 3 addtl appointments, addtl education, feedback of HAM-D, list of hazardous meds	100 intervention 75 control	9 months	65	76%	48% White 51% Black	9 years mean
Int.: Physicians received HAM-D scores plus explicit treatment recommendations	3767 screened 515 eligible with CES-D >=16 254 enrolled 222 analyzed	1 year	65.9	76%	46%	10% greater than HS
Int.: Feedback of Beck results at 0 and 6 months Ctrl: No feedback	116 patients rated as not depressed and with Beck >14 52 to intervention 64 to control	12 months	NR	NR	NR	NR
Feedback of GHQ results	1,093 screened 60 intervention 59 control	1 year	NR	NR	NR	NR
Complicated 6 group design, results here for Grp. one (screening) vs Grps. 5 and 6 (no screening)	150 total	No follow-up	56	71%	60% White 40% Non-White	NR

Table 4. Screening Outcomes for Depression Studies (continued)

Author, Year	Outcome Defined: (1) Main 1 (2) Sec. 2 (3) Sec. 3 (4) Sec. 4	Outcome Main 1:	Outcome Sec 2:	Outcome Sec 3:	Outcome Sec 4:
Callahan et al, 1994[91]	1) New chart documentation of depression 2) Started antidepressant 3) Change in HAM-D 0-9 months	Int.: 32% Ctrl:12%	Int.: 26% Ctrl: 8%	Int.: -7 Ctrl: -7	NR
Callahan et al, 1996[92]	1) Chart documentation of depression or depressive symptoms 2) Likelihood of depression rated >50% 3) Intention to treat (any form) 4) Initiation of treatment	Int.: 86.7% Ctrl: 40.4% *P* =.001	Int.: 49.6% Ctrl: 54.8% *P* =0.44	Int.: 63.8% Ctrl: 18.2% *P* =.001	Int.: 46.5% Ctrl: 29.0% *P* =.001
Dowrick, 1995[93]	1) % diagnosed as depressed 2) Odds ratio for detection 3) % with def. plan to treat depression	Int.: 35% Ctrl: 21 %	2.10 (0.84, 5.28)	Int.: 27% Ctrl: 21% *P* =NS	NR
Johnstone and Goldberg, 1976[94]	1. Mean GHQ scores at 12 months overall 2. Mean GHQ scores at 12 months (severe cases) 3. Duration of first episode (full sample) 4. Total duration of depression (full sample)	No difference	Int.: 11.0 Ctrl: 22.7	Int.: 2.8 months Ctrl: 5.3 months *P* <.01	Int.: 4.2 months Ctrl: 6.3 months *P* <0.01
Linn and Yager, 1980[95]	1) Recognition and notation 2) Treatment	Int.: 29% Ctrl: 8%	Int.: 13% Ctrl: 8%	NR	NR

Table 4. Screening Outcomes for Depression Studies (continued)

Adequate Inclusion Criteria?	Drop-outs similar in Intervention & Control Groups?	Adequate Randomization & Conceal-ment?	Intention to Treat Analysis?	Internal Validity?	External Validity?	Overall Quality?	Comments
Yes	Yes	No	Yes	Fair	Fair	Fair	
Yes	Yes	No	Yes	Fair	Fair	Fair	
Yes	Yes	Yes	Yes	Good	Fair	Good	
DK	Yes	No	DK	Fair	Fair	Fair	
Yes	Yes	Yes	Yes	Good	Good	Good	

Appendix D. Evidence Tables

Table 4. Screening Outcomes for Depression Studies

Author, Year	Screening Setting	Definition of Population	Exclusions	Screening Instrument	Other Rating	Confirmatory Exam
Lewis et al, 1996[96]	Primary care (low income population in London)	Consecutive attenders 18-70	Unable to complete screening in English	GHQ	NR	PROQSY
Magruder-Habib et al, 1990[97]	Primary Care (VA)	All patients during study period	Age >90; Women; Hospital employee; Depression noted within 6 months in chart	Zung SDS	NR	DIS
Moore et al, 1978[98]	Primary Care (academic)	All patients seen during 8 weeks study, ages 20-60 years	NR	Zung SDS	Chart audit to determine if depression noted	NR
Rand et al, 1988[99]	Primary Care (academic)	Consecutive patients attending family practice training clinics	<18 years old "Too sick" Non-English speaking "Mentally disabled"	GHQ 28	Chart review of physician notation of symptoms, diagnoses, treatment	NR
Reifler et al, 1996[100]	Primary Care (academic)	Random sample of scheduled patients	Missed appointment; unable to read English	SDDS-PC	SF-36 Zung SDS	SDDS diagnostic module (for those screening positive)

Appendix D. Evidence Tables

Table 4. Screening Outcomes for Depression Studies (continued)

Study Groups & Intervention Defined	Number of Subjects (intervention/ control)	Study Duration	Mean Age (years)	% Female	Ethnicity	Education
Patients scoring <2 assigned to: Grp 1: GHQ feedback Grp 2: PROQSY feedback Grp 3: Control - no feedback	1,937 screened 851 screened positive 146 refused 227 assigned to each group	6 months	Grp 1: 39.5 Grp 2: 38.7 Grp 3: 37.4	Grp 1: 70% Grp 2: 61% Grp 3: 68%	NR	NR
Feedback of SDS scores	880 screened 100 depressed Grp 1: 48 feedback grp. Grp 2: 52 control grp.	1 year	Grp 1: 57.9 Grp 2: 61.9	Grp 1: 0% Grp 2: 0%	Grp 1: 68.8% White Grp 2: 67.3% White	NR
Int.: Feedback on screening results for scores >50, o/w same as controls Ctrl: Notification that screening performed	213 contacted 212 completed SDS	None	NR	NR	NR	NR
Feedback of GHQ results to intervention physicians vs usual care	434 eligible pts. 356 did GHQ 260 enrolled	3 months	NR	50%	50%	NR
Int.: Received screening results and asked to complete diagnostic module Ctrl: No feedback	605 contacted 501 eligible 358 enrolled 186 intervention 172 control	3 months	40.8	76%	56% White; 32% Black	NR

Appendix D. Evidence Tables

Table 4. Screening Outcomes for Depression Studies (continued)

Author, Year	Outcome Defined: (1) Main 1 (2) Sec. 2 (3) Sec. 3 (4) Sec. 4	Outcome Main 1:	Outcome Sec 2:	Outcome Sec 3:	Outcome Sec 4:
Lewis et al, 1996[96]	1) Mean GHQ at 6 weeks 2) Mean GHQ at 3 months 3) Mean GHQ at 6 months	1) 27.2 2) 25.7 3) 26.6 *P*=0.04	1) 27.0 2) 25.5 3) 26.4 *P*=0.07	1) 26.8 2) 25.4 3) 28.9	NR
Magruder-Habib et al, 1990[97]	1) % recognized as depressed initially 2) % recognized at 1 year 3) % treated initially 4) % treated at 1 year	Int.: 25% Ctrl: 7.7%	Int: 42% Ctrl: 21%	Int.: 28% Ctrl: 3.8%	Int.: 56% Ctrl: 42%
Moore et al, 1978[98]	1) Chart documentation of diagnosis for subjects with SDS ≥50	Int.: 56% Ctrl: 22%	NR	NR	NR
Rand et al, 1988[99]	1) % of residents increasing rate of making diagnosis of depression (pre and post) 2) % of patients with any psychiatric diagnosis 3) % patients with depression who received treatment	Int.: 75% Ctrl: 25%	Int.: pre 7%, post 16% Ctrl: pre 13%, post 12%	"Significant" increase in the use of antidepressants for experimental group	NR
Reifler et al, 1996[100]	1) Health care visits for those screening positive (mean ±SD) 2) Change in SF-36 3) Change in Zung SDS 4) Satisfaction	Int.: 3.7 ±3.9 Ctrl.: 5.3 ±6.7 *P*=0.06	No difference	No difference	No difference

Table 4. Screening Outcomes for Depression Studies (continued)

Adequate Inclusion Criteria?	Drop-outs similar in Intervention & Control Groups?	Adequate Randomization & Conceal-ment?	Intention to Treat Analysis?	Internal Validity?	External Validity?	Overall Quality?	Comments
Yes	Yes	Yes	Yes	Good	Fair	Fair	General mental health screening, not specific to depression
Yes	Yes	Yes	Yes	Good	Fair	Good	
Yes	Yes	Yes	Yes	Good	Good	Good	
Yes	Yes	No	NA	Fair	Fair	Fair	
Yes	Yes	Yes	No	Good	Good	Good	18% prevalence of MDD or subsyndromal depression

Appendix D. Evidence Tables

Table 4. Screening Outcomes for Depression Studies

Author, Year	Screening Setting	Definition of Population	Exclusions	Screening Instrument	Other Rating	Confirmatory Exam
Wells et al, 2000[101]	Multiple primary care clinics	Consecutive clinic patients with current depressive symptoms	<18 years old; acute medical emergency; no insurance; non-English or Spanish speaker	2 question screener	CES-D	CIDI
Williams et al, 1999[102]	Primary Care	Adult primary care attenders	NR	1) Single question 2) CES-D 20	SF-36 CAGE DUSOI rand visit rating scale MD recognition from chart review	DIS
Zung and King, 1983[103]	Primary Care (private practice)	Consecutive clinic patients	<20 years old Mentally retarded Pregnant Prev dx of dep	Zung SDS	CGI severity of illness CGI improvement	DSM-IIIR

Table 4. Screening Outcomes for Depression Studies (continued)

Study Groups & Intervention Defined	Number of Subjects (intervention/ control)	Study Duration	Mean Age (years)	% Female	Ethnicity	Education
Randomization by clinic; intervention clinics received feedback on screening and educational intervention	46 clinics 27,332 patients screened 1,356 randomized 913 intervention 443 controls	12 months	43.7	71%	57% White 30% Hispanic 7% African Amercican 6% Other	81% High school or greater
Grp 1: Single question vs Grp 2: CES-D vs Grp 3: Usual care	Grp 1: 330 Grp 2: 323 Grp 3: 316	3 months	Grp 1: 58 Grp 2: 59 Grp 3: 56	Grp 1: 68% Grp 2: 74% Grp 3: 71%	30% White 60 % Hispanic 10% African American	Grp 1: 11 years Grp 2: 10 years Grp 3: 11 years
Feedback of SDS scores and treatment with alprazolam vs usual care	499 patients screened 60 scored >55 49 + on DSM-III Grp 1: 23 to feedback Grp 2: 26 to control	4 weeks	Grp 1: 53.4 Grp 2: 65.9	NR	NR	NR

Table 4. Screening Outcomes for Depression Studies (continued)

Author, Year	Outcome Defined: (1) Main 1 (2) Sec. 2 (3) Sec. 3 (4) Sec. 4	Outcome Main 1:	Outcome Sec 2:	Outcome Sec 3:	Outcome Sec 4:
Wells et al, 2000[101]	1) Appropriate care at 12 months 2) Depressed on CES-D at 12 months 3) Still (+) on screener at 12 months 4) Employment among those initially employed	Int: 59% Ctrl: 50% *P*=0.006	Int: 54% Ctrl: 61% *P*=0.04	Int: 42% Ctrl: 51% *P*=0.005	Int: 90% Ctrl: 85% *P*=0.05
Williams et al, 1999[102]	1) Sensitivity/specificity of screening tests 2) Recognition of depression 3) Treatment or referral rates 4) Depression at 3 months	1) Single question sens = 85% spec = 66% 2) CES-D sens = 88% spec = 75%	Int.: 39% Ctrl: 29% *P*=NS	Int.: 44% Ctrl: 43% *P*=NS	Int.: 37% Ctrl: 46% (*P*=0.19)
Zung and King, 1983[103]	1) Improvement (decrease) on SDS ≥12	Int.: 66% Ctrl: 35%	NR	NR	NR

Table 4. Screening Outcomes for Depression Studies (continued)

Adequate Inclusion Criteria?	Drop-outs similar in Intervention & Control Groups?	Adequate Randomization & Conceal-ment?	Intention to Treat Analysis?	Internal Validity?	External Validity?	Overall Quality?	Comments
Yes	Yes	Yes	N/A	Good	Good	Good	
Yes	Yes	Yes	Yes	Good	Good	Good	
Yes	Yes	Yes	DK	Good	Good	Good	