

Draft Comparative Effectiveness Review

Number XX

**Nonpharmacological Versus Pharmacological
Treatments for Adult Patients With Major Depressive
Disorder**

Prepared for:

Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. xxx-xx-xxxx

Prepared by:

<Name> Evidence-based Practice Center
<City, State>

Investigators:

First and Last Names, X.X.
First and Last Names, X.X.

AHRQ Publication No. xx-EHCxxx

<Month Year>

This information is distributed solely for the purposes of pre-dissemination peer review. It has not been formally disseminated by the Agency for Healthcare Research and Quality. The findings are subject to change based on the literature identified in the interim and peer-review/public comments and should not be referenced as definitive. It does not represent and should not be construed to represent an Agency for Healthcare Research and Quality or Department of Health and Human Services (AHRQ) determination or policy.

This report is based on research conducted by an Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. XXX-XXXX-XXXXX). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

AHRQ or U.S. Department of Health and Human Services endorsement of any derivative products that may be developed from this report, such as clinical practice guidelines, other quality enhancement tools, or reimbursement or coverage policies may not be stated or implied.

This document is in the public domain and may be used and reprinted without permission except those copyrighted materials that are clearly noted in the document. Further reproduction of those copyrighted materials is prohibited without the specific permission of copyright holders.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact [insert program email address].

<p>None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.</p>

Suggested citation: <Authors>. <Topic in Title Caps>. <Report Series Name> <#>. (Prepared by the <EPC Name> Evidence-based Practice Center under Contract No. <##>.) AHRQ Publication No. XX-EHCXXX-EF. Rockville, MD: Agency for Healthcare Research and Quality. <Month Year>. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of systematic reviews to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. These reviews provide comprehensive, science-based information on common, costly medical conditions, and new health care technologies and strategies.

Systematic reviews are the building blocks underlying evidence-based practice; they focus attention on the strength and limits of evidence from research studies about the effectiveness and safety of a clinical intervention. In the context of developing recommendations for practice, systematic reviews can help clarify whether assertions about the value of the intervention are based on strong evidence from clinical studies. For more information about AHRQ EPC systematic reviews, see www.effectivehealthcare.ahrq.gov/reference/purpose.cfm

AHRQ expects that these systematic reviews will be helpful to health plans, providers, purchasers, government programs, and the health care system as a whole. Transparency and stakeholder input are essential to the Effective Health Care Program. Please visit the Web site (www.effectivehealthcare.ahrq.gov) to see draft research questions and reports or to join an e-mail list to learn about new program products and opportunities for input.

We welcome comments on this systematic review. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to epc@ahrq.hhs.gov.

Richard G. Kronick, Ph.D.
Director
Agency for Healthcare Research and Quality

David Meyers, M.D.
Acting Director
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director, EPC Program
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Aysegul Gozu, M.D., M.P.H.
Task Order Officer
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this project: <Acknowledgments>.

Key Informants

In designing the study questions, the EPC consulted a panel of Key Informants who represent subject experts and end-users of research. Key Informant input can inform key issues related to the topic of the technical brief. Key Informants are not involved in the analysis of the evidence or the writing of the report. Therefore, in the end, study questions, design, methodological approaches and/or conclusions do not necessarily represent the views of individual Key Informants.

Key Informants must disclose any financial conflicts of interest greater than \$10,000 and any other relevant business or professional conflicts of interest. Because of their role as end-users, individuals with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any conflicts of interest.

The list of Key Informants who participated in developing this report follows:

<Name>
<Place>
<City>, <ST>

Technical Expert Panel

In designing the study questions and methodology at the outset of this report, the EPC consulted several technical and content experts. Broad expertise and perspectives were sought. Divergent and conflicted opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Therefore, in the end, study questions, design, methodologic approaches, and/or conclusions do not necessarily represent the views of individual technical and content experts.

Technical Experts must disclose any financial conflicts of interest greater than \$10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

The list of Technical Experts who participated in developing this report follows:

<Name>
<Place>
<City>, <ST>

<Name>
<Place>
<City>, <ST>

Peer Reviewers

Prior to publication of the final evidence report, EPCs sought input from independent Peer Reviewers without financial conflicts of interest. However, the conclusions and synthesis of the scientific literature presented in this report does not necessarily represent the views of individual reviewers.

Peer Reviewers must disclose any financial conflicts of interest greater than \$10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential non-financial conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential non-financial conflicts of interest identified.

The list of Peer Reviewers follows:

<Name>
<Place>
<City>, <ST>

<Name>
<Place>
<City>, <ST>

Nonpharmacological Versus Pharmacological Treatments for Adult Patients With Major Depressive Disorder

Structured Abstract

Objective. To compare the benefits and harms of pharmacological, psychological, complementary and alternative medicine, and exercise treatment options as first-line interventions for adult outpatients with acute phase major depressive disorder (MDD), and as second-line interventions for patients with MDD who did not achieve remission after a first treatment attempt with second-generation antidepressants (SGAs).

Data sources. MEDLINE® (via PubMed), EMBASE, the Cochrane Library, AMED (Allied and Complementary Medicine Database), PsycINFO, and CINAHL (Cumulative Index to Nursing and Allied Health Literature) from January 1, 1990, through May 2, 2014.

Review method. Two investigators independently selected, extracted data from, and rated risk of bias of studies. We graded strength of evidence based on established guidance.

Results. Forty-five trials met inclusion criteria. Across all interventions, we graded the strength of evidence as moderate for only two comparisons, namely SGAs compared with cognitive behavioral therapy (CBT) and St. John's wort. Results indicate that CBT and St. John's wort have levels of effectiveness regarding symptomatic relief similar to those of SGAs. The overall risk for adverse events or discontinuation of treatment because of adverse events, however, is lower for these non-SGA therapies.

Our confidence in findings from comparisons of SGAs with the remaining treatment options was low or insufficient, indicating that these bodies of evidence had major or unacceptable deficiencies. Nevertheless, for most comparisons the overall findings indicated no statistically significant differences in effectiveness but lower risk of adverse events for nonpharmacological treatment options. Exceptions are omega-3-fatty acids, which appear to have lower effectiveness than SGAs; and the combination treatment of SGAs and acupuncture which appears to have better effectiveness than SGA monotherapy. Our confidence in these findings, however, is low and results have to be interpreted cautiously.

For second-line therapies (i.e. therapy for patients with MDD who did not achieve remission after a first treatment attempt with SGAs), although evidence is limited, no clear benefit emerges to suggest either switching to a particular SGA or to cognitive therapy or augmenting with a particular medication or cognitive therapy.

Conclusions. Overall, the available evidence does not support the superiority of SGAs over CBT and St. John's wort as first line treatments for patients with moderate to severe MDD. Given no clear differences in beneficial treatment effect among treatment options, the choice of the initial treatment of MDD should be strongly based on patient preferences and the feasibility (e.g., costs, likely adherence) following a discussion of the advantages and disadvantages of each treatment option. Differences with respect to adverse events, personal engagement, and costs

may be taken into consideration for the choice of a first-line treatment. Such shared and informed decisionmaking might enhance treatment adherence and improve treatment outcomes for patients with MDD, especially because treatment continuity is one of the main challenges in treating such patients.

Contents

Executive Summary	ES-1
Introduction	1
Background	1
Clinical and Methodological Issues	1
Scope and Key Questions	6
Scope of This Review	6
Key Questions	8
Organization of This Report	8
Methods	9
Topic Refinement and Review Protocol	9
Literature Search Strategy.....	9
Search Strategy	9
Inclusion and Exclusion Criteria.....	10
Study Selection	14
Data Extraction	15
Risk-of-Bias Assessment of Individual Studies.....	15
Data Synthesis.....	15
Meta-Analysis of Direct Comparisons.....	16
Network Meta-Analyses	17
Strength of Evidence of the Body of Evidence.....	17
Applicability	18
Peer Review and Public Commentary	18
Results	19
Introduction.....	19
Results of Literature Searches	19
Description of Included Trials	19
KQ 1: First-Line Therapy: Second-Generation Antidepressants Compared With Nonpharmacologic Therapies	22
Key Points: Second-Generation Antidepressants Compared With Psychological Interventions	22
Detailed Synthesis: KQ 1	25
Impact of Severity on Various Outcomes	48
KQ 2: Second-Line Therapy—Switching or Augmenting Strategies Involving a Second- Generation Antidepressant.....	51
Key Points: Switching Strategies.....	52
Key Points: Augmentation Strategies	52
Key Points: Severity as a Moderator of Treatment Effectiveness of Second-line Therapies	52
Detailed Synthesis: KQ 2.....	52
KQ 3: Comparative Risks of Treatment Harms.....	58
Overview	58
Key Points.....	59
KQ 4: Comparative Benefits and Risks of Harms for Selected Subgroups.....	82
Overview	82
Key Points: Common Accompanying Psychiatric Symptoms.....	83

Key Points: Age	83
Key Points: Sex	83
Key Points: Race or Ethnicity	83
Discussion.....	86
Key Findings and Strength of Evidence	86
Comparative Benefits and Harms of Treatment Options for Initial Treatment of Patients With Major Depressive Disorder	86
Comparative Benefits and Harms as a Function of Baseline Depressive Severity.....	92
Comparative Benefits and Harms of Alternative Pharmacologic and Nonpharmacologic Options for Patients Whose Depression Did Not Achieve Remission Following Initial Treatment With a Second-Generation Antidepressant.....	93
Comparative Benefits and Harms of Second-line Therapies as a Function of Baseline Depressive Severity	94
Findings in Relationship to What is Already Known	95
Applicability	97
Implications for Clinical and Policy Decisionmaking	99
Limitations of the Comparative Effectiveness Review Process	100
Limitations of the Evidence Base	100
Research Gaps.....	101
Conclusions.....	102
References	104

Tables

Table A. Inclusion/exclusion criteria	ES-3
Table 1. SGAs: Usual dosing range and frequency of administration for adults	4
Table 2. Inclusion/exclusion criteria.....	11
Table 3. Categories of depressive severity ⁶²	11
Table 4. Eligible interventions for major depressive disorders	12
Table 5. Possible comparisons of eligible interventions with second-generation antidepressants.....	13
Table 6. Outcomes rated as critical or important for decisionmaking.....	14
Table 7. Definition of strength of evidence grades.....	18
Table 8. Abbreviations and full names of mental health and other assessment tools.....	19
Table 9. Reported sources of funding for included trials.....	20
Table 10. Number of included trials by type of comparison	25
Table 11. Second-generation antidepressants versus cognitive behavioral therapy: Trial characteristics, main outcomes, and risk of bias ratings	27
Table 12. Second-generation antidepressants versus interpersonal psychotherapy: Trial characteristics, main outcomes, and risk of bias ratings	32
Table 13. Second-generation antidepressants versus psychodynamic therapies: Trial characteristics, main outcomes, and risk of bias ratings	34
Table 14. Second-generation antidepressants versus third-wave cognitive behavioral therapy: Trial characteristics, main outcomes, and risk of bias ratings of trials.....	35
Table 15. Second-generation antidepressant versus acupuncture: Study characteristics, main outcomes, and risk of bias ratings.....	36
Table 16. Second-generation antidepressants versus omega-3 fatty acids: Study characteristics, main outcomes, and risk of bias ratings.....	38

Table 17. Second-generation antidepressants versus SAME: Study characteristics, main outcomes, and risk of bias ratings.....	39
Table 18. Second-generation antidepressants versus St. John’s wort: Trial characteristics, main outcomes, and risk of bias ratings.....	40
Table 19. Second-generation antidepressants versus exercise: Study characteristics, main outcomes, and risk of bias ratings.....	43
Table 20. SGAs versus psychological interventions by depression severity: Trial characteristics, main outcomes, and risk of bias ratings of trials	45
Table 21. Comparative efficacy for critical efficacy outcomes by baseline severity for psychological interventions and second-generation antidepressants.....	49
Table 22. SGAs versus SAME by depression severity: Trial characteristics, main outcomes, and risk of bias ratings of trials.....	50
Table 23. Comparative efficacy for critical efficacy outcomes by baseline severity for complementary and alternative interventions, exercise, and second-generation antidepressants	51
Table 24. Number of included trials by type of comparison	53
Table 25. Second-generation antidepressant switch versus another second-generation switch strategy: Trial characteristics, main outcomes, and risk of bias ratings	55
Table 26. Second-generation antidepressant augmentation versus another second-generation augmentation strategy: Trial characteristics, main outcomes, and risk of bias ratings.....	57
Table 27. Number of trials for each comparison of interest	58
Table 28. Number of included trials for all subgroups by type of comparison	82
Table 29. Second-generation antidepressants versus nonpharmacologic therapies in subgroups: Trial characteristics, main outcomes, and risk of bias ratings	84
Table 30. Summary of findings with strength of evidence: Comparative benefits and harms of second-generation antidepressants and other treatment options as an initial choice for the treatment of patients with major depressive disorders.....	87
Table 31. Summary of findings with strength of evidence: Variation in effectiveness by severity for second-generation antidepressants compared with other treatments for patients with major depressive disorder.....	93
Table 32. Summary of findings with strength of evidence: Comparative benefits of second-generation antidepressants and other treatment options as an initial choice for the treatment of major depressive disorder (KQ 2a).....	94
Table 33. Summary of findings with strength of evidence: Second-generation antidepressants compared with other treatments for major depressive disorder: Does effectiveness vary by severity? (KQ 2b).....	95

Figures

Figure A. Relative risks of response of SGAs compared with other eligible interventions	ES-7
Figure B. Relative risks of discontinuation because of adverse events of SGAs compared with other eligible interventions	ES-8

Figure 1. Phases of treatment for major depression.....	3
Figure 2. Analytic framework for treatment of major depressive disorder	7
Figure 3. PRISMA diagram for treatment of major depressive disorders	20
Figure 4. Network of trials included for network meta-analyses.....	21
Figure 5. Relative risks of response of SGAs compared with other eligible interventions	24
Figure 6. Relative risks of remission of SGAs compared with other eligible interventions.	24

Figure 7. SGA versus cognitive behavioral therapy: Remission	29
Figure 8. SGA versus cognitive behavioral therapy: Response.....	30
Figure 9. SGA versus cognitive behavioral therapy: Change in HAM-D-17.....	30
Figure 10. SGAs versus St. John’s Wort: Response.....	41
Figure 11. SGA versus St. John’s Wort: Remission.....	42
Figure 12. SGA versus St. John’s Wort: Change in HAM-D-17.....	42
Figure 13. Relative overall risk of harms of SGAs compared with other eligible interventions..	61
Figure 14. Relative risks of overall discontinuation of SGAs compared with other eligible interventions.....	62
Figure 15. Relative risks of discontinuation because of adverse events rates of SGAs compared with other eligible interventions	63
Figure 16. Overall discontinuation rates comparing second-generation antidepressants with psychological interventions as a class.....	64
Figure 17. Discontinuation rates because of lack of efficacy comparing second-generation antidepressants with psychological interventions as a class	65
Figure 18. Discontinuation rates because of adverse events comparing second-generation antidepressants with psychological interventions as a class	65
Figure 19. Overall discontinuation rates comparing second-generation antidepressants with cognitive behavioral therapies	67
Figure 20. Discontinuation rates because of lack of efficacy comparing second-generation antidepressants with cognitive behavioral therapies.....	67
Figure 21. Discontinuation rates because of adverse events comparing second-generation antidepressants with cognitive behavioral therapy	67
Figure 22. Meta-analysis of overall discontinuation rates comparing second-generation antidepressants with combination of acupuncture and SGA	71
Figure 23. Meta-analysis of overall risk for adverse events comparing second-generation antidepressants with St. John’s wort.....	73
Figure 24. Meta-analysis of overall discontinuation rates comparing second-generation antidepressants with St. John’s wort.....	74
Figure 25. Meta-analysis of discontinuation because of adverse events comparing second-generation antidepressants with St. John’s wort.....	74
Figure 26. Meta-analysis of discontinuation rates because of lack of efficacy comparing second-generation antidepressants with St. John’s wort.....	75
Figure 27. Meta-analysis of suicidality data comparing second-generation antidepressants with any psychological treatment	77
Figure 28. Meta-analysis of serious adverse events data comparing second-generation antidepressants with St. John’s wort.....	81

Appendixes

Appendix A. Search Strategy

Appendix B. Cochrane Depression, Anxiety and Neurosis Group (CCDAN) Topic List: Intervention – Psychological Therapies

Appendix C. Studies Excluded at the Full-Text Review Stage

Appendix D. Risk of Bias Evaluations

Appendix E. Summary of Findings Tables

Appendix F. Studies Included in Network Meta-Analyses

Executive Summary

Background

Condition and Therapeutic Strategies

Major depressive disorder (MDD)¹ is the most prevalent and disabling form of depression, affecting more than 16 percent of U.S. adults (lifetime).² MDD can be characterized as mild, moderate, or severe based on symptom severity, functional impairment, and level of patient distress;¹ in clinical trials, these distinctions are typically made by scores on a depressive rating instrument.³ Approximately one-third of patients with MDD are severely depressed,⁴ which is associated with a harder to treat depression.⁵

In any given year, nearly 7 percent of the U.S. adult population (approximately 17.5 million people in 2014) experiences an episode of MDD that warrants treatment.² Most patients receiving care obtain treatment in primary care settings,⁶ where second-generation antidepressants (SGAs) are the most commonly prescribed agents.⁷ Nonetheless, primary care patients and clinicians may prefer other options (or at least want to be able to consider them). These include psychological interventions, complementary and alternative medicine (CAM) options, and exercise.

The psychological interventions used to treat depressed patients include acceptance and commitment therapy, cognitive therapy (CT), cognitive behavioral therapy (CBT), interpersonal therapy, and psychodynamic therapies. All may have different customary lengths of treatment.

Commonly used CAM interventions for the treatment of patients with MDD include acupuncture, meditation, omega-3 fatty acids, S-adenosyl-L-methionine (SAME), St. John's wort, and yoga.

Exercise covers a broad range of activities; they can be done over varying durations of time and singly, in classes, or in informal groups. Compared with no treatment, exercise has a moderate clinical benefit for patients with MDD.⁸

Nevertheless, about 40 percent of patients treated with SGAs do not respond to initial treatment; approximately 70 percent do not achieve remission during the first-line treatment.⁹ Those who do not achieve remission following initial pharmacological treatment require a different treatment strategy. Accordingly, various other interventions—such as medication combinations, psychotherapy, or CAM treatments—are important options for patients and clinicians.

Scope and Key Questions

This review for the Evidence-based Practice Center (EPC) program of the U.S. Agency for Healthcare Research and Quality (AHRQ) examines the evidence base for primary care management of MDD for the first two treatment attempts, after which primary care clinicians would consider referral to or consultation by a mental health professional. The specific Key Questions (KQs) are listed below:

- KQ 1a. In adult patients with MDD who are undergoing an initial treatment attempt, what is the effectiveness of second-generation antidepressant (SGA) monotherapy compared with the effectiveness of either nonpharmacological monotherapy or combination therapy (involving nonpharmacological treatments with or without an SGA)?
- KQ 1b. Does comparative treatment effectiveness vary by MDD severity?

KQ 2a. In adult patients with MDD who did not achieve remission following an initial adequate trial with one SGA, what is the comparative effectiveness of second-line therapies*?

* Any comparison that involves an eligible intervention (whether as a monotherapy or a combination therapy) and compares an intervention to one involving an SGA is eligible.

KQ 2b. Does comparative treatment effectiveness vary by MDD severity?

KQ 3a. In adult patients with MDD, what are the comparative risks of harms of these treatment options:

(1) for those undergoing an initial treatment attempt or

(2) for those who did not achieve remission following an initial adequate trial with an SGA?

KQ 3b. Do the comparative risks of treatment harms vary by MDD severity?

KQ 4. Do the benefits and risks of harms of these treatment options differ by subgroups of patients with MDD defined by common accompanying psychiatric symptoms (coexisting anxiety, insomnia, low energy, or somatization) or demographic characteristics (age, sex, race, or ethnicity)?

Methods

Literature Search Strategy

We searched MEDLINE[®] (via PubMed), EMBASE, the Cochrane Library, AMED (Allied and Complementary Medicine Database), PsycINFO, and CINAHL (Cumulative Index to Nursing and Allied Health Literature) from January 1, 1990, through May 2, 2014. We used a combination of medical subject headings (MeSH) and title and abstract key keywords, focusing on terms to describe the relevant population and interventions of interest. We limited the electronic searches to English-, German-, and Italian-language and human-only studies.

In addition, we manually searched reference lists of pertinent reviews, included trials, and background articles and searched for “gray literature” relevant to this review following guidance from the *Methods Guide for Effectiveness and Comparative Effectiveness Reviews* for these steps.¹⁰

Inclusion and exclusion criteria are presented in Table A.

Table A. Inclusion/exclusion criteria

	Inclusion	Exclusion
Population	Adult (18 years or older) outpatients of all races and ethnicities with MDD during either an initial treatment attempt or a second treatment attempt in patients who did not remit following an initial adequate trial with an SGA	<ul style="list-style-type: none">• Children under age 18• Patients with perinatal depression, seasonal affective disorder, psychotic depression, or treatment-resistant depression (i.e., two or more treatment failures)
Interventions	<p>Second-Generation Antidepressants:</p> <ul style="list-style-type: none">• Bupropion• Citalopram• Desvenlafaxine• Duloxetine• Fluoxetine• Escitalopram• Fluvoxamine• Levomilnacipran• Mirtazapine• Nefazodone• Paroxetine• Sertraline• Trazodone• Venlafaxine• Vilazodone• Vortioxetine <p>Common Depression-Focused Psychotherapies:</p> <ul style="list-style-type: none">• Acceptance and commitment therapy• Cognitive and behavioral approaches• Interpersonal therapy• Psychodynamic and attachment-based approaches <p>Complementary and Alternative Medicines:</p> <ul style="list-style-type: none">• Acupuncture• Meditation (e.g., mindfulness-based stress reduction)• Omega-3 fatty acids• S-adenosyl-L-methionine (SAME)• St. John's wort (<i>Hypericum</i>)• Yoga <p>Exercise:</p> <ul style="list-style-type: none">• Any formal exercise program <p>Other pharmacotherapies for combination or augmentation:</p> <ul style="list-style-type: none">• Atypical antipsychotics (aripiprazole, asenapine maleate, clozapine, iloperidone, lurasidone, olanzapine, paliperidone, quetiapine, risperidone, ziprasidone)• Psychostimulants (amphetamine-dextroamphetamine, armodafinil, dexamfetamine, dextroamphetamine, lisdexamfetamine, methylphenidate, modafinil)• Buspirone• Levothyroxine (T4)• Lithium• Pindolol• Triiodo-thyronine (T3)	Ineligible interventions

Table A. Inclusion/exclusion criteria (continued)

	Inclusion	Exclusion
Control interventions	<p>For all populations of interest (i.e., KQ 1, KQ 3, and KQ 4):</p> <ul style="list-style-type: none"> • SGAs vs. psychotherapies • SGAs vs. CAM • SGAs vs. exercise • SGAs vs. SGA + psychotherapies • SGAs vs. SGA + CAM • SGAs vs. SGA + exercise • SGAs vs. combinations of eligible interventions <p>In addition for populations who did not achieve remission following an initial adequate trial with an SGA (i.e., KQ 2, KQ 3, and KQ 4):</p> <ul style="list-style-type: none"> • SGA switch^a vs. SGA switch^a • SGA switch^a vs. nonpharmacologic • SGA switch^a vs. SGA augmentation^b • SGA augmentation^b vs. SGA augmentation^b • SGA augmentation^b vs. nonpharmacologic <p>In addition for network meta-analyses:</p> <ul style="list-style-type: none"> • Placebo or other inactive control • Comparisons of eligible interventions without an SGA arm 	Ineligible interventions, such as placebo arms
Outcomes	<ul style="list-style-type: none"> • Benefits: response, remission, speed of response, speed of remission, relapse, quality of life, functional capacity, reduction of suicidality, reduction of hospitalization • Harms: overall adverse events, withdrawals because of adverse events, serious adverse events, specific adverse events (including hyponatremia, seizures, suicidality, hepatotoxicity, weight gain, gastrointestinal symptoms, sexual side effects), withdrawals because of specific adverse events, or drug interactions (pharmacologic and complementary and alternative treatments) 	<ul style="list-style-type: none"> • Studies that do not include at least one of the outcomes listed under the inclusion criteria
Timing of intervention	<ul style="list-style-type: none"> • No limitations 	Not applicable
Publication language	English, German, Italian	All other languages
Study design	<ul style="list-style-type: none"> • Original research • Eligible study designs include: • For efficacy/effectiveness <ul style="list-style-type: none"> ○ RCTs ○ SRs and meta-analyses • In addition for harms <ul style="list-style-type: none"> ○ Nonrandomized controlled trials ○ Prospective controlled cohort studies ○ Retrospective controlled cohort studies ○ Case-control studies ○ Nonrandomized studies must have a minimum sample size of 500 participants 	<ul style="list-style-type: none"> • Case series • Case reports • Nonsystematic reviews • Studies without a control group • Nonrandomized studies with fewer than 500 participants • Post hoc or secondary analyses • Pooled studies

Table A. Inclusion/exclusion criteria (continued)

Inclusion		Exclusion
Publication type	Any publication reporting primary data	Publications not reporting primary data

CAM = complementary and alternative medicine; KQ = Key Question; MDD = major depressive disorder; PICOTS = patients, interventions, comparators, outcomes, timing, setting; RCT = randomized controlled trial; SAME = S-adenosylmethionine; SGA = second-generation antidepressant; SR = systematic review; vs. = versus.

Two trained research team members independently reviewed all titles, abstracts, and eligible full-text articles. We designed, pilot tested, and used a structured data abstraction form to ensure consistency of data abstraction. Trained reviewers initially abstracted data from each study. A senior reviewer then read each abstracted article and evaluated the completeness and accuracy of the data abstraction. We resolved discrepancies by consensus or by involving a third, senior reviewer.

Risk of Bias Assessment of Individual Studies

To assess the risk of bias of studies, we used definitions based on AHRQ guidance.¹¹ We rated the risk of bias for each relevant outcome of a study as low, moderate, or high. To determine risk of bias in a standardized way, we used the Cochrane Risk of Bias tool to appraise RCTs.¹² Two independent reviewers assigned risk of bias ratings. They resolved any disagreements by discussion and consensus or by consulting a third, independent party.

Data Synthesis

Throughout this review we synthesized the literature qualitatively. When data were sufficient, we augmented findings with quantitative analyses.

For meta-analyses, we used random- (DerSimonian-Laird) and fixed-effects models to estimate comparative effects. We assessed statistical heterogeneity in effects between studies by calculating the chi-squared statistic and Cochran's q . We used the I^2 statistic to estimate the magnitude of heterogeneity. We examined potential sources of heterogeneity using sensitivity analysis or analysis of subgroups. We assessed publication bias by checking study registries and using funnel plots and Kendall's tests. However, given the small number of component studies in our meta-analyses, these tests have low sensitivity to detect publication bias.

Because of the dearth of studies directly comparing interventions of interest, we a priori planned network meta-analyses. Our outcome measure of choice was the rate of response on the Hamilton Depression Rating Scale (HAM-D, defined as at least a 50 percent improvement of scores from baseline). We included all placebo- and active-controlled RCTs detected through our searches that were homogenous in study populations and outcome assessments and were part of a connected network. We employed a hierarchical frequentist approach using random-effects models^{13,14}

Strength of the Body of Evidence

We graded the strength of evidence (SOE) based on AHRQ guidance established for the EPC program.¹⁵ This approach incorporates five key domains: risk of bias, consistency, directness, precision, and reporting bias. Grades (high, moderate, low, insufficient) reflect the strength of the body of evidence for a specific outcome on the comparative benefits and harms of the interventions in this review. During the protocol development, we asked the Technical Expert Panel (TEP) and the Key Informants to rank the relative importance of outcomes following a

process proposed by the GRADE Working Group.¹⁶ We graded only those outcomes that TEP and Key Informants deemed as important or critical for decisionmaking.

Applicability

We assessed applicability of the evidence following guidance from the *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*.¹⁷ We used the PICOTS ((populations, interventions, comparators, outcomes, timing, settings) framework to explore factors that affect applicability.

Results

We document the outputs of our literature searches and then describe included trials in general terms. We also summarize findings by KQ, dealing with KQ 1 and KQ 3 (respectively, benefits and harms) together and organized by intervention comparisons.

Results of Literature Searches

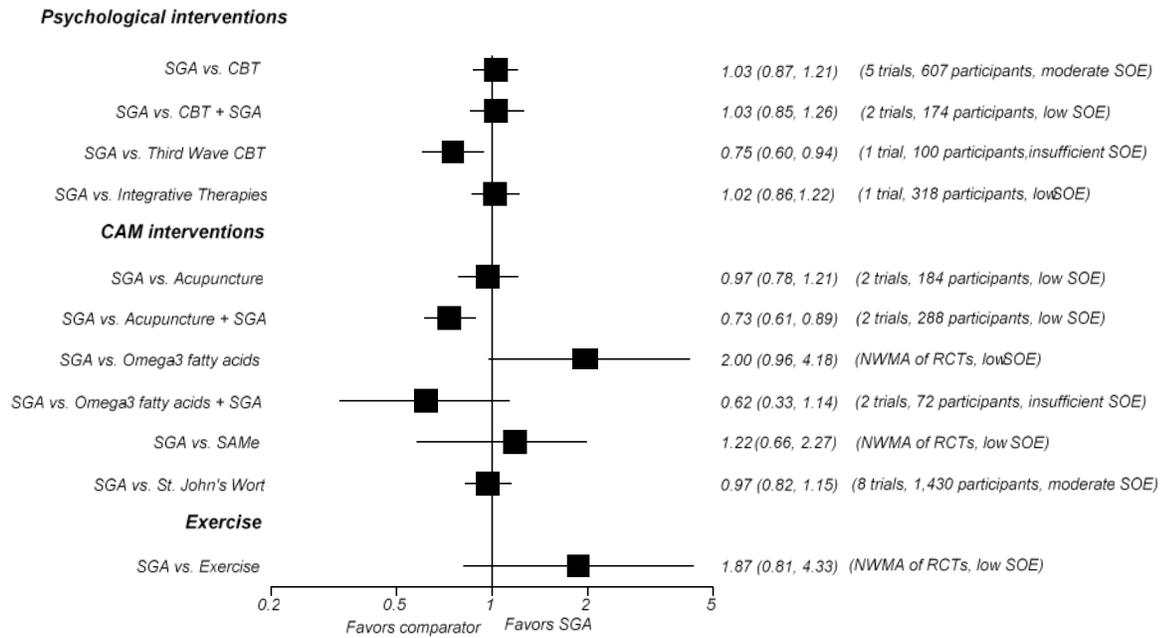
Our search strategies identified 7,377 possible articles. We excluded 6,947 references following independent dual title and abstract review and another 375 references at the full-text review stage. Reasons for exclusion were based on eligibility criteria. Overall, we included 45 trials reported in 55 published articles. Of these, 42 trials pertained to KQ 1a and five to KQ 1b. Two trials pertained to KQ 2a, and no trials were identified for KQ 2b. In addition, of the 45 trials, 44 trials pertained to KQ 3a and one to KQ 3b; three pertained to KQ 4.

We included data from 97 additional published trials and data from 27 unpublished trials for network meta-analyses. These trials addressed comparisons of interventions of interest that did not meet eligibility criteria for this report; they did, however, provide common comparators that we could use for network meta-analyses.

Comparative Benefits and Harms of Treatment Options for Initial Treatment of Patients With Major Depressive Disorder

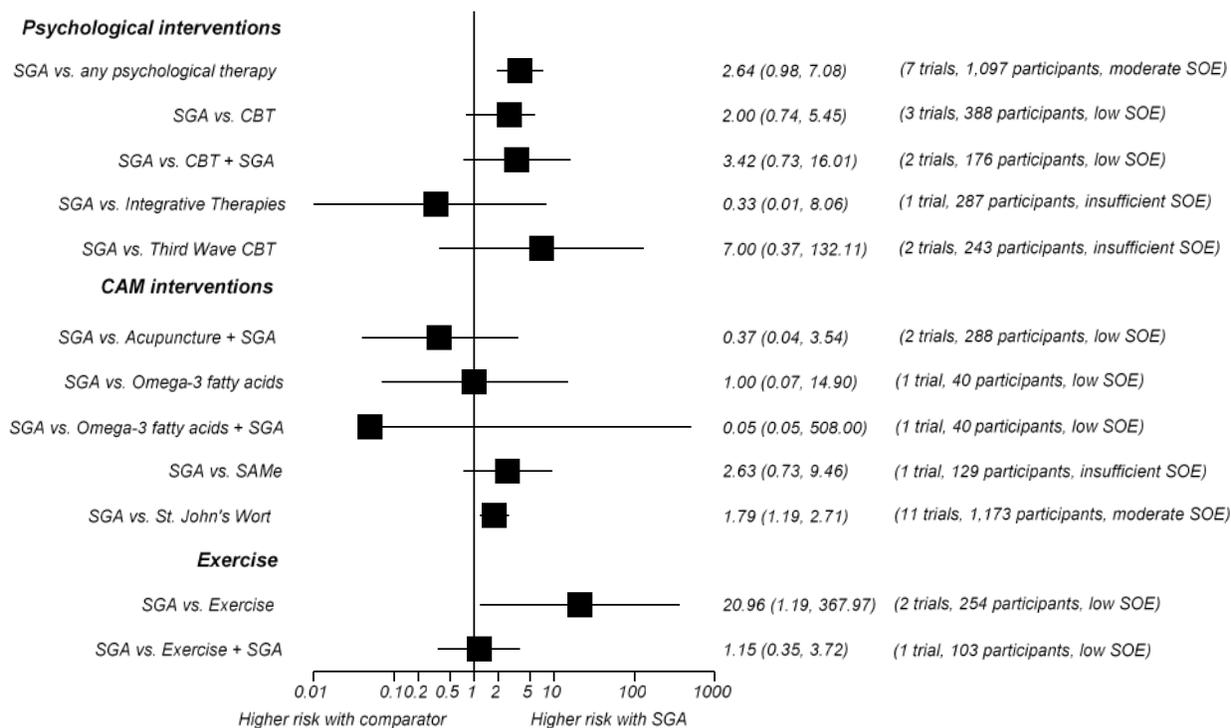
In all, 43 trials comparing SGAs with nonpharmacological treatment options for MDD provided direct evidence on acute-phase outcomes. Study durations ranged from 4 to 96 weeks. Most patients suffered from moderate to severe major depression. Many of the available trials had serious methodological limitations; few trials reported information on quality of life or functional capacity. Figures A and B provide graphical overviews of response rates and discontinuation rates because of adverse events (respectively) of SGAs compared with psychological interventions, CAM therapies, and exercise.

Figure A. Relative risks of response of SGAs compared with other eligible interventions



CAM = complementary and alternative medicine; CBT = cognitive behavioral therapy; KQ = Key Question; NWMA = Network Meta-analysis; SAmE = S-adenosyl-L-methionine; SGA = second-generation antidepressant; SOE = strength of evidence; vs. = versus.

Figure B. Relative risks of discontinuation because of adverse events of SGAs compared with other eligible interventions



CAM = complementary and alternative medicine; CBT = cognitive behavioral therapy; KQ = Key Question; SAMe = S-adenosyl-L-methionine; SGA = second-generation antidepressant; SOE = strength of evidence; vs. = versus.

Second-Generation Antidepressants Compared With Psychological Interventions

Second-Generation Antidepressant Versus Cognitive Behavioral Therapy

We identified 11 trials (1,393 participants) of interventions categorized by the CCDAN (Appendix B) as cognitive behavioral therapies. Six trials employed CBT, four used CT, and one each used problem-solving therapy and rational emotive behavior therapy. Three trials included a combination SGA plus CBT arm. Overall, SGAs and CBT monotherapies led to similar rates of response (moderate SOE), remission (low SOE), and overall discontinuation (moderate SOE). Patients receiving CBT had a 3 times lower risk of discontinuing treatment because of adverse events than those on SGAs (low SOE).

Adding CBT to SGA did not show any benefit in response and had similar rates of discontinuation due to adverse events (low SOE). The evidence was insufficient to conclude about differences in functional capacity, quality of life, and risk of adverse events.

Second-Generation Antidepressants Versus Integrative Therapies

We identified four trials (872 participants) that compared SGA monotherapy with interpersonal psychotherapy alone. One trial also examined the effect of adding interpersonal psychotherapy to the SGA regimen.

SGAs and interpersonal psychotherapy did not lead to statistically different response, remission, and discontinuation rates (low SOE). The combination of SGA and interpersonal psychotherapy had 25 percent higher remission rates than SGA monotherapy (low SOE). Discontinuation rates were similar between SGA monotherapy and the combination of SGA and interpersonal therapy (low SOE). The evidence was insufficient to conclude anything about differences in functional capacity, quality of life, and risk of serious adverse events.

Second-Generation Antidepressants Versus Psychodynamic Therapies (PSYD)

Four trials (570 participants) compared SGA monotherapy with PSYD. One trial also examined the effect of adding PSYD to the SGA regimen. SGA monotherapy and PSYD monotherapy did not lead to statistically different rates of remission (low SOE) and improvements in functional capacity (low SOE). SGAs and PSYD also led to similar rates of overall discontinuation over 48 weeks (low SOE) and 96 weeks of followup (low SOE).

Adding PSYD to SGAs led to overall discontinuation rates that did not differ statistically for those patients receiving SGA monotherapy (low SOE). Suicidality did not differ statistically for patients on SGAs, PSYD, or a combination of the two (low SOE). The evidence was insufficient to conclude anything about differences in functional capacity, quality of life, risk of serious adverse events, and discontinuation due to adverse events.

Second-Generation Antidepressants Versus Third-Wave Cognitive Behavioral Therapy

One randomized trial (272 participants) compared an SGA with behavioral activation, a type of third-wave behavioral therapy. The evidence was insufficient to conclude about differences in response, remission, overall discontinuation, discontinuation of treatment because of adverse events, and suicidality.

Severity as a Moderator of Treatment Effectiveness

Four trials yielded insufficient evidence to determine whether the comparative effectiveness of SGAs versus any psychological treatments changes as a function of MDD severity.

Second-Generation Antidepressants Compared With CAM Interventions

Second-Generation Antidepressants Versus Acupuncture

Three trials (263 participants) compared an SGA with either full-body or scalp electroacupuncture. For treatment response, pooled results from direct comparisons and indirect comparisons using network meta-analysis demonstrated no differences in effectiveness (low SOE). Two trials (237 participants) examined the effect of adding acupuncture to the SGA treatment regimen. Acupuncture in combination with an SGA had 37 percent higher response rates than SGAs alone (low SOE) but did not differ statistically in remission rates (low SOE). The combination of SGAs and acupuncture led to overall discontinuation rates (moderate SOE) and discontinuation rates because of adverse events (low SOE) that did not differ statistically from those among patients on SGA monotherapy.

The evidence was insufficient to conclude anything about differences in functional capacity, quality of life, and risk of harms. Indirect evidence, however, indicates lower adverse event rates for acupuncture than SGAs.

Second-Generation Antidepressants Versus Omega-3 Fatty Acids

One trial (40 participants) compared an SGA with omega-3-fatty acids. Network meta-analysis indicated a response rate that was twice as high for patients treated with SGAs than for those receiving omega-3-fatty acids (low SOE). SGAs and omega-3 fatty acids did not lead to significantly different rates of overall discontinuation (low SOE) or discontinuation because of adverse events (low SOE). Evidence was insufficient to draw conclusions about differences in remission, functional capacity, quality of life, and overall risks of adverse events.

Two trials (72 participants) examined the effect of adding omega-3-fatty acids to the SGA regimen. Because of methodological shortcomings, the evidence was insufficient to draw conclusions.

Second-Generation Antidepressants Versus S-Adenosylmethionine (SAME)

One trial (129 participants) compared an SGA with SAME. Network meta-analysis indicated response rates that did not differ statistically for patients on SGAs or SAME (low SOE). Overall discontinuation rates were also similar among patients treated with SGAs or SAME.

The evidence was insufficient to conclude about differences in remission, functional capacity, quality of life, and risk of harms.

Second-Generation Antidepressants Versus St. John's Wort

We identified 12 trials (1,430 participants) comparing SGAs with St. John's wort monotherapy. Meta-analysis of eight trials (1,179 participants) indicated similar response rates between SGAs and St. John's wort. Meta-analysis of four trials (683 participants) demonstrated similar remission rates for the two treatments (moderate SOE). SGAs led to 29 percent higher rates of overall discontinuation (moderate SOE) and 79 percent higher discontinuation rates because of adverse events (moderate SOE) than did St. John's wort. The overall risk of adverse events was 19 percent higher among patients receiving SGAs than those receiving St. John's wort (moderate SOE). In contrast, the risk of serious adverse events did not differ between patients receiving SGAs or St. John's wort (low SOE).

The evidence was insufficient to conclude anything about differences in functional capacity, quality of life, and risk of serious adverse events.

Second-Generation Antidepressants Versus Yoga or Meditation

We identified no eligible trial that compared an SGA with yoga or meditation.

Severity as a Moderator of Treatment Effectiveness

One trial yielded insufficient evidence to determine whether the comparative effectiveness of SGAs versus any CAM treatments changes as a function of MDD severity.

Second-Generation Antidepressants Compared With Exercise

Two trials (358 participants) compared an SGA with aerobic exercise. One trial also examined the effects of adding exercise to the SGA regimen. Rates of remission and discontinuation did not statistically differ for patients treated with SGAs and patients treated with

exercise monotherapy (low SOE). Estimates based on network meta-analysis indicate no significant difference in response for patients treated with SGAs and those treated with exercise (low SOE). Discontinuation rates because of adverse events were 20 times higher for patients on SGAs than for those on exercise (low SOE).

The combination treatment of SGAs and exercise led to remission, overall discontinuation rates and discontinuation rates because of adverse events that did not differ statistically from those among patients receiving SGA monotherapy (low SOE).

Second-Line Therapy: Comparative Effectiveness of Switching or Augmenting Strategies Involving a Second-Generation Antidepressant

Switch: Second-Generation Antidepressant Versus Second-Generation Antidepressant

Results from two direct comparisons involving 1,123 patients who were switched to different SGAs indicate no substantial differences in response rates between SGAs (moderate SOE). Results from one direct comparison involving 727 patients indicate no substantial difference in remission rates or in the decrease in depressive severity between SGAs (low SOE).

Switch: Second-Generation Antidepressant Versus Cognitive Therapy

Results from one direct comparison involving 122 patients who were assigned to switch to a different SGA or to CT indicate no substantial differences in rates of response or remission or in the decrease in depressive severity (low SOE).

Switch: Second-Generation Antidepressant Versus CAM or Exercise

We did not find any eligible switch evidence comparing an SGA strategy with either CAM or exercise.

Augment: Second-Generation Antidepressant Versus Second-Generation Antidepressant

Results from one direct comparison involving 565 patients indicate no substantial differences in rates of response or remission between SGAs (low SOE). However, results from one direct comparison involving 565 patients indicate a greater decrease in depressive severity after adding bupropion than buspirone (low SOE).

Augment: Second-Generation Antidepressant Versus Cognitive Therapy

Results from one direct comparison involving 182 patients whose treatment was augmented with a second medication versus augmented with CT indicate no substantial differences in rates of response or remission or in the decrease in depressive severity (low SOE).

Effect of Depressive Severity on the Comparative Effectiveness of Second-Line Therapies

One industry-supported secondary analysis involving 396 patients found an insignificant trend toward differences in remission rates for those with severe depression. By contrast, a

second secondary analysis involving 727 patients, which was government funded, found that having mild or moderate rather than severe depression did not change the likelihood of remitting after treatment with one versus another SGA (insufficient evidence).

Comparative Benefits and Risks of Harms for Selected Subgroups

No trials were specifically designed to assess differences in our specified subgroups. Overall, only three trials addressing a subgroup of interest met the criteria for inclusion: one in subgroups defined by common accompanying psychiatric symptoms and two subgroups defined by demographic characteristics. For common accompanying psychiatric symptoms, SGAs produced slightly higher remission rates than interpersonal psychotherapy in patients with a comorbid anxiety disorder but not in those without co-occurring anxiety (insufficient SOE). We had no evidence for any other common accompanying symptoms (insomnia, low energy, or somatization).

For subgroups defined by demographic characteristics, we included two trials. In one trial conducted in older adults, SGAs and St. John's wort led to similar response rates and discontinuation rates because of adverse events (low SOE). The other trial included only minority (predominantly black and Latina) women and showed similar reduction in depressive symptoms between SGAs and CBT (insufficient SOE). We did not identify any trials assessing differences between men and women in efficacy or harms (insufficient SOE).

No trials at all addressed efficacy or harms in selected subgroups of patients who did not achieve remission following an initial adequate trial with one SGA (insufficient SOE).

Discussion

Key Findings and Strength of Evidence

Across all interventions, we graded the strength of evidence as moderate for only two comparisons, namely SGAs compared with cognitive behavioral therapy (CBT) and St. John's wort. Results from trials of these comparisons indicate that CBT and St. John's wort have levels of effectiveness regarding symptomatic relief similar to those of SGAs. The overall risk for adverse events or discontinuation of treatment because of adverse events, however, is lower for these non-SGA therapies. Our confidence in findings from the remaining treatment options was low or insufficient, indicating that these bodies of evidence had major or unacceptable deficiencies. Nevertheless, for most comparisons the overall findings did not detect statistically significant differences in effectiveness but did indicate a lower risk of adverse events for nonpharmacological treatment options. Notable exceptions are omega-3-fatty acids, which appear to have lower effectiveness than SGAs and the combination of SGAs with acupuncture which appear to have better effectiveness than SGA monotherapy. Our confidence in these findings, however, is low and results have to be interpreted cautiously. In addition, for many comparisons that are limited to single trials, determining whether similar treatment effects between SGAs and other interventions are based on similar effectiveness or high placebo response rates is impossible.

The limited amount of comparative intervention data addressing whether depressive severity moderates the comparative effectiveness offers no conclusions on how selection of treatment strategies might differ based on a patient's severity of depression. Overall, the available data

does not indicate differences in the comparative effectiveness between SGAs and non-pharmacological interventions for patients with severe MDD. This important question, however, raised by a few systematic reviews,¹⁸⁻²⁰ remains without a clear answer.

Beyond the two articles identified comparing switch and augmentation strategies employing a limited number of medication options or CT, the absence of relevant comparative data about which treatment options are most effective for those needing second-line treatment (about 70 percent of patients with MDD)^{21,22} was striking.

Our findings are consistent with several prior systematic reviews and meta-analyses that compared SGAs with nonpharmacological interventions. Most of these reviews, however, included populations that were not eligible for our review, such as patients with minor depression, bipolar disorder, or dysthymia.

Our results are partially consistent with the recommendations of both the American Psychiatric Association²³ and the Department of Veterans Affairs/Department of Defense.²⁴ These consider both pharmacotherapy and psychotherapy to be appropriate individual first-line treatments for patients with mild to moderate MDD, and state that the combination of pharmacotherapy and psychotherapy may be necessary in cases of moderate to severe depression.

Results with moderate SOE indicating similar effectiveness can serve as a reasonable starting place for providers and patients for starting a course of medication, CBT, or St. John's wort to treat MDD. Patients who strongly prefer one or the other therapy can be allowed freedom to choose among available options, while under a physician's supervision and monitoring. Moreover, patients who would like to maintain or start an exercise regimen in addition to undergoing SGA therapy can be encouraged to do so. The enhanced potential for increasing physical well-being and expanding social interactions may be an added incentive to encourage an exercise regimen.

Applicability

The scope of this review was limited to trials that enrolled adult patients with MDD. We did not attempt to review literature on interventions for MDD in children or for patients with subthreshold depression, dysthymia, or perinatal depression. The included trials covered populations with mild, moderate, and severe MDD; the majority of participants were women. Most trial populations, however, excluded patients with medical comorbidities; few trials included elderly patients. We did not find evidence to confirm or refute whether treatments are more or less efficacious for various subgroups (i.e., patients characterized by sex, race, or ethnicity or individuals with coexisting psychiatric conditions).

With few exceptions, interventions in included trials were in line with clinical practice. Except for some CAM trials in which patients received SGA dosages at the lower end of the recommended range, prescribing patterns and doses in the SGA arms of our evidence base were consistent with clinical practice. Some newer SGAs such as desvenlafaxine, levomilnacipran, vilazodone, or vortioxetine, however, have never been compared with psychological or CAM treatments or exercise. Nevertheless, reliable evidence indicates that the comparative effectiveness of SGAs is similar.²⁵ Consequently, we believe that our findings are applicable across the class of SGAs.

As noted above, detecting no statistically significant difference does not necessarily mean the outcomes are equivalent. The studies involved were designed to test whether an outcome for one intervention was different from another rather than to test equivalence, which would generally

require a much larger sample size. This point is especially relevant for those findings with a moderate SOE. While confidence intervals were relatively narrow, and risk ratios were often close to 1 (findings consistent with equivalent outcomes), a conclusion of equivalence cannot be made. Further, while comparative effectiveness at a group level did not detect a difference between SGA and CBT or St John's wort, how best to tailor this information to an individual patient is still not clear. Indeed, other potentially relevant indicators (e.g., depressive severity, comorbid psychiatric illness) may favor one over another, but the current evidence base (as indicated in the KQ 1b and 2b findings) is quite limited.

Finally, many trials, particularly for CAM interventions, were conducted outside the United States. Whether and how differences in ethnic or cultural backgrounds and health systems affect the applicability of results to U.S. populations remain uninvestigated and unanswered.

Research Gaps

Across all comparisons of interventions, major research gaps pertain to information about the comparative risk of harms and patient-relevant outcomes such as functional capacity and quality of life. Lack of information about harms can lead to a biased knowledge base and the potential for decisions that cause more harm than good.

We found no eligible studies that compared SGAs with behavior therapy or behavior modification, humanistic therapies, yoga, or mindfulness interventions. Given the wide use of these types of psychotherapies in clinical practice, further research into their comparative effectiveness with SGAs in treating MDD patients is desirable. For many psychotherapies and all CAM therapies that have been evaluated against an SGA, the data were insufficient because trials did not report important outcomes, most notably quality of life and functional capacity. Future studies should assess remission, response to treatment, quality of life, and functional capacity using standardized measures to allow for more direct comparisons across studies using the same or similar SGAs and psychological interventions. These same deficiencies in the literature extend to the comparative effectiveness of SGAs and both psychological and CAM interventions for treating MDD as a function of depression severity.

Finally, a major gap in the evidence is the lack of studies addressing different treatment options for patients who have not achieved remission with first-line therapy. No second-line therapy data at all exist comparing SGA with CAM or exercise treatments. This void in the evidence base is a major one that will perplex and confound clinicians, patients, policymakers, and guideline developers alike.

Conclusions

Overall, the available evidence does not support the superiority of SGAs over CBT and St. John's wort as first line treatments for patients with moderate to severe MDD. Given no clear differences in beneficial treatment effect among treatment options, the choice of the initial treatment of MDD should be strongly based on patient preferences and the feasibility (e.g., costs, likely adherence) following a discussion of the advantages and disadvantages of each treatment option.

Most comparisons of SGAs with other treatment options also did not detect statistically significant differences, however, these findings have to be viewed more cautiously because of methodological limitations. Only omega-3-fatty acids appear to have lower effectiveness than SGAs.

Differences with respect to adverse events, personal engagement, and costs may be taken into consideration for the choice of a first-line treatment. Such shared and informed decisionmaking might enhance treatment adherence and improve treatment outcomes for patients with MDD, especially because treatment continuity is one of the main challenges in treating such patients. For second-line therapies, although evidence is limited, no clear benefit emerges to suggest either switching to a particular SGA or CT or augmenting with a particular medication or CT. The more important decision appears to be simply to try a different evidence-based approach.

References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed., Arlington, VA: American Psychiatric Publishing; 2013.
2. Kessler RC, Berglund P, Demler O, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA*. 2003 Jun 18;289(23):3095-105. Epub: 2003/06/19. PMID: 12813115.
3. Khan A, Sambunaris A, Edwards J, et al. Vilazodone in the treatment of major depressive disorder: efficacy across symptoms and severity of depression. *Int Clin Psychopharmacol*. 2014 Mar;29(2):86-92. Epub: 2013/11/20. PMID: 24247740.
4. Nemeroff CB. The burden of severe depression: a review of diagnostic challenges and treatment alternatives. *J Psychiatr Res*. 2007 Apr-Jun;41(3-4):189-206. Epub: 2006/07/28. PMID: 16870212.
5. Thase ME. Treatment of severe depression. *J Clin Psychiatry*. 2000;61 Suppl 1:17-25. Epub: 2000/03/07. PMID: 10703759.
6. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: The National Academies Press; 2001.
7. Mojtabai R, Olfson M. National patterns in antidepressant treatment by psychiatrists and general medical providers: results from the national comorbidity survey replication. *J Clin Psychiatry*. 2008 Jul;69(7):1064-74. Epub: 2008/04/11. PMID: 18399725.
8. Rimer J, Dwan K, Lawlor DA, et al. Exercise for depression. *Cochrane Database Syst Rev*. 2012;7:CD004366. Epub: 2012/07/13. PMID: 22786489.
9. Gartlehner G, Hansen RA, Morgan LC, et al. Second-Generation Antidepressants in the Pharmacologic Treatment of Adult Depression: An Update of the 2007 Comparative Effectiveness Review. (Prepared by the RTI International–University of North Carolina Evidence-based Practice Center, Contract No. 290-2007-10056-I.) AHRQ Publication No. 12-EHC012-EF. Rockville, MD: Agency for Healthcare Research and Quality; December 2011. www.effectivehealthcare.ahrq.gov/reports/final.cfm.
10. AHRQ Methods for Effective Health Care. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008.
11. Viswanathan M, Ansari MT, Berkman ND, et al. Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Agency for Healthcare Research and Quality Methods Guide for Comparative Effectiveness Reviews AHRQ Publication No. 12-EHC047-EF. Rockville MD: Agency for Healthcare Research and Quality; March 2012. www.effectivehealthcare.ahrq.gov/.
12. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928. Epub: 2011/10/20. PMID: 22008217.
13. Hong H, Carlin BP, Shamliyan TA, et al. Comparing Bayesian and frequentist approaches for multiple outcome mixed treatment comparisons. *Med Decis Making*. 2013 Jul;33(5):702-14. Epub: 2013/04/04. PMID: 23549384.
14. Jones B, Roger J, Lane PW, et al. Statistical approaches for conducting network meta-analysis in drug development. *Pharm Stat*. 2011 Nov-Dec;10(6):523-31. Epub: 2012/01/04. PMID: 22213533.
15. Berkman ND, Lohr KN, Ansari M, et al. Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update. Methods Guide for Effectiveness and Comparative Effectiveness Reviews (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290-2007-10056-I) AHRQ Publication No. 13(14)-EHC130-EF. Rockville MD: Agency for Healthcare Research and Quality; November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.
16. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol*. 2011 Apr;64(4):395-400. Epub: 2011/01/05. PMID: 21194891.

17. Atkins D, Chang S, Gartlehner G, et al. Assessing the Applicability of Studies When Comparing Medical Interventions Methods Guide for Comparative Effectiveness Reviews AHRQ Publication No. 11-EHC019-EF. Rockville MD: Agency for Healthcare Research and Quality; January 2011.
18. Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA*. 2010 Jan 6;303(1):47-53. Epub: 2010/01/07. PMID: 20051569.
19. Kirsch I, Deacon BJ, Huedo-Medina TB, et al. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med*. 2008 Feb;5(2):e45. Epub: 2008/02/29. PMID: 18303940.
20. Khan A, Leventhal RM, Khan SR, et al. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J Clin Psychopharmacol*. 2002 Feb;22(1):40-5. Epub: 2002/01/19. PMID: 11799341.
21. Gaynes BN, Rush AJ, Trivedi MH, et al. Primary versus specialty care outcomes for depressed outpatients managed with measurement-based care: results from STAR*D. *J Gen Intern Med*. 2008 May;23(5):551-60. Epub: 2008/02/06. PMID: 18247097.
22. Gaynes BN, Lux LJ, Lloyd SW, et al. Nonpharmacologic Interventions for Treatment-Resistant Depression in Adults. Comparative Effectiveness Reviews, No. 33. (Prepared by RTI International-University of North Carolina (RTI-UNC) Evidence-based Practice Center under Contract No. 290-02-0016I.) AHRQ Publication No. 11-EHC056-EF. Rockville, MD: Agency for Healthcare Research and Quality; Sep 2011. www.effectivehealthcare.ahrq.gov/reports/final.cfm.
23. American Psychiatric Association. Practice guideline for the treatment of patients with major depressive disorder. Arlington (VA): American Psychiatric Association; Oct 2010. http://psychiatryonline.org/pb/assets/raw/site_wide/practice_guidelines/guidelines/mdd.pdf.
24. Department of Veteran Affairs, Department of Defense. VA/DoD clinical practice guideline for management of major depressive disorder (MDD). Washington (DC): Department of Veteran Affairs, Department of Defense; May 2009. http://www.healthquality.va.gov/mdd/mdd_full09_c.pdf.
25. Gartlehner G, Hansen RA, Morgan LC, et al. Comparative benefits and harms of second-generation antidepressants for treating major depressive disorder: an updated meta-analysis. *Ann Intern Med*. 2011 Dec 6;155(11):772-85. Epub: 2011/12/08. PMID: 22147715.

Introduction

Background

Clinical and Methodological Issues

Context

Depressive disorders can be serious, disabling illnesses. Major depressive disorder (MDD),¹ defined as the presence of depressed mood or loss of interest or pleasure, along with at least four additional MDD diagnosis criteria or symptoms for at least 2 weeks, is the most prevalent and disabling, affecting more than 16 percent of U.S. adults (lifetime).² MDD can be characterized as mild, moderate, or severe based on symptom severity, functional impairment, and level of patient distress;¹ in clinical trials, these distinctions are typically made by scores on a depressive rating instrument.³ Approximately one-third of patients with MDD are severely depressed,⁴ which is associated with a harder to treat depression.⁵

The burden of depressive illnesses, in both human and financial terms, is enormous; by the year 2020, depression is expected to become the second leading cause of disability throughout the world, trailing only ischemic heart disease.⁶ MDD, in particular, exerts a negative impact on physical health. It reduces participation in preventive health care activities^{7,8} and adherence to medical treatment.⁹ It increases the likelihood of chronic conditions such as obesity, smoking, sedentary lifestyles, and hypertension,^{10,11} as well as amplifies the risk of cancer¹² and death following myocardial infarction.¹⁰ Mortality rates attributable to MDD and other depressive illnesses are high; approximately 4 percent of adults with a mood disorder commit suicide, and about two-thirds of suicides are preceded by depression.¹³

In 2000, the U.S. economic burden associated with depressive disorders was estimated to be \$83.1 billion, a figure that has likely increased during the ensuing 10 years. More than 30 percent of these costs are attributable to direct medical expenses.¹³

In any given year, nearly 7 percent of the U.S. adult population (approximately 17.5 million people in 2014) experiences an episode of MDD that warrants treatment.² Approximately one-half of these patients seek care. Most patients receiving care obtain treatment in primary care settings,¹⁴ where second-generation antidepressants (SGAs) are the most commonly prescribed agents.¹⁵ Patients who initially present to a psychiatric clinic are, in general, similar to those who seek treatment in primary care settings.^{16,17}

For patients who do receive care, only 20 percent receive a minimal degree of adequate treatment, based on available evidence-based guidelines as receiving either pharmacotherapy (at least 2 months of an appropriate medication for MDD plus more than four visits to any type of physician) or psychotherapy (at least eight visits with any health care professional lasting an average of at least 30 minutes).¹⁸ Relative to these guidelines,^{19,20} for the general population of patients with MDD, the risk of *undertreatment* can be substantial.

In contrast, for the group receiving pharmacotherapy treatment, *overtreatment* with antidepressant medications poses another potential risk. Several recent studies involving comparisons with placebo controls have highlighted differences in response to pharmacotherapy based on baseline depression severity, suggesting a risk of excessive use of these treatment interventions for patients with mild disease.²¹⁻²⁴ Eligibility criteria for most clinical trials require severely or very severely depressed patients, raising concerns about the generalizability of their

results to populations with milder degrees of MDD (which are commonly seen in primary care settings).

Several meta-analyses have reported that as baseline depressive symptoms increase, response to pharmacotherapy improves. One meta-analysis of patient-level data from six randomized controlled trials (RCTs) of antidepressants reported that response to two types of antidepressants (imipramine or paroxetine) begins to outpace placebo response only when baseline scores on the 17-item version Hamilton Depression Rating Scale (HAM-D) exceed 25.²¹ In other words, patients with mild MDD who are identified and treated may be at risk of antidepressant overtreatment. Therefore, considering the role of depression severity in MDD on treatment outcomes can be crucial in guiding treatment selection.

Outcomes following an initial, evidence-based treatment with antidepressants in primary care settings are equivalent to those in tertiary care psychiatric clinics. In each of these types of settings, approximately 30 percent of patients will experience symptom remission (usually defined as a HAM-D score of ≤ 7); about 70 percent will have an inadequate treatment response.^{25,26} Providing this latter group (i.e., the remaining 70 percent) with a second treatment attempt led to similar rates of improvement;²⁷ such interventions can include switching antidepressants or augmenting with a second medication.

These data suggest that outcomes achieved in psychiatric clinics for both an initial treatment attempt and a second attempt are applicable to primary care settings. However, remission decreases to 15 percent for patients who have not yet recovered following two adequate antidepressant trials. This pattern suggests that patients experiencing treatment failure following two adequate trials of antidepressants would benefit from referral to a psychiatric clinic where clinicians can try more complicated treatment regimens.²⁸ Accordingly, this systematic review (SR) will focus on the initial two treatment attempts for depressive illness.

Purpose of this Report

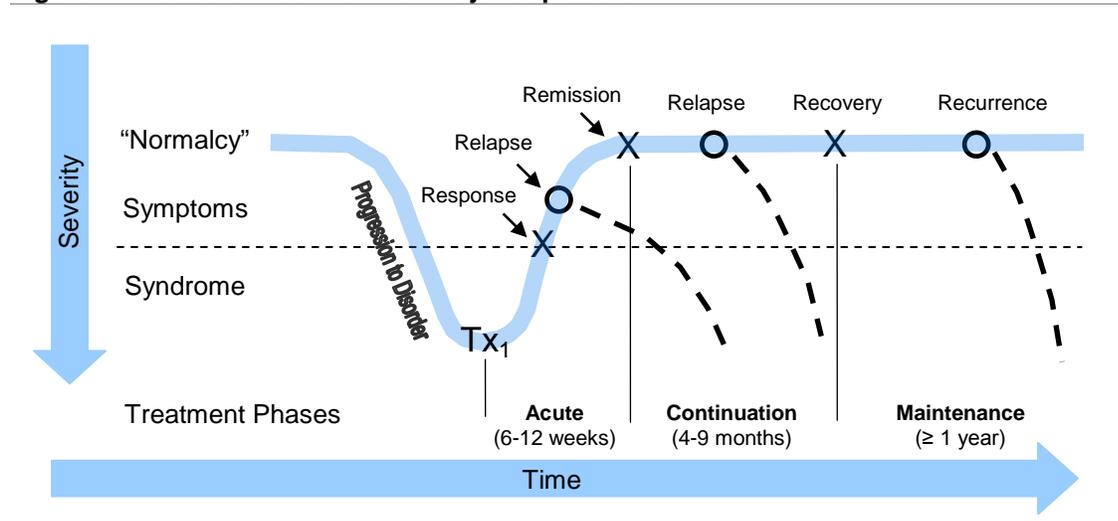
Primary care physicians provide the largest number of antidepressant prescriptions and account for most of the near doubling in the use of antidepressants over the past decade.²⁹ Accordingly, much of this treatment may be for patients with either threshold or mild MDD, suggesting a risk of overtreatment for this group. At the same time, primary care physicians appreciate that other potentially effective interventions are available. According to the topic nominators, primary care physicians require an evidence base identifying the comparative effectiveness of the available treatments for depression to increase the likelihood that treatments are selected and managed correctly. This review will focus on two key issues facing primary care physicians:

1. As an initial treatment choice, how effective are SGAs compared with nonpharmacologic interventions?
2. For patients whose depression did not achieve remission following initial treatment with an SGA, what is the comparative effectiveness of alternative pharmacologic and nonpharmacologic options? These options include adding a pharmacologic or nonpharmacologic treatment to the initial medication choice (which we refer to as augmentation) or switching to a different SGA or to a nonpharmacologic treatment.

Interventions for MDD

Management of MDD involves three treatment phases (see Figure 1): the acute phase, in which symptoms are treated to remission; the continuation phase, during which remission is sustained until the episode has resolved (ranging from 4 to 9 months); and the maintenance phase, in which treatment is maintained to prevent recurrence of another episode of MDD.

Figure 1. Phases of treatment for major depression



Source: Recreated based on Kupfer, 1991.³⁰ Tx₁=treatment attempt 1; dashed lines indicate hypothetical worsening of depressive severity. Remission, the goal of for treatment, refers to the resolution of depressive symptoms and return to premorbid functioning; response refers to substantial clinical improvement which may or may not reach remission.

Pharmacotherapy remains the primary intervention for MDD patients in primary care. Nonetheless, primary care patients and clinicians may prefer other options (or at least want to be able to consider them). These include psychotherapeutic interventions, complementary and alternative medicine (CAM) options, or exercise. As noted above, clinicians want comparative effectiveness data to help guide treatment selection across these various choices.³¹

We review below the treatment options relevant to this comparative effectiveness review. Given the likelihood of greater benefit of pharmacotherapy for more severely depressed than mildly depressed patients, an important clinical issue is to determine the comparative benefits and harms of SGAs with other treatment options such as psychotherapy, CAM interventions, or exercise as potential monotherapy for patients with mild to severe MDD. A related issue concerns their roles as potential adjuncts to antidepressants for patients with more severe MDD.

Pharmacotherapy for MDD

Pharmacotherapy (e.g., SGAs) dominates the medical management of depressive disorders. This SR will focus on SGAs, which we define as including selective serotonin reuptake inhibitors, serotonin and norepinephrine reuptake inhibitors, bupropion, mirtazapine, nefazodone, and trazodone. Standard dosing for SGAs is shown in Table 1.

Focusing solely on SGAs more accurately represents the pharmaceutical therapies that primary care clinicians prescribe most often.^{15,32} Furthermore, because SGAs are most frequently used as first-line therapy, we will examine only comparisons that include SGAs in at least one arm of any given comparative study.

Table 1. SGAs: Usual dosing range and frequency of administration for adults

Generic Name	U.S. Trade Name ^a	Usual Daily Dosing Range	Frequency
Bupropion	Wellbutrin®	200–450 mg	Three times daily
	Wellbutrin SR®	150–400 mg	Twice daily
	Wellbutrin XL®	150–450 mg	Once daily
Citalopram	Celexa®	20–40 mg	Once daily
Desvenlafaxine	Pristiq®	50 mg	Once daily
Duloxetine	Cymbalta®	40–60 mg ^b	Once or twice daily
Escitalopram	Lexapro®	10–20 mg	Once daily
Fluoxetine	Prozac®	10–80 mg	Once or twice daily
	Prozac Weekly®	90 mg (weekly)	Once weekly
Fluvoxamine	Luvox®	50–300 mg	Once or twice daily
Levomilnacipran	Fetzima®	40–120 mg	Once daily
Mirtazapine	Remeron®	15–45 mg	Once daily
	Remeron Sol tab®	15–45 mg	Once daily
Nefazodone	Serzone®	200–600 mg	Twice daily
Paroxetine	Paxil®	20–60 mg	Once daily
	Paxil CR®	12.5–75 mg	Once daily
Sertraline	Zoloft®	50–200 mg	Once daily
Trazodone	Desyrel®	150–400 mg	Three times daily
Venlafaxine	Effexor®	75–375 mg	Two to three times daily
	Effexor XR®	75–225 mg	Once daily
Vilazodone	Viibryd®	40 mg	Once daily
Vortioxetine	Brintellix®	10–20 mg	Once daily

^a CR, SR, XL, and XR are registered trademarks referring to controlled-, sustained-, or extended-release dosage forms, respectively.

^b Doses of duloxetine up to 120 mg were studied in clinical trials, although doses above 60 mg are not believed to have additional efficacy.

Available evidence for MDD does not warrant choosing one SGA over another based on either greater efficacy or greater effectiveness.³² Only about 60 percent of patients treated with SGAs respond to treatment (meaning specifically that their depressive severity decreases by at least half, an improvement that may or may not meet criteria for remission); approximately 30 percent achieve remission during the first-line treatment.³³

More than 60 percent of patients experience at least one adverse effect during treatment. Although most adverse effects are minor, such as constipation, diarrhea, and dizziness, they frequently lead to discontinuation of treatment.³⁴

As documented above, 70 percent of MDD patients do not achieve remission following initial pharmacological treatment, and available data indicate that no one antidepressant performs better than any other. Accordingly, various other interventions—such as medication combinations, psychotherapy, or CAM treatments—are important options for patients and clinicians. In addition, lifestyle changes, for example, increased exercise, have been recommended as adjunctive treatments for MDD.^{35,36} Finally, strategies to augment antidepressant medications for those failing an initial treatment attempt may provide better treatment response than single medications alone.³⁷

Psychotherapy for MDD

The American Psychological Association recently concluded that the general benefits of the major psychotherapies that have been studied are significant and large.^{38,39} Some effects of psychotherapy tend to last longer and to be less subject to relapse requiring additional treatment than outcomes following pharmacological interventions;⁴⁰ however, the effect of depressive severity on these results is not clear. The psychological interventions used to treat depressed

patients include acceptance and commitment therapy, cognitive therapy, cognitive behavioral therapy, interpersonal therapy, psychodynamic therapies, and other talk therapies, which may have different customary lengths of treatment. Of note, the optimal frequency of psychotherapy has not been rigorously studied in controlled trials, so there is no clear evidence for what might be considered adequate or standard dosing.¹⁹

In general, these interventions potentially help people identify how past and present factors may contribute to their depression and teach them how to deal effectively with them. Certain psychological interventions can help individuals identify negative or distorted thought patterns that contribute to feelings of hopelessness and helplessness that accompany depression. These interventions can also help people acquire skills to relieve suffering and prevent later bouts of depression. Among them are developing or strengthening social networks, creating new ways to cope with challenges, and following self-care plans that include positive lifestyle changes. To date, however, little is known about the comparative efficacy and effectiveness or harms of psychological interventions to treat depression.

CAM for MDD

CAM interventions are a growing area of both treatment and research. They are most often used in conjunction with conventional treatments (as complementary medicine) rather than as instead of conventional therapies. Although evidence-based standard dosing schedules for most dietary supplements do not currently exist, the European Union has produced some guidelines for dosing of St. John's wort.^{41,42} Most sources suggest using an extract standardized to 0.1 percent to 0.3 percent hypericin with a dose of 900 mg daily, usually divided into three doses, to deliver a daily hypericin dose of 1 to 2 mg. Although some clinical trials have demonstrated the importance of additional standardization to 3 percent to 5 percent hyperforin, no guidelines for hyperforin content currently exist because of inconsistent results among trials.⁴³ In the absence of clear guidelines, protocols followed in clinical trials may define standard practice.

Numerous clinical trials and reviews of CAM therapies for depression exist, including several Cochrane reviews.⁴⁴⁻⁴⁷ In addition to SRs, the American Psychiatric Association Task Force and the Canadian Network for Mood and Anxiety Treatments have issued practice guidelines that incorporate the adjunctive use of several CAM interventions.^{48,49} Although the evidence base from high-quality RCTs is limited, sufficient placebo-controlled evidence exists to support St. John's wort for mild to moderate MDD.⁵⁰ The evidence base is not as robust for the use of yoga, acupuncture, meditation, S-adenosyl-L-methionine, and omega-3 fatty acids.^{46,51-55}

Adverse events are uncommon for most CAM treatments, but potential drug interactions between some dietary supplements and other medications are of some concern. For example, St. John's wort should not be recommended to patients taking any pharmaceutical medications without the advice of a medical provider or pharmacist with expertise in evaluating herb-drug interactions. Importantly, more than one-half of patients with depression are estimated to use some form of CAM therapy, and the majority of patients do not spontaneously disclose CAM use to their care providers.⁵⁶

The comparative effectiveness (either benefits or harms) of CAM and other therapies is not known. As noted for other interventions, the role of depressive severity on these outcomes remains unclear as well.

Exercise for MDD

The use of exercise as either a primary treatment or an augmentation strategy for depression has a growing literature and evidence base. The most comprehensive Cochrane review identified 32 trials involving 1,858 participants with diagnosed MDD;⁵⁷ the authors found a moderate clinical benefit of exercise versus no treatment or control. Although small in number, some studies compare exercise with cognitive therapy, medications, and alternative therapies; most find no clear differences in benefits.

This literature continues to evolve. SRs of exercise versus an inactive control suggest small but clinically meaningful benefits (in the elderly a reduction of approximately 20 percent in depressive severity).⁵⁸ In addition, recently published clinical trial data indicate that the benefit from exercise is similar to that from sertraline in terms of reducing depressive symptoms in patients with cardiovascular disease and elevated depressive symptoms (but not necessarily MDD), with additional improvements in cardiovascular biomarkers; these findings suggest benefit for both clinical outcomes and quality of life.⁵⁹

Nevertheless, the comparative effectiveness of exercise as either a primary treatment for MDD or an augmentation therapy is unknown. Several clinical trials addressing MDD and exercise are currently under way (<http://ccdan.cochrane.org/specialised-register>; <http://clinicaltrials.gov/>), suggesting a need for a review of this area.

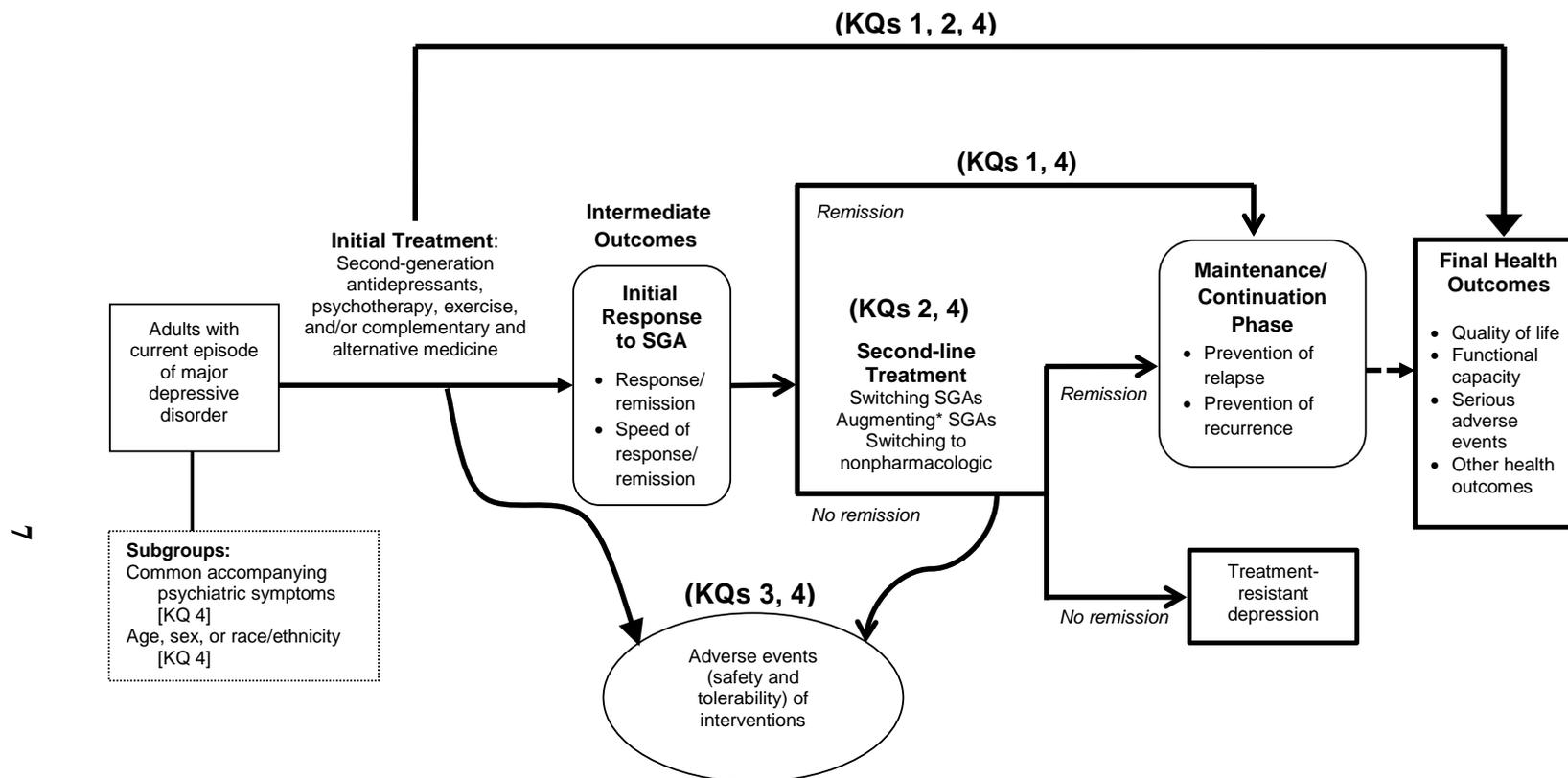
Exercise covers a broad range of activities done over varying durations of time and done singly, in classes, or in informal groups. This SR will focus on the benefits and harms of formal exercise activities (a prescribed exercise regimen, either supervised or unsupervised) that enroll people with an explicit diagnosis of MDD because these interventions are the ones most likely to be studied in trials.

Scope and Key Questions

Scope of This Review

This review will examine the evidence base for primary care management of MDD for the first two treatment attempts, after which primary care clinicians would consider referral to or consultation by a mental health professional. The specific Key Questions (KQs) are listed below, and Figure 2 displays the analytic framework that guided our work.

Figure 2. Analytic framework for treatment of major depressive disorder



*Augmenting with a second SGA, additional medication or a nonpharmacologic treatment

KQ = Key Question; SGA = second-generation antidepressant

Key Questions

KQ 1a: In adult patients with MDD who are undergoing an initial treatment attempt, what is the effectiveness of second-generation antidepressant (SGA) monotherapy compared with the effectiveness of either nonpharmacological monotherapy or combination therapy (involving nonpharmacological treatments with or without an SGA)?

KQ 1b. Does comparative treatment effectiveness vary by MDD severity?

KQ 2a. In adult patients with MDD who did not achieve remission following an initial adequate trial with one SGA, what is the comparative effectiveness of second line therapies*?

* Any comparison that involves an eligible intervention (whether as a monotherapy or a combination therapy) and compares an intervention to one involving an SGA is eligible. Examples of potential comparisons are listed below.

KQ 2b. Does comparative treatment effectiveness vary by MDD severity?

KQ 3a. In adult patients with MDD, what are the comparative risks of harms of these treatment options:

(1) for those undergoing an initial treatment attempt or

(2) for those who did not achieve remission following an initial adequate trial with an SGA?

KQ 3b. Do the comparative risks of treatment harms vary by MDD severity?

KQ 4. Do the benefits and risks of harms of these treatment options differ by subgroups of patients with MDD defined by common accompanying psychiatric symptoms (coexisting anxiety, insomnia, low energy, or somatization) or demographic characteristics (age, sex, race, or ethnicity)?

Organization of This Report

The remainder of the review first describes our methods in detail; it then presents the results of our synthesis of the literature with summary tables and the strength of evidence grades for major comparisons and outcomes. The discussion section offers our conclusions, summarizes our findings, and provides other information relevant to interpreting this work for clinical practice and future research. References, a list of acronyms and abbreviations, and a glossary of terms follow the Discussion section.

Appendix A contains the exact search strings for our literature searches. Appendix B presents the typology used to categorize common, depression-focused psychotherapies. Appendix C lists the studies excluded at the stage of reviewing full-text articles with reasons for exclusion. Risk-of-bias assessments of individual studies in this review are presented in Appendix D. Strength of evidence profiles appear in Appendix E. Published and unpublished trials included in the network meta-analyses on response to treatment are listed in Appendix F.

Methods

The methods for this comparative effectiveness review follow the guidance provided in the Agency for Healthcare Research and Quality (AHRQ) *Methods Guide for Effectiveness and Comparative Effectiveness Reviews* (www.effectivehealthcare.ahrq.gov/methodsguide.cfm) for the Evidence-based Practice Center (EPC) program. The main sections in this chapter reflect the elements of the protocol established for this review. Certain methods map to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist.⁶⁰ All methods and analyses were determined a priori.

The AHRQ Effective Health Care (EHC) program's Topic Triage group developed and reviewed the topic; because this group deemed the topic sufficiently relevant, they moved it forward for the Topic Refinement phase. All topics are reviewed and assessed for appropriateness for systematic review (SR) (see EHC Web site for information on the process for selecting topics: <http://www.effectivehealthcare.ahrq.gov/index.cfm/submit-a-suggestion-for-research/how-are-research-topics-chosen/>). Once a topic is assessed and determined to be appropriate for further product development in the EHC program, AHRQ assigns it to a research team. Further development of the topic occurs with the input of key informants and technical experts (see the EHC Web site for information on the research process: <http://www.effectivehealthcare.ahrq.gov/index.cfm/submit-a-suggestion-for-research/what-is-the-research-process/>).

Topic Refinement and Review Protocol

During the topic refinement, we engaged in a public process to develop a draft and final protocol for the review. We generated an analytic framework, preliminary Key Questions (KQs), and preliminary inclusion/exclusion criteria in the form of PICOTS (populations, interventions, comparators, outcomes, timing, settings). Information provided by the topic nominator helped guide our processes. Initially a panel of eight Key Informants gave input on the KQs to be examined; these KQs were posted on AHRQ's Web site for public comment (www.effectivehealthcare.ahrq.gov) from February 3, 2014, through February 24, 2014, and revised as needed. We then drafted a protocol for the SR.

In addition, we recruited a panel of technical experts (TEP) to provide high-level content and methodological expertise throughout the development of the review. They represented consumer perspective and professional organizations, researchers, and payers with expertise in psychopharmacology, psychotherapy, complementary and alternative medicine (CAM), and exercise therapies for depression. TEP members participated in one conference call to review the analytic framework, KQs, and PICOTS and in several discussions through email.

Literature Search Strategy

Search Strategy

To identify articles relevant to each KQ, we searched MEDLINE[®] (via PubMed), EMBASE, the Cochrane Library, AMED (Allied and Complementary Medicine Database), PsycINFO, and CINAHL (Cumulative Index to Nursing and Allied Health Literature) from January 1, 1990, through May 2, 2014, using analogous search terms (Appendix A). We used a combination of medical subject headings (MeSH) and title and abstract key keywords, focusing on terms to describe the relevant population and interventions of interest. An experienced information

scientist ran the searches; another information scientist (EPC librarian) peer-reviewed the searches. We limited the electronic searches to English-, German-, and Italian-language and human-only studies.

In addition to electronic searches, we manually searched reference lists of pertinent reviews, included trials, and background articles on this topic to identify any relevant citations that our searches might have missed. We imported all citations into an EndNote[®]X6 electronic database.

We searched for “gray literature” relevant to this review following guidance from the *Methods Guide for Effectiveness and Comparative Effectiveness Reviews* for these steps.⁶¹ Sources of gray literature included ClinicalTrials.gov, the World Health Organization’s International Clinical Trials Registry Platform, Drugs@FDA, the European Medicines Agency, the National Institute of Mental Health Web site, the American Psychological Association Web site, Scopus, and the Conference Proceedings Citation Index.

The AHRQ Scientific Resource Center requested scientific information packets from relevant pharmaceutical manufacturing companies, asking for any unpublished studies or data relevant for this SR. The AHRQ Scientific Resource Center managed the process of submitting requests for scientific information packets, which contain information about drugs and CAM interventions. We received information packets from Eli Lilly and Company and Merck & Co., Inc.

We investigated any literature suggested by the peer reviewers or the public and, when appropriate using the same methods as described below, incorporated additional studies into the final review.

Inclusion and Exclusion Criteria

We specified our inclusion and exclusion criteria based on the PICOTS identified in topic refinement. Table 2 specifies inclusion and exclusion criteria; subsequent sections define the PICOTS in more detail.

Population(s)

For this review, we included adult (18 years or older) outpatients of all races and ethnicities with MDD during either an initial treatment attempt (KQ 1) or a second treatment attempt in patients who did not achieve remission following an initial adequate trial with a second-generation antidepressant (SGA) (KQ 2).

Subgroups of interest are based on

- common accompanying psychiatric symptoms (anxiety, insomnia, low energy, somatization),
- age,
- sex, and
- race or ethnicity.

We did not include patients with bipolar depression, perinatal depression, chronic depression, seasonal affective disorder, psychotic depression, or treatment-resistant depression (i.e., two or more treatment failures). We classified severity of depression of patients following a categorization that is outlined in Table 3.

Table 2. Inclusion/exclusion criteria

	Inclusion	Exclusion
Population	Adult (18 years or older) outpatients of all races and ethnicities with MDD during either an initial treatment attempt or a second treatment attempt in patients who did not remit following an initial adequate trial with an SGA	<ul style="list-style-type: none"> • Children under age 18 • Patients with perinatal depression, seasonal affective disorder, psychotic depression, or treatment-resistant depression (i.e., two or more treatment failures)
Geography	No limit	<ul style="list-style-type: none"> • No limit
Date of search	Searches went back until 1990	<ul style="list-style-type: none"> • Articles published before January 1990
Settings	<ul style="list-style-type: none"> • Primary, secondary, and tertiary care outpatient settings 	<ul style="list-style-type: none"> • Inpatient settings
Interventions	<ul style="list-style-type: none"> • As defined in the PICOTS criteria 	<ul style="list-style-type: none"> • First-generation antidepressants • Any other interventions not defined in the PICOTS criteria
Control interventions	<ul style="list-style-type: none"> • As defined in the PICOTS criteria 	<ul style="list-style-type: none"> • Ineligible interventions (see PICOTS criteria)
Outcomes	<ul style="list-style-type: none"> • As defined in the PICOTS criteria 	<ul style="list-style-type: none"> • Studies that do not include at least one of the outcomes listed under the inclusion criteria
Timing of intervention	<ul style="list-style-type: none"> • No limitations 	<ul style="list-style-type: none"> • NA
Publication language	<ul style="list-style-type: none"> • English, German, Italian 	<ul style="list-style-type: none"> • All other languages
Study design	<ul style="list-style-type: none"> • Original research • Eligible study designs include: <ul style="list-style-type: none"> • For efficacy/effectiveness <ul style="list-style-type: none"> - RCTs - SRs and meta-analyses • In addition for harms <ul style="list-style-type: none"> - Nonrandomized controlled trials - Prospective controlled cohort studies - Retrospective controlled cohort studies - Case-control studies - Nonrandomized studies must have a minimum sample size of 500 participants 	<ul style="list-style-type: none"> • Case series • Case reports • Nonsystematic reviews • Studies without a control group • Nonrandomized studies with fewer than 500 participants • Post hoc or secondary analyses • Pooled studies
Publication type	<ul style="list-style-type: none"> • Any publication reporting primary data 	<ul style="list-style-type: none"> • Publications not reporting primary data

MDD = major depressive disorder; PICOTS = populations, interventions, comparators, outcomes, timing, and setting; RCT = randomized controlled trial; SGA = second-generation antidepressant; SR = systematic review.

Table 3. Categories of depressive severity⁶²

Instrument	None/Mild	Moderate	Severe/Very Severe
HAM-D ₁₇	≤ 13	14–19	≥ 20
HAM-D ₂₁	≤ 15	16–22	≥ 23
HAM-D ₂₄	≤ 18	19–26	≥ 27
MADRS	≤ 19	20–34	≥ 35
BDI	≤ 18	18–29	≥ 30
QID-SR	≤ 10	11–15	≥ 16

BDI = Beck Depression Inventory; HAM-D = Hamilton Rating Scale for Depression; MADRS = Montgomery-Åsberg Depression Rating Scale; QID-SR = Quick Inventory of Depressive Symptomatology—Self-Report.

Interventions

For patients with acute-phase MDD and an initial treatment attempt, we were interested in common depression-focused psychotherapies, common CAM interventions, and exercise

- (1) as monotherapies
- (2) in combination with one another, or
- (3) in combination with SGAs.

For patients who did not achieve remission following an initial adequate trial with an SGA, we were also interested in second line therapies that involve an eligible intervention (whether as a monotherapy or a combination therapy). Table 4 presents interventions that were eligible for this report. Appendix B gives a more detailed description of common depression-focused psychotherapies.

Table 4. Eligible interventions for major depressive disorders

Second-Generation Antidepressants	Common Depression-Focused Psychotherapies	Complementary and Alternative Medicines	Exercise	Other Pharmacotherapies for combination or augmentation
<ul style="list-style-type: none"> • Bupropion • Citalopram • Desvenlafaxine • Duloxetine • Fluoxetine • Escitalopram • Fluvoxamine • Levomilnacipran • Mirtazapine • Nefazodone • Paroxetine • Sertraline • Trazodone • Venlafaxine • Vilazodone • Vortioxetine 	<ul style="list-style-type: none"> • Acceptance and commitment therapy • Cognitive and behavioral approaches • Interpersonal therapy • Psychodynamic and attachment-based approaches 	<ul style="list-style-type: none"> • Acupuncture • Meditation (e.g., mindfulness-based stress reduction) • Omega-3 fatty acids • S-adenosyl-L-methionine (SAME) • St. John's wort (<i>Hypericum</i>) • Yoga 	Any formal exercise program	<ul style="list-style-type: none"> • Atypical antipsychotics (aripiprazole, asenapine maleate, clozapine, iloperidone, lurasidone, olanzapine, paliperidone, quetiapine, risperidone, ziprasidone) • Psychostimulants (amphetamine-dextroamphetamine, armodafinil, dexmethylphenidate, dextroamphetamine, lisdexamfetamine, methylphenidate, modafinil) • Buspirone • Levothyroxine (T4) • Lithium • Pindolol • Triiodo-thyronine (T3)

One difficulty that arises with systematic reviews that include a variety of psychological interventions is how to categorize them. When different frameworks are used to organize and categorize the interventions in systematic reviews, the ability to draw conclusions between them can be substantially diminished. The Cochrane Collaborative Depression, Anxiety and Neurosis (CCDAN) Group has developed a framework for categorizing psychological interventions which it uses in its reviews.⁶³ In an effort to enhance consistency of categorization of psychotherapies in this review and our ability to compare our findings to those of other large reviews, we have used the first six categories and descriptions of the CCDAN Group's framework: Behavior/Behavior Modification, Cognitive Behavioral Therapy, Third Wave Cognitive Behavioral Therapy, Psychodynamic Therapies, Humanistic Therapies, and Integrative Therapies.⁶³ We did not include the categories of Systemic Therapies or Other Psychologically-

Oriented Interventions because these categories reflect the mode of delivery as opposed to the type of therapy. Appendix B presents the CCDAN classification in more detail.

Comparators

For KQ 1, we were interested in direct comparisons of eligible interventions with SGAs as single interventions. Except for network meta-analyses, we excluded studies that did not include SGA monotherapies in at least one arm of the study. For KQ 2, we were also interested in studies that modified an existing SGA strategy and compared it with nonpharmacological interventions other pharmacological treatment strategies, or combinations of nonpharmacological and pharmacological strategies. These second line therapies could involve a switch to a new treatment or augmentation of an existing treatment. We excluded studies that did not involve an SGA (whether as a new monotherapy or as part of a combination therapy). Table 5 lists possible head-to-head comparisons of eligible interventions with SGAs.

Table 5. Possible comparisons of eligible interventions with second-generation antidepressants for all populations of interest (i.e., KQ 1, KQ 3, and KQ 4)

SGAs vs. psychotherapies
SGAs vs. CAM
SGAs vs. exercise
SGAs vs. SGA + psychotherapies
SGAs vs. SGA + CAM
SGAs vs. SGA + exercise
SGAs vs. combinations of eligible interventions
In addition for populations who did not achieve remission following an initial adequate trial with an SGA (i.e., KQ 2, KQ 3, and KQ 4):
SGA switch ^a vs. SGA switch ^a
SGA switch ^a vs. nonpharmacologic
SGA switch ^a vs. SGA augmentation ^b
SGA augmentation ^b vs. SGA augmentation ^b
SGA augmentation ^b vs. nonpharmacologic
In addition for network meta-analyses (KQ1):
Any eligible intervention vs. placebo
Any eligible intervention vs. any other eligible intervention

^a Switching to another SGA.

^b Augmenting with a second SGA, for an additional non-SGA medication, or a nonpharmacologic treatment.

CAM = complementary and alternative medicine; KQ = Key Question; SGA = second-generation antidepressant; vs. = versus.

Outcomes

In general, we were interested in patient-relevant health outcomes. In collaboration with the Technical Expert Panel (TEP) and the Key Informants, we selected the following outcomes as relevant for this report.

- **Benefits:** response, remission, speed of response, speed of remission, relapse, quality of life, functional capacity, reduction of suicidality, reduction of hospitalization
- **Harms:** overall adverse events, withdrawals because of adverse events, serious adverse events, specific adverse events (including hyponatremia, seizures, suicidality, hepatotoxicity, weight gain, gastrointestinal symptoms, sexual side effects), or drug interactions (pharmacologic and complementary and alternative treatments)

In addition, during the protocol development, we asked the TEP and the Key Informants to rank the relative importance of these outcomes following a process proposed by the GRADE

Working Group.⁶⁴ We used SurveyMonkey[®] for an anonymous ranking of the relative importance of outcomes. Participants used a 9-point Likert scale to rank outcomes into three categories: (1) critical for decisionmaking, (2) important but not critical for decisionmaking, and (3) of low importance for decisionmaking. Table 6 lists the 11 outcomes (seven benefits, four harms) that respondents viewed as either critical or important for decisionmaking. For average ratings, 9 would indicate greatest importance and 1 least importance.

Table 6. Outcomes rated as critical or important for decisionmaking

Category for Decisionmaking	Outcomes	Average Ratings
Critical	Reduction of suicidality	8.00
	Quality of life	7.57
	Response to treatment	7.43
	Remission	7.29
	Functional capacity	7.29
	Risk of serious adverse events	7.14
Important	Overall risk of adverse events	6.43
	Speed of remission	6.14
	Risk of drug interactions	5.71
	Speed of response	5.71
	Risk of discontinuing treatment because of adverse events	5.43

Timing

We had no limitations on study duration or length of followup.

Setting

We included outpatients from primary, secondary, and tertiary care settings.

Study Selection

Two trained research team members independently reviewed all titles and abstracts identified through searches for eligibility against our inclusion/exclusion criteria using AbstrackR[®].⁶⁵ Studies marked for inclusion underwent full-text review. For studies without adequate information at the title/abstract stage to determine inclusion or exclusion, we retrieved the full text and then made the determination. All results at both title/abstract and full-text review stages were tracked in an EndNote[®] bibliographic database (Thomson Reuters, New York, NY).

We retrieved and reviewed the full text of all articles retained during the title/abstract phase. Two trained team members independently reviewed each full-text article for inclusion or exclusion based on the eligibility criteria described above. If both reviewers agreed that a study did not meet the eligibility criteria, we excluded the study. If the reviewers disagreed, conflicts were resolved by discussion and consensus or by consulting a third member of the review team. We recorded the reason that each excluded full-text publication did not satisfy the eligibility criteria. If the information in published articles was insufficient to permit us to decide about inclusion or exclusion, we contacted authors for further clarification. Appendix C gives the bibliography of excluded studies and reasons for exclusion.

For this review, results from low risk-of-bias head-to-head trials provide the strongest evidence to compare interventions of interest with respect to benefits and harms. In addition to head-to-head studies, we included placebo-controlled trials for network meta-analysis. For harms (i.e., evidence pertaining to safety, tolerability, and adverse events), we intended to examine data

from both randomized and nonrandomized studies; however, we found no eligible nonrandomized studies. (Throughout this report we use “harms” as a summary term for adverse events and unwanted effects, as suggested by the CONSORT [Consolidated Standards of Reporting Trials] statement.⁶⁶)

Data Extraction

We designed, pilot-tested, and used a structured data abstraction form to ensure consistency of data abstraction. Trained reviewers initially abstracted data from each study. A senior reviewer then read each abstracted article and evaluated the completeness and accuracy of the data abstraction. We resolved discrepancies by consensus or by involving a third, senior reviewer.

We abstracted the following data from included trials: study design, eligibility criteria, intervention, additional medications allowed, funder of the study, methods of outcome assessment, population characteristics (such as age, sex, race, ethnicity, or coexisting anxiety), sample size, attrition, and outcomes of interest. We recorded intention-to-treat results (ITT; i.e., all patients are analyzed as randomized with missing values imputed) if available. For studies eligible for quantitative analyses, we contacted authors if reported data were incomplete or missing.

Risk-of-Bias Assessment of Individual Studies

To assess the risk of bias of studies, we used definitions based on AHRQ guidance.⁶⁷ We rated the risk of bias for each relevant outcome of a study as low, moderate, or high. In general terms, results of a study with low risk of bias are considered to be valid. Medium risk of bias implies some confidence that the results represent true treatment effect. The study is susceptible to some bias, but the problems are not sufficient to invalidate the results (i.e., no flaw is likely to cause major bias). A study with high risk of bias has significant methodological flaws (e.g., stemming from serious errors in design or analysis) that may invalidate its results. Ratings of risk of bias are not comparable across study designs. That is, a low risk-of-bias nonrandomized study does not necessarily equal a low risk-of-bias randomized controlled trial (RCT). We take limitations of certain study designs into consideration when we grade the strength of the evidence.

We included all eligible studies regardless of risk of bias in this review. For quantitative analyses, however, we used studies with high risk of bias only for sensitivity analyses.

To determine risk of bias in a standardized way, we used the Cochrane Risk of Bias tool to appraise RCTs.⁶⁸ For nonrandomized studies, we employed criteria outlined by Deeks et al.⁶⁹ For SRs with meta-analyses we used the AMSTAR (Assessing the Methodological Quality of Systematic Reviews) tool.⁷⁰

Two independent reviewers assigned risk-of-bias ratings. They resolved any disagreements by discussion and consensus or by consulting a third, independent party. Time constraints precluded our contacting study authors for clarification of methodological questions. Appendix D presents risk-of-bias assessments of individual studies included in this review.

Data Synthesis

Throughout this review we synthesized the literature qualitatively. When data were sufficient, we augmented findings with quantitative analyses. We conducted meta-analyses of

data for head-to-head comparisons for trials that were fairly homogenous in study populations and outcome assessments. We also conducted network meta-analyses to compare pharmacologic with nonpharmacological interventions when direct head-to-head evidence was sparse or entirely lacking.

Meta-Analysis of Direct Comparisons

To determine whether quantitative analyses were appropriate, we assessed the clinical and methodological heterogeneity of the studies under consideration following established guidance.⁷¹

For all analyses, we used random- and fixed-effects models to estimate comparative effects. We used DerSimonian & Laird models for random effects analyses. For efficacy, we were able to conduct meta-analyses on four outcomes relating to benefits:

1. the relative risk of achieving response (as defined by authors, most commonly defined as a 50 percent or greater improvement from baseline) on the Hamilton Depression Rating Scale (HAM-D) or the Montgomery-Åsberg Depression Rating Scale (MADRS) at study endpoint
2. the relative risk of achieving remission (as defined by authors, most commonly defined as a HAM-D score of < 7) at study endpoint
3. the weighted mean difference of changes on HAM-D
4. the standardized mean difference of changes in cases where studies used both HAM-D and MADRS

For harms, we conducted meta-analyses on the relative risk of

1. experiencing an adverse event
2. experiencing a serious adverse event,
3. discontinuing treatment,
4. discontinuing treatment because of harms
5. discontinuing treatment because of lack of efficacy, and
6. risk of suicidality

Evidence indicates that no substantial differences in benefits and harms exist among SGAs,³² therefore in all meta-analyses we compared SGAs as a class with other interventions of interest. When we conducted meta-analyses, we assessed statistical heterogeneity in effects between studies by calculating the chi-squared statistic and Cochran's q . We used the I^2 statistic (the proportion of variation in study estimates attributable to heterogeneity) to estimate the magnitude of heterogeneity. We examined potential sources of heterogeneity using sensitivity analysis or analysis of subgroups. For quantitative analyses, we conducted sensitivity analyses including high risk-of-bias studies. Planned stratifications or categories for subgroup analyses included the subgroups listed in the analytic framework (Figure 2).

We assessed publication bias using funnel plots and Kendell's tests. However, given the small number of component studies in our meta-analyses, these tests have low sensitivity to detect publication bias.

We report the results from random-effects models because, in all our meta-analyses, the results from random- and fixed-effects models were very similar. All meta-analyses were conducted using Comprehensive Meta-analysis, version 3.2.

Network Meta-Analyses

Because we were aware of the dearth of studies directly comparing some interventions of interest, we planned a priori with pre-specified criteria to conduct network meta-analyses with a hierarchical frequentist approach using random effects models.^{72,73} Evidence suggests that network meta-analyses agree with head-to-head trials if component studies are similar and treatment effects are expected to be consistent in patients in different trials.⁷⁴ Nevertheless, results have to be interpreted cautiously.

To conduct network meta-analyses, we included all placebo- and active-controlled RCTs that were homogenous in study populations and outcome assessments and were part of a connected network. We built on a database of relevant RCTs of a previous report on the comparative efficacy and safety of SGAs.³² For drugs and most CAM interventions, we included only double-blinded RCTs. For interventions where double blinding was not possible (e.g., psychological intervention or yoga), we required that outcomes assessors had to be blinded. For network meta-analyses, we excluded studies conducted exclusively in subjects who were older than 55 years of age because evidence indicates that older patients have a smaller treatment benefit than younger patients.

Our outcome measure of choice was the rate of response on the HAM-D (defined as a 50 percent improvement of scores from baseline). We recalculated response rates for each study using the number of all randomized patients as the denominator to reflect a true ITT analysis. With this approach, we attempted to correct variations in results of modified ITT analyses encountered in individual studies.

The data provided information on the probability of the response of treatment j out of K possible treatments in study i (p_{ij}). We applied a generalized linear model with random effects. The logit for the random effects model can be expressed as (Hong et al., 2013;⁷² Jones et al., 2011;⁷⁵ Lu & Ades, 2004⁷⁵):

$$\text{logit}(p_{ij}) = \mu_i + \delta_{ij} + \sum_{k=1}^K \frac{\delta_{ik}}{K}$$

where all $\delta_{i1}=1$ and $(\delta_{i2}, \dots, \delta_{ik}) \sim N[(d_2, \dots, d_k), \Sigma]$.

We fit all models using PROC GLIMMIX in SAS version 9.3, specifying a binomial likelihood and logit link function. For ease of interpretation, we present the relative risks and 95 percent confidence intervals of outcomes of interest for all possible comparisons among our treatments of interest.

Strength of Evidence of the Body of Evidence

We graded the strength of evidence based on the guidance established for the EPC Program.⁷⁶ Developed to grade the overall strength of a body of evidence, this approach incorporates five key domains: risk of bias (includes study design and aggregate quality), consistency, directness, precision, and reporting bias. For some scenarios, it also considers other optional domains that may be relevant: a dose-response association, plausible confounding that would decrease the observed effect, and strength of association (magnitude of effect). We asked for input from the TEP to determine minimally important differences, which we used to score assess precision.

Grades reflect the strength of the body of evidence to answer KQs on the comparative benefits and harms of the interventions in this review. Table 7 defines the four grades of strength

of evidence.⁷⁶ Two trained reviewers assessed each domain for each key outcome; differences were resolved by consensus. One of the two reviewers was always a senior researcher with experience in grading strength of evidence. Following GRADE guidance, we graded the strength of evidence for eight outcomes deemed by the TEP and the Key Informants to be of most importance for decisionmaking (see section on outcomes in Inclusion and Exclusion Criteria). Because we found little evidence on overall risk of adverse events, we also graded overall discontinuation rates and discontinuation rates because of adverse events. We used the Guideline Development Tool (<http://www.guidelinedevelopment.org/>) to grade the strength of evidence in a standardized manner and to develop summary of findings tables.

Table 7. Definition of strength of evidence grades

Grade	Definition
High	We are very confident that the estimate of effect lies close to the <i>true effect</i> for this outcome. The body of evidence has few or no deficiencies. We believe that the findings are stable (i.e., another study would not change the conclusions).
Moderate	We are moderately confident that the estimate of effect lies close to the <i>true effect</i> for this outcome. The body of evidence has some deficiencies. We believe that the findings are likely to be stable, but some doubt remains.
Low	We have limited confidence that the estimate of effect lies close to the <i>true effect</i> for this outcome. The body of evidence has major or numerous deficiencies (or both). We believe that additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the <i>true effect</i> .
Insufficient	We have no evidence, we are unable to estimate an effect, or we have no confidence in the estimate of effect for this outcome. No evidence is available or the body of evidence has unacceptable deficiencies, precluding reaching a conclusion.

Applicability

We assessed applicability of the evidence following guidance from the *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*.⁷⁷ We used the PICOTS framework to explore factors that affect applicability. Some factors identified a priori that may limit the applicability of evidence include the following: age of enrolled populations, sex of enrolled populations (e.g., fewer men may be enrolled in some studies), and race or ethnicity of enrolled populations.

Peer Review and Public Commentary

The AHRQ Task Order Officer and an AHRQ associate editor (a senior member of another EPC) reviewed the draft report before peer review and public comment. The draft report (revised as needed) was sent to invited peer reviewers and simultaneously uploaded to the AHRQ Web site where it was available for public comment for 28 days.

We collated all reviewer comments (both invited and from the public) and addressed them individually. We documented all our responses to these comments in a disposition of comments document, which will be posted on the AHRQ EHC program Web site about 3 months after Web publication of the evidence report. The authors of the report have final discretion as to how the report will be revised based on the reviewer comments, with oversight by the Task Order Officer and associate editor.

Results

Introduction

This chapter begins with the results of our literature search and a general description of the included trials. It is then organized by Key Question (KQ 1 through KQ 4). For each KQ, we give an overview, the key points, and more detailed syntheses of the literature organized by intervention comparisons. We also restate the actual issue for that particular KQ.

In each KQ section, we present a table with characteristics of included trials and results of the main outcomes. More details about included trials can be found at the Systematic Review Data Repository (<http://srdhr.gov/>). In Appendix E, we also present “summary of findings” tables that give the main results (effect sizes) for outcomes ranked as critical or important for decisionmaking and the respective strength of evidence (SOE) grades.

Trials that we reviewed reported outcomes data based on an array of commonly used mental health–related measures and assessment tools. Table 8 lists abbreviations of mental health assessment tools encountered in this literature. Important outcomes typically encountered included response to treatment, remission, and changes on depression measures and occasionally quality of life or functional status.

Table 8. Abbreviations and full names of mental health and other assessment tools

Abbreviation	Full Name of Instrument
BDI	Beck Depression Inventory
BDI-II	Beck Depression Inventory-II
HAM-A-X ^a	Hamilton Rating Scale for Anxiety
HAM-D-X ^a	Hamilton Rating Scale for Depression
MADRS	Montgomery-Åsberg Depression Rating Scale
QIDS-SR-X ^a	Quick Inventory of Depressive Symptomatology-Self Report
WAIS-III	Wechsler Adult Intelligence Scale-III

^a X indicates the number of items in the scale.

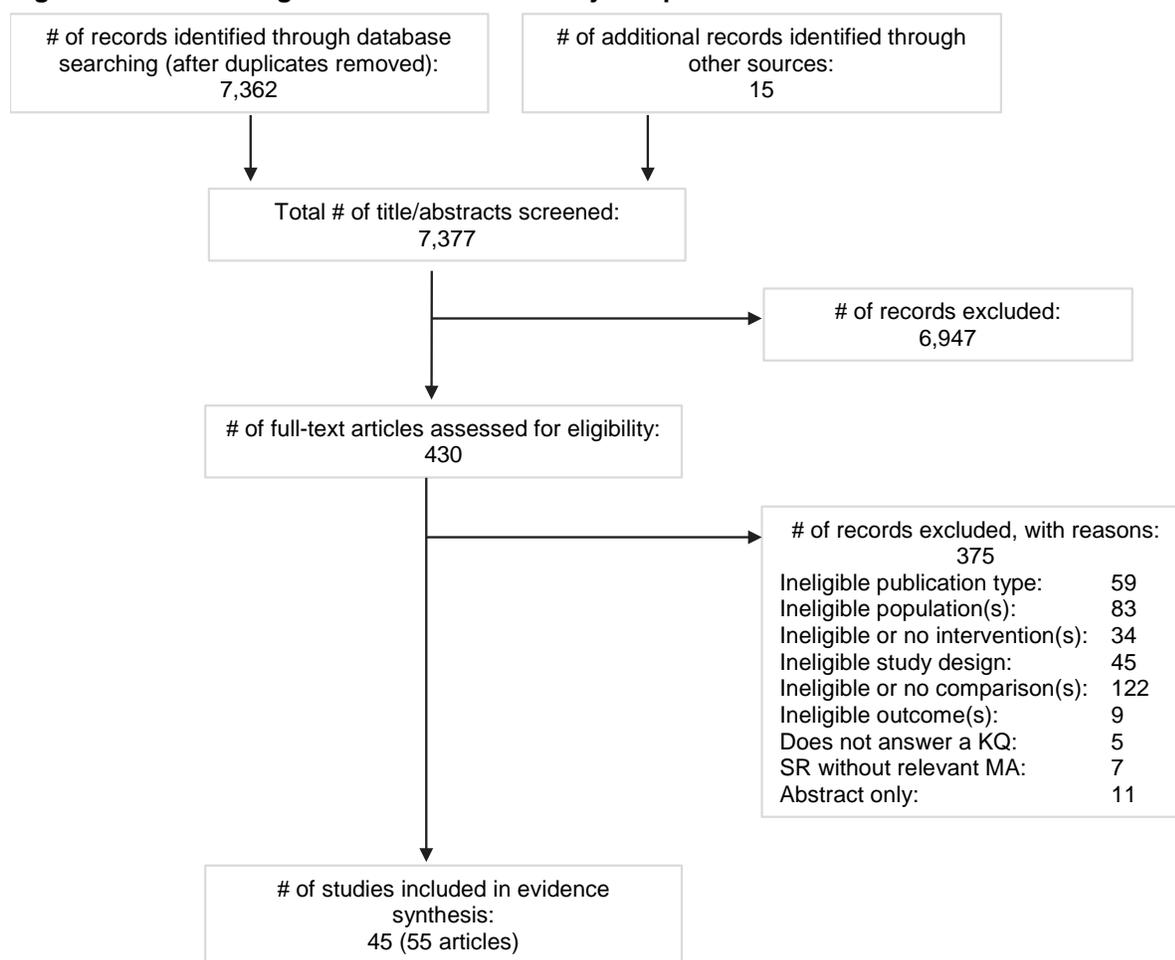
Results of Literature Searches

Our search strategies identified 7,377 possible articles. From that pool, we excluded 6,947 references following independent dual title and abstract review and another 375 references at the full-text review stage. Reasons for exclusion were based on eligibility criteria. Appendix C lists articles excluded during full-text review with reasons for exclusion. Figure 3 documents the disposition of the articles identified from searches.

Description of Included Trials

Overall, we included 45 trials reported in 55 published articles. Of these, 42 trials pertained to KQ 1a and five to KQ 1b. Two trials pertained to KQ 2a, and none was identified for KQ 2b. In addition, of the 45 trials, 44 trials pertained to KQ 3a and one to KQ 3b. Finally, three trials pertained to KQ 4.

Figure 3. PRISMA diagram for treatment of major depressive disorders



KQ = Key Question; MA = meta-analysis; SR = systematic review; PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

To obtain unreported data of interest from included published trials, we sent e-mails soliciting additional data to 31 authors (current contact information for three authors was unavailable). Sixteen authors responded to our query, but many could not provide data because they were no longer available. Ultimately, we obtained additional outcomes data from ten authors.

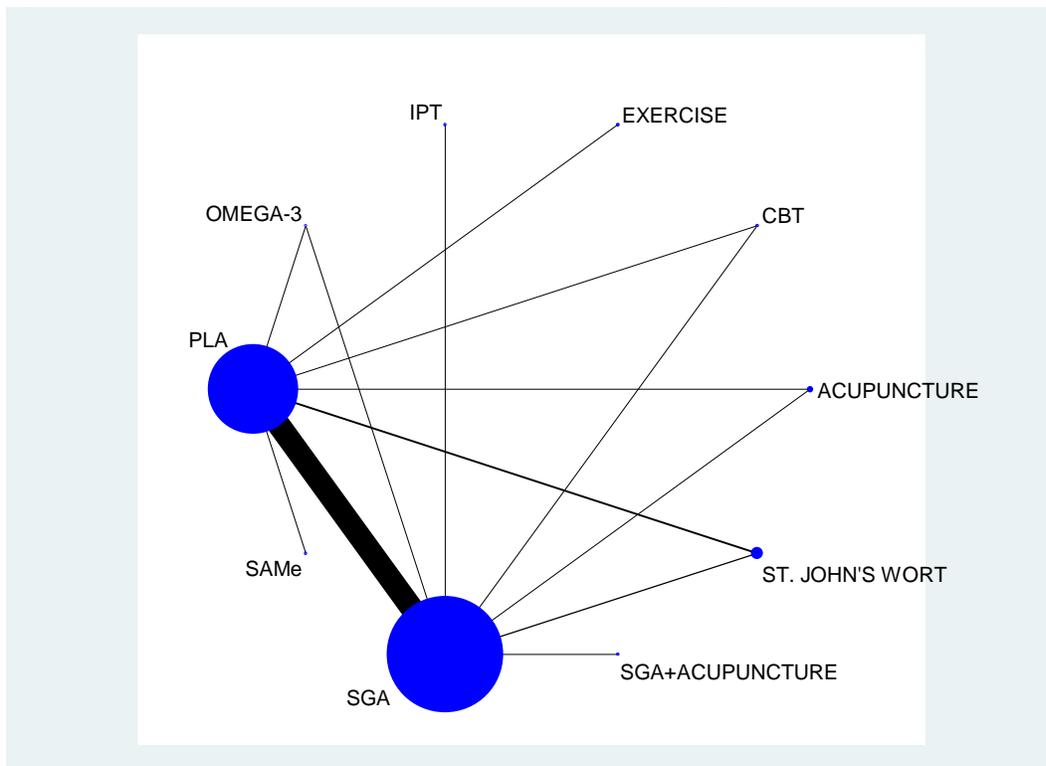
Trials included for this report had various funding sources. The majority of funding came from government agencies and industry sources. Table 9 describes funding sources for each included trial.

Table 9. Reported sources of funding for included trials

Funding Categories	Number of Trials
Government	23 ⁷⁸⁻¹⁰⁰
Industry	16 ^{80,83-85,97-99,101-109}
Academic	3 ¹¹⁰⁻¹¹²
Foundation or nonprofit organization	7 ^{80-82,96,113-115}
Professional organization	0
Funding source not reported	7 ¹¹⁶⁻¹²²

We included data from 124 published and unpublished trials for network meta-analyses. Fourteen of these trials also provide direct evidence for KQ1a; the remaining 110 trials (83 published, 27 unpublished) are included in network meta-analyses only. These trials addressed comparisons of interventions of interest that did not meet eligibility criteria for this report (e.g., SSRIs vs. SNRIs or placebo-controlled trials); they did, however, provide common comparators that we could use for network meta-analyses. Appendix F lists published and unpublished trials included in the network meta-analyses. Figure 4 is a visual presentation of the network of trials included for network meta-analyses. Nodes are weighted according to the number of studies including the respective interventions. Lines represent the available direct comparisons. In this network, SGAs were the most commonly available comparator, followed by placebo (abbreviated as PLA in the figure).

Figure 4. Network of trials included for network meta-analyses



CBT = cognitive behavioral therapy; IPT = Interpersonal psychotherapy; PLA = placebo; SAME = S-Adenosylmethionine; SGA = second-generation antidepressant.

KQ 1: First-Line Therapy: Second-Generation Antidepressants Compared With Nonpharmacologic Therapies

KQ 1a deals with adult patients with acute-phase MDD receiving an initial treatment attempt (also referred to as first-line therapy) with an SGA. It examines the effectiveness of the SGA compared with i) the effectiveness of either nonpharmacological interventions used alone or ii) various combinations of SGAs and one of the nonpharmacological treatments. KQ 1b examines whether treatment effectiveness varies by MDD severity. The nonpharmacologic interventions for this KQ are psychological interventions, complementary and alternative medicine (CAM) interventions, and exercise.

In all, 42 trials comparing SGAs with nonpharmacological treatment options provided direct evidence on acute-phase outcomes (as depicted in Figure 1 in the introduction). Study durations ranged from 4 to 96 weeks. Most patients suffered from moderate to severe major depression. Many of the available trials had serious methodological limitations; few trials reported information on quality of life or functional capacity. We present results from network meta-analyses on response to treatment if we could not find sufficient eligible head-to-head evidence or if direct head-to-head evidence had substantial flaws or limitations (insufficient SOE) and network meta-analyses yielded findings with stronger SOE. For network meta-analyses we utilized 124 placebo- or active-controlled trials; 14 provided direct evidence as well.

Key Points: Second-Generation Antidepressants Compared With Psychological Interventions

- SGAs and cognitive behavioral therapy (CBT) monotherapy led to similar response rates after 8 to 52 weeks of treatment in patients with moderate to severe MDD (five RCTs, moderate SOE); patients treated with SGAs had numerically higher but not significantly different remission rates (three RCTs, low SOE).
- Adding CBT to SGA treatment did not lead to statistically different response and remission rates compared with SGA monotherapies in patients with moderate to severe MDD after 12 to 52 weeks of treatment (two RCTs, low SOE).
- SGAs and integrative therapies (interpersonal psychotherapy [IPT]) did not lead to statistically different response rates (one RCT, low SOE) and remission rates (two RCTs, low SOE) in patients with moderate to severe MDD after 8 to 12 weeks of treatment.
- Adding IPT to SGA treatment resulted in higher remission rates compared with SGA monotherapy in patients with moderate to severe MDD after 12 weeks of treatment (one RCT, low SOE).
- SGAs and psychodynamic therapy (PSYD) monotherapy did not lead to statistically different remission rates in patients with moderate MDD following 16 weeks of treatment (one RCT, low SOE).
- We did not find any eligible trials comparing SGAs with behavior therapies or humanistic therapies (insufficient SOE).

Key Points: Second-Generation Antidepressants Compared With Complementary and Alternative Medicine Interventions

- SGAs and acupuncture monotherapy did not lead to statistically different response rates in patients with severe MDD following 6 weeks of treatment (two RCTs, network meta-analysis, low SOE).
- Adding acupuncture to SGA treatment improved treatment responses compared with SGAs alone in patients with severe MDD after 6 weeks of treatment (2 RCTs, low SOE), but did not lead to statistically different rates of remission (1 RCT, low SOE).
- SGAs led to higher response rates than monotherapy with omega-3-fatty acids in patients with severe MDD (network meta-analysis, low SOE).
- SGAs and S-Adenosyl methionine (SAME) did not lead to statistically different response rates in patients with moderate MDD following 12 weeks of treatment (one RCT, network meta-analysis, low SOE).
- SGAs and St. John's wort monotherapy led to similar response (eight trials, moderate SOE) and remission rates (four trials, moderate SOE) in patients with moderate to severe MDD after 4 to 12 weeks of treatment
- We did not find any eligible trials comparing SGAs with meditation or yoga (insufficient SOE).

Key Points: Second-Generation Antidepressants Compared With Exercise

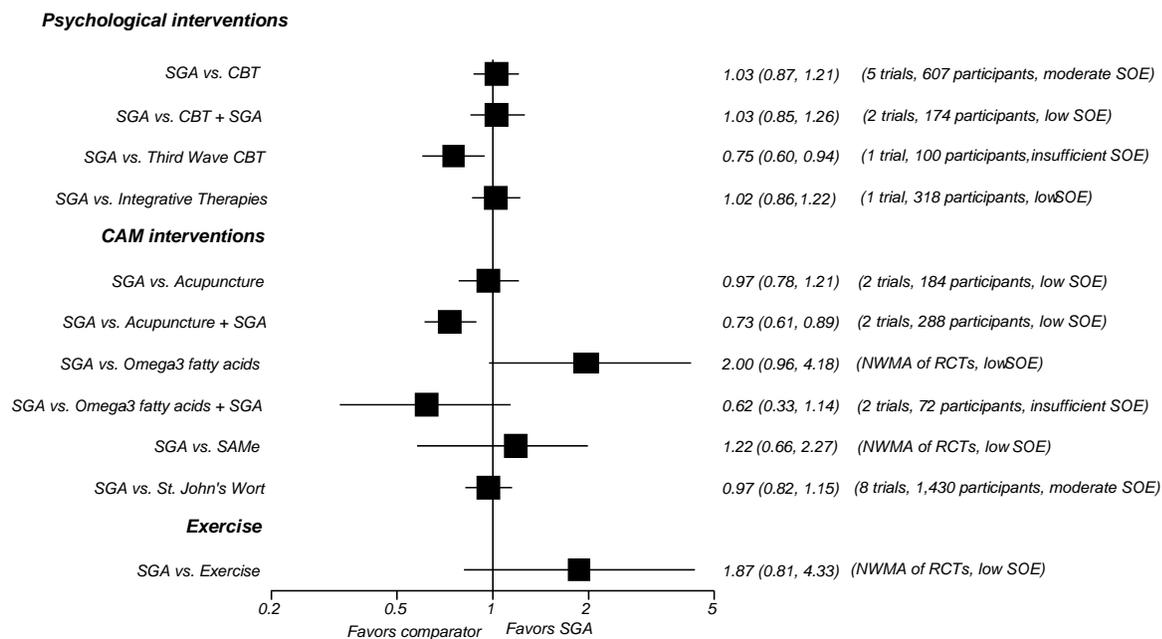
- SGAs and exercise did not lead to statistically different rates of response (network meta-analysis, low SOE) or remission in patients with moderate MDD, following 16 weeks of treatment (two trials, low SOE).
- Adding exercise with to SGA treatment did not lead to statistically different remission rates compared with SGA monotherapy in patients with moderate MDD, following 16 weeks of treatment (one trial, low SOE).

Key Points: Severity as a Moderator of Treatment Effectiveness

- The evidence is inconclusive as to whether the comparative effectiveness of SGAs versus psychological treatments changes as a function of MDD severity (four trials, insufficient SOE).
- The evidence is insufficient to draw conclusions about the effect of severity of disease on the comparative effectiveness SGAs and CAM interventions (one RCT, insufficient SOE).

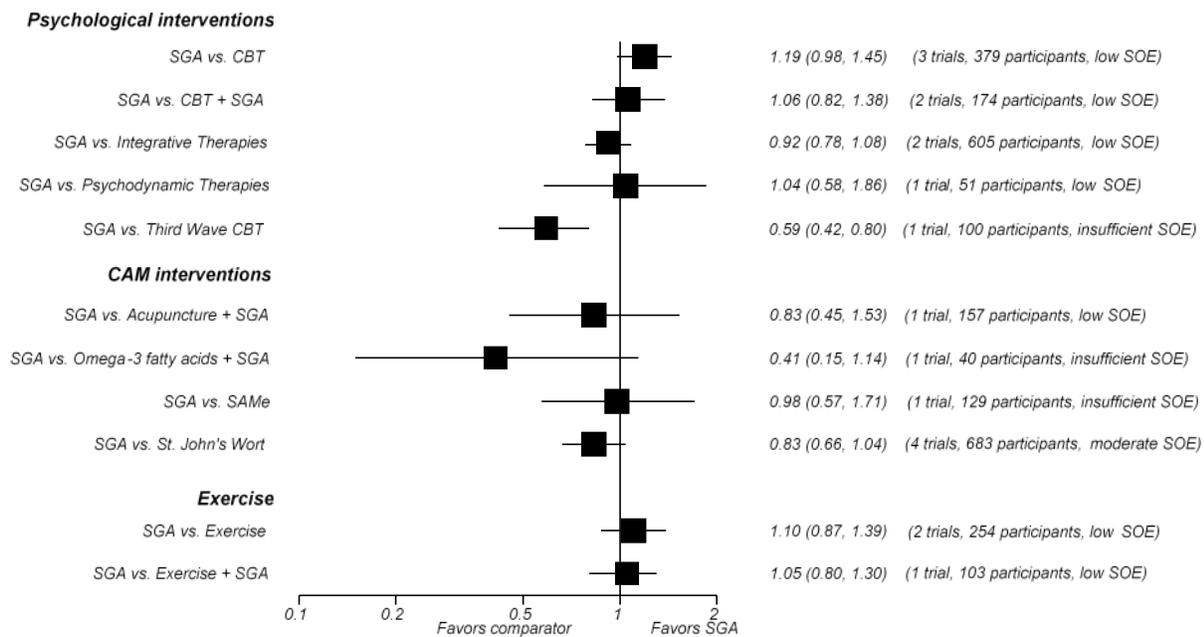
Figures 5 and 6 graphically display relative risks of response and remission rates of SGAs compared with other interventions.

Figure 5. Relative risks of response of SGAs compared with other eligible interventions



CAM = complementary and alternative medicine; CBT = cognitive behavioral therapy; KQ = Key Question; NWMA = network meta-analysis; RCT = randomized controlled trial; SAMe = S-adenosyl-L-methionine SGA = second-generation antidepressant; SOE = strength of evidence; vs. = versus.

Figure 6. Relative risks of remission of SGAs compared with other eligible interventions.



CAM = complementary and alternative medicine; CBT = cognitive behavioral therapy; KQ = Key Question; SAMe = S-adenosyl-L-methionine SGA = second-generation antidepressant; SOE = strength of evidence; vs. = versus.

Detailed Synthesis: KQ 1

In this section, we present findings for both KQs 1a and 1b. The first subsection below (KQ 1a) concerns comparisons of SGAs with various other therapeutic interventions—namely, psychological therapies, CAM interventions, and exercise—as initial options for treating patients with acute-phase MDD (KQ 1a). In all cases, comparisons involve monotherapies for both the SGAs and the alternative interventions. In some cases, the comparisons involve SGA monotherapy with various combinations of SGAs and the alternative. The second subsection below (KQ 1b) examines the question of whether outcomes differ by the severity of MDD.

Table 10 provides the number of included trials by eligible comparison. We included any trial that met eligibility criteria, regardless of the risk of bias rating. In our syntheses, however, we place more emphasis on trials with low or medium risk of bias because of the presumed higher certainty of findings. In Appendix E we present “summary of findings” tables of important outcomes. These tables are intended for guideline development and give basic information on the available evidence, show absolute and relative effect measures, and present SOE grades for outcomes that the TEP and key informants deemed as most important for decisionmaking.

Table 10. Number of included trials by type of comparison

Comparison Category	Comparisons for KQ 1	Number of Trials and Citations
SGA vs. Psychological interventions	SGA vs. Behavior therapies/behavior modification	0
	SGA vs. CBT	12 ^{78,82,85-89,91,102,111,113,116}
	SGA vs. Humanistic therapies	0
	SGA vs. Integrative therapies	4 ^{80,83,84,92}
	SGA vs. Psychodynamic therapies	4 ^{79,81,90,101}
	SGA vs. Third-wave CBTs	2 ^{86,110}
SGA vs. CAM	SGA vs. Acupuncture	5 ^{94,96,114,117,118}
	SGA vs. Omega-3 fatty acids	2 ^{95,112}
	SGA vs. SAMe	1 ⁹³
	SGA vs. St. John's wort	12 ^{97,103-108,119-123}
	SGA vs. Meditation	0
SGA vs. Exercise	SGA vs. Yoga	0
	SGA vs. Exercise	2 ^{98,99}

CAM = complementary and alternative medicine; CBT = cognitive behavioral therapy; KQ = Key Question; SAMe = S-adenosyl-L-methionine SGA = second-generation antidepressant.

KQ 1a: Second-Generation Antidepressants Compared With Psychological Interventions

In this section, we categorize types of psychotherapy according to the Cochrane Collaboration Depression, Anxiety and Neurosis Review Group (CCDAN) classification system (see Appendix B).⁶³ We address CBT, integrative therapies, psychodynamic therapies, and third-wave CBTs. Most of these trials compare monotherapies; when relevant, we also present information about an SGA monotherapy with some form of a combination of SGA and the relevant psychological treatment.

We are aware of a new trial that likely meets inclusion criteria that was published while we were developing this draft.¹²⁴ We will evaluate that trial and, if appropriate, incorporate its results into this review before it is published.

Description of Included Trials

In all, 20 primary RCTs (reported in 24 articles) compared SGAs with a psychological treatment and provided data for KQ 1a. Trials are grouped according to the type of psychotherapy compared with the SGA. They are listed within this chapter's tables first by subtype of psychotherapy (if applicable) and then alphabetically by SGA. We found no trials eligible for KQ 1a that compared an SGA with behavior therapy or behavior modification or with humanistic therapies.

Five trials^{81,83,89,92,113} were conducted in primary care settings; the remainder took place in mental health care locations. Most trials were funded by the government; seven trials^{80,83-85,89,101,102} received at least partial funding from the pharmaceutical industry. Six trials^{78,79,83,88,89,116} took place solely in the United States; other countries included Brazil,⁹⁰ Canada,^{85,87,102} England,¹¹³ Finland,⁸¹ Germany,⁹¹ Iran,^{110,111} Italy,⁹² Romania,⁸² and The Netherlands.^{84,101} One trial was conducted in both the United States and Italy.⁸⁰

Generally, patients were between 18 and 65 years of age; most trials reported a mean age between 35 and 45 years. In all trials, the majority of patients were female. One trial enrolled only women.⁸⁹ In the few trials that reported race or ethnicity, three^{79,83,89} included more than 33 percent nonwhite patients. All trials reported mean baseline depressive severity of at least a moderate degree; most trials reported mean baseline HAM-D-17 scores between 16 (moderate depression) and 23 (severe). The total daily dose of each SGA medication was within the usual ranges prescribed for adults.

Second-Generation Antidepressants Compared With Behavior Therapies/Behavior Modification Therapies

We found no eligible trials that compared an SGA with behavior therapy/behavior modification.

Second-Generation Antidepressants Compared With Cognitive Behavioral Therapy

Table 11 describes the 11 included trials (13 publications) of an SGA compared with a CBT (grouped by therapy subtype and in alphabetical order by first author). Six trials employed CBT,^{78,85,87,89,91,102} four used cognitive therapy (CT),^{82,88,111,116} and one each used problem solving therapy (PST)¹¹³ and rational emotive behavior therapy (REBT).^{82,125} Trial counts exceed 11 because one trial had both CT and REBT arms.⁸² All but one trial compared SGA monotherapy with CBT alone; Lam and colleagues compared SGA monotherapy with SGA plus CBT.¹⁰² Two trials included an additional comparison of SGA monotherapy with a combination of SGA and CBT.^{111,113} Treatment duration ranged from 8 weeks to 1 year; some trials also reported follow-up results once patients were off-treatment.

One trial was rated overall low risk of bias,¹⁰² five were rated medium risk,^{82,88,89,113,116} and five trials were rated high.^{78,85,87,91,111} Reasons for high risk of bias ratings included high attrition without proper handling of missing data, potentially meaningful differences in baseline characteristics between treatment groups, potential reporting bias, and little or no information on randomization and allocation procedures. In two cases, we applied a second risk of bias rating for specific outcomes: one medium-risk trial⁸⁸ was rated high for change in HAM-D score, and one overall high-risk trial⁹¹ was rated medium for remission and response because we could use data from the full sample for those outcomes.⁹¹ Full risk of bias assessments for included trials are found in Appendix D.

Table 11. Second-generation antidepressants versus cognitive behavioral therapy: Trial characteristics, main outcomes, and risk of bias ratings

Trial and Type of Psychotherapy	N	Total Sample Mean Baseline Severity	SGA Type: mg/day Type of Psychotherapy: Number of Sessions	Response ^a and Significance Level	Remission ^a and Significance Level	Mean Change in HAM-D Score from Baseline and Significance Level	Risk of Bias Rating
David et al., 2008 ⁸²	112	HAM-D-17: 22.5	Fluoxetine: 40 to 80	At 14 weeks: 58% vs. 63% p>0.05	At 14 weeks: 47% vs. 50% p>0.05	-12.6 vs. -14.3 p>0.05	Medium
Sava et al., 2009 ¹²⁵	14 treatment; 36 follow-up		CT: 20				
CT							
David et al., 2008 ⁸²	113	HAM-D-17: 22.5	Fluoxetine: 40 to 80	At 14 weeks: 58% vs. 65% p>0.05	At 14 weeks: 47% vs. 44% p>0.05	-12.6 vs. -14.3 p>0.05	Medium
Sava et al., 2009 ¹²⁵	14 treatment; 36 followup		REBT: 20				
REBT							
DeRubeis et al., 2005 ⁸⁸	180	HAM-D-17: 23.4	Paroxetine: 10 to 50	50% vs. 43% p=0.40	NR	Effect size estimate: 0.16 (favors SGA) p=0.46NR	Medium for response and remission; high for change in HAM-D ⁱ
Leykin et al., 2007 ¹²⁶	8 ^h		CT: 20 to 28				
CT							
Hegerl, 2010 ⁹¹	48	HAM-D-17: 16.1	Sertraline: 50 to 200	38% vs. 50% p=NR	NR	-6.5 vs. -8.8 p=NR	Medium for response and remission; high for change in HAM-D
	10		CBT: 14				
CBT							
Kennedy et al., 2007 ⁸⁵	31	HAM-D-17: 20.5	Venlafaxine: 75 to 225	64% vs. 41% p=NR	57% vs. 29% p=NR	-12.9 vs. -10.8 p=NR	High ^b
	16		CBT: 16				
CBT							
Lam et al., 2013 ¹⁰²	105	MADRS: 27.6	Escitalopram: 10 to 20	61% vs. 63% p=0.86	53% vs. 56% ^c p=0.74	MADRS: -14.3 vs. -15.7 p=0.60	Low
	12		CBT (via telephone): 8 + escitalopram: 10 to 20				
CBT							
Landenberger et al., 2002 ¹¹⁶	92	HAM-D-17: 23.3	Paroxetine: 10 to 50	75% vs. 62% ^j p=0.20	67% vs. 50% p=NR	NR	Medium
	16		CT: 24				
CT							
McGrath et al., 2013 ⁷⁸	82	HAM-D-17: 18.8	Escitalopram: 10 to 20	60% vs. 57% p=NR	28% vs. 29% p=NR	NR	High ^d
	12		CBT: 16				
CBT							

Table 11. Second-generation antidepressants versus cognitive behavioral therapy: Trial characteristics, main outcomes, and risk of bias ratings (continued)

Trial and Type of Psychotherapy	N	Total Sample Mean Baseline Severity	SGA Type: mg/day Type of Psychotherapy: Number of Sessions	Response ^a and Significance Level	Remission ^a and Significance Level	Mean Change in HAM-D Score from Baseline and Significance Level	Risk of Bias Rating
Mynors-Wallis et al., 2000 ¹¹³	151	HAM-D-17: 20.3	Fluvoxamine: 100 to 150 or Paroxetine: 10 to 40	At 12 weeks: 78% vs. 64% vs. 69% vs. 74% p=NR	At 12 weeks: 67% vs. 51% vs. 54% vs. 60% p=NR	-14.0 vs. -12.0 vs. -11.8 vs. -12.3 p>0.05	Medium
PST			PST (provided by GP): 6 PST (provided by nurse): 6 PST (provided by nurse): 6 + fluvoxamine: 100 to 150 or paroxetine: 10 to 40				
Segal et al., 2006 ⁸⁷	301	HAM-D-17: 19.5	Sertraline: 50 to 200 or paroxetine: 20 to 50 or venlafaxine: 75 to 225 CBT: 20	At 24 weeks: 80% vs. 72% p=NR	At 24 weeks: ^e 71% vs. 61% p=NR	NR	High ^f
CBT	24 treatment; 96 followup						
Shamsaei et al., 2008 ¹¹¹	120	BDI: 42.8	Citalopram: 20	NR	NR	NR	High ^k
CT	8		CT: 8				
WECare ⁸⁹	178	HAM-D (version NR): 16.9	Paroxetine: 10 to 50 CBT: 8	NR	NR	-5.0 vs. -2.1 p=0.17	Medium
CBT	4 ^g						

^a Response (≥ 50 percent decrease in depressive severity) and remission (as defined by authors of individual trials) are measured using the HAM-D unless indicated otherwise.

^b High attrition, completers analysis, difference in baseline age between groups.

^c Response was defined as ≥ 50 percent decrease in MADRS; remission was defined as MADRS ≤ 12 .

^d High attrition, completers analysis, no baseline data for part of the population.

^e Definition of response was not reported.

^f Very high attrition, completers analysis, unclear randomization method.

^g Although patients received SGA for 8 weeks, only the 4-week time point was reported.

^h Nonresponders were switched to another pharmacotherapy at 8 weeks.

ⁱ For dropouts, only the data gathered prior to attrition were used in continuous outcome models.

^j Response was defined as "sufficiently low symptom level."

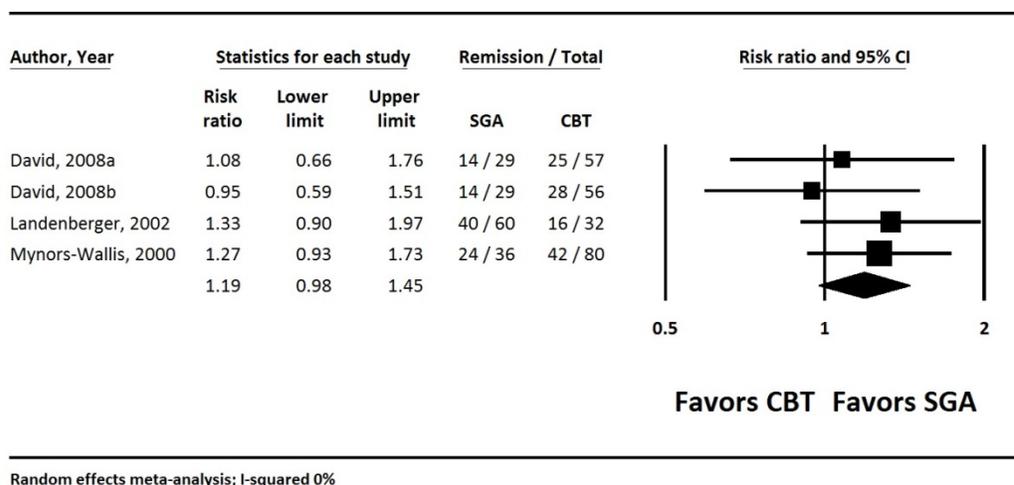
^k Several important aspects of study design and analysis not reported. BDI = Beck Depression Inventory; CBT = cognitive behavioral therapy; CT = cognitive therapy; GP = general practitioner; HAM-D = Hamilton Depression Rating Scale; MADRS = Montgomery-Åsberg Depression Rating Scale; N = number; NR = not reported; PST = problem solving therapy; REBT = rational emotive behavior therapy; SGA = second-generation antidepressant; vs. = versus.

Second-Generation Antidepressant Versus Cognitive Behavioral Therapy: Monotherapy Comparisons

We conducted random-effects meta-analyses of trials rated low or medium risk of bias for three outcomes: (1) remission (three trials [four comparisons];^{82,113,116} 379 patients), (2) response (five trials [six comparisons];^{82,88,91,113,116} 607 patients), and (3) change in HAM-D-17 score (three trials [four comparisons];^{82,89,113} 427 patients). We also performed sensitivity analyses including three additional trials rated high risk of bias (414 patients).^{78,85,87}

For remission, we included results measured between 12 and 16 weeks; all trials compared an SGA with CBT. The effect size (risk ratio [RR]) favored SGAs but was not significantly different from CBT (57 percent versus 50 percent; RR, 1.19; 95% CI, 0.98 to 1.45; Figure 7). We found similar results when we stratified by subtype of CBT (CT versus PST versus REBT). Our sensitivity analysis included one additional SSRI trial,⁷⁸ a trial of an SNRI (venlafaxine),⁸⁵ and a trial that allowed patients to receive either an SSRI or an SNRI.⁸⁷ Our sensitivity analysis yielded an overall significant difference favoring the SGAs (55 percent versus 45 percent; RR, 1.18; 95% CI, 1.04 to 1.33), owing largely to the single high risk of bias trial⁸⁷ with a large sample size (N=301) that allowed use of either an SSRI or an SNRI. Reasons for that trial's rating included very high attrition and very few details on trial methods.

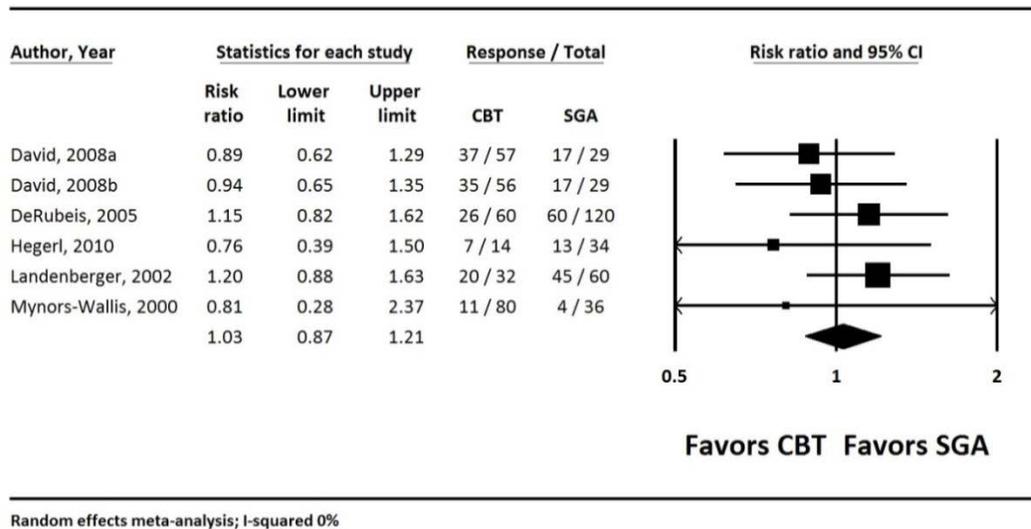
Figure 7. SGA versus cognitive behavioral therapy: Remission



CBT = cognitive behavioral therapy; CI = confidence interval; SGA = second-generation antidepressant;

For response, we included results measured between 8 and 16 weeks. Treatment effects were similar for SGAs and CBT (48 percent versus 49 percent; RR, 1.03; 95% CI, 0.87 to 1.21; Figure 8). We found similar results when we stratified by subtype of CBT and by time point (<12 weeks versus 12 to 16 weeks). The sensitivity analysis did not yield a statistically significant difference in response between SGAs and CBT.

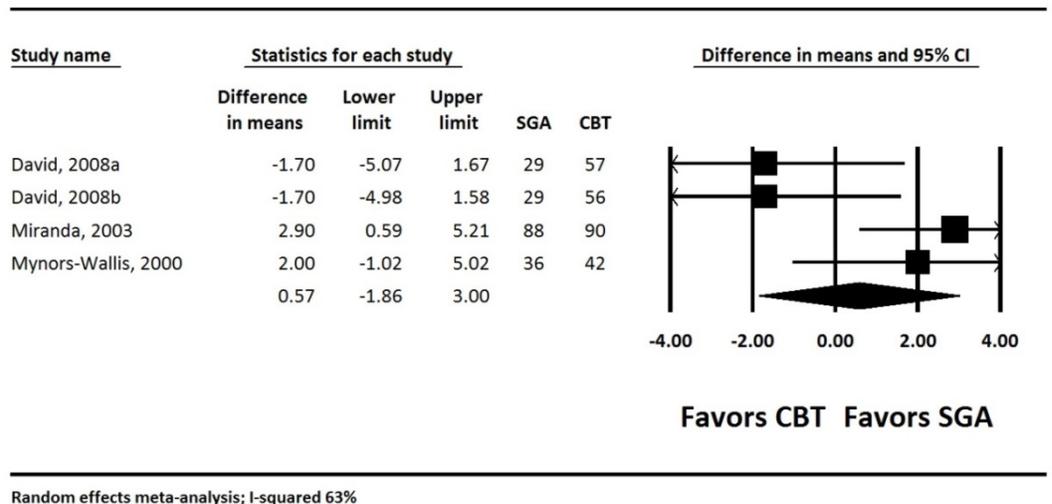
Figure 8. SGA versus cognitive behavioral therapy: Response



CBT = cognitive behavioral therapy; CI = confidence interval; SGA = second-generation antidepressant.

Our weighted mean difference analysis of change in HAM-D-17 scores found no statistically significant difference between SGAs and CBT (-11.0 versus -10.7; WMD, 0.57; 95% CI, -1.86 to 3.00; Figure 9), although heterogeneity was somewhat high ($I^2 = 63$ percent). Potential sources of heterogeneity include the number of psychotherapy sessions received at the time point reported (ranging from 4⁸⁹ to at least 14⁸²), variation between CBT subtypes (included trials used CBT,⁸⁹ CT,⁸² PST,¹¹³ and REBT⁸²), and type of provider delivering the psychotherapy (general practitioner¹¹³ versus psychologists or psychiatrists^{82,89}). Removing the trial that reported data at 4 weeks reduced the I^2 to 44 percent. Further sensitivity analyses by those factors were not possible owing to too few trials. Adding the high risk of bias trials to the model yielded no difference in comparative effectiveness.

Figure 9. SGA versus cognitive behavioral therapy: Change in HAM-D-17



CBT = cognitive behavioral therapy; CI = confidence interval; HAM-D = Hamilton Depressive Scale; SGA = second-generation antidepressant.

Two trials, both rated medium risk of bias, reported depression results at time points beyond 16 weeks. In one,⁸² patients receiving either REBT or CT reported higher rates of remission and response at 6 months than patients taking fluoxetine, although neither difference was statistically significant. At 6 months, patients receiving REBT or CT reported significantly lower HAM-D-17 scores than the patients taking fluoxetine. In the trial that compared either fluvoxamine or paroxetine with PST,¹¹³ rate of remission at 1 year was higher in the PST arms, although rate of response at 1 year was higher in the SGA arm. In that trial, patients' HAM-D-17 scores continued to decline, with 1-year scores being lower in the PST arms than the SGA arm. Again, these differences failed to reach statistical significance.

With respect to other health outcomes, two trials reported relapse rates during off-treatment followup.^{82,87} In the medium risk of bias trial,⁸² 11 percent of patients treated with fluoxetine relapsed within 6 months, compared with 2 percent and 6 percent of patients treated with REBT and CT, respectively. In the trial rated high risk of bias,⁸⁷ 48 percent and 39 percent of remitted patients relapsed within 18 months.

The single trial that reported measures of functional capacity used the Social Adjustment Scale;¹¹³ SGA and PST did not differ at end of treatment or at 40-week off-treatment followup.

Second-Generation Antidepressant Versus Cognitive Behavioral Therapy: Combination Comparisons

The three trials that compared SGA monotherapy with a combination of SGA and CBT reported no statistically significant between-group differences in rates of either remission or response.^{102,111,113} Table 11 also presents effect estimates and the respective SOE grades for response and remission. All three trials reported change in depression scale score between baseline and endpoint, but only one¹¹¹ reported a significant between-group difference—namely, a smaller decrease in scores on the Montgomery-Åsberg Depression Rating Scale (MADRS) for patients on citalopram alone compared with patients treated with citalopram plus CT. This trial, however, was rated high risk of bias, whereas the other two were rated low¹⁰² and medium¹¹³ risk of bias.

The trial that compared escitalopram alone with escitalopram plus telephone CBT measured several work-related outcomes.¹⁰² Patients receiving the combination of escitalopram and telephone CBT reported greater improvement on three of four work functioning measures. The authors reported found no between-group differences in reduction of hours of work missed, although both groups reported a decrease at the end of treatment. In the trial that compared SGA alone with the combination of SGA and PST, there was no between-group difference in the Social Adjustment Scale at end of treatment or at 40-week off-treatment followup.¹¹³

Second-Generation Antidepressants Compared With Humanistic Therapies

We found no eligible trials that compared an SGA with humanistic therapies.

Second-Generation Antidepressants Compared With Integrative Therapies

Table 12 describes the four included trials (five publications) of an SGA compared with one specific type of integrative therapy—IPT.^{80,83,84,92,127} One trial also included a combination SGA+IPT arm.⁸⁴ Two trials took place outside the United States;^{84,92} two were conducted in outpatient primary care clinics.^{83,92} Three of the four trials received a combination of industry and government funding.^{80,83,84,127}

Table 12. Second-generation antidepressants versus interpersonal psychotherapy: Trial characteristics, main outcomes, and risk of bias ratings

Trial	N	Total Sample Mean Baseline Severity	SGA Type: (mg/day) Type of Psychotherapy: Number of Sessions	Response ^a and Significance Level	Remission ^a and Significance Level	Mean Change in HAM-D Score from Baseline and Significance Level	Risk of Bias Rating
Blom et al., 2007 ⁸⁴	207	HAM-D-17: 20.1	Nefazodone: 400 to 600 IPT: 12 Nefazodone: 400 to 600 + IPT: 12	NR	Nefazodone + IPT vs. nefazodone: OR (95% CI) 3.22 (1.02 to 10.12) Other comparisons NR p>0.10	-5.4 vs. -6.9 vs. -8.1 p=NR	Medium
Frank et al., 2011 ⁸⁰ Rucci, 2011 ¹²⁷	318 12	HAM-D-17: 20.0	Escitalopram: 10 to 20 IPT: NR	At 6 weeks: 62.7% vs. 61.3% p=NR At 12 weeks: NR	At 12 weeks: 46.8% vs. 42.5% p=NR	NR	High ^b
Menchetti et al., 2014 ⁹²	287	HAM-D-21: 17.3	Citalopram: 10 to 60 or Sertraline: 25 to 200 IPT: 6 to 8	NR	45% vs. 59% p=0.021	NR	Medium
Raue et al., 2009 ⁸³	60	HAM-D-24: 23.7	Escitalopram: 10 to 20 IPT: 14	NR	At 12 weeks: NR; p=NS At 24 weeks: 18.9 vs. 14.0 p=0.05	NR	High ^c

^a Response (≥ 50 percent decrease in depressive severity) and remission (as defined by authors of individual trials) are measured on the Hamilton Depression Rating Scale.

^b No methods of randomization/allocation reported, unclear if outcome assessors were masked, and median duration of illness was much higher in SGA arm (10.8 years) than in IPT arm (3.5 years).

^c Very little information provided about procedures/methods, randomization was to a treatment by way of preference congruence.

CI = confidence interval; HAM-D = Hamilton Depression Rating Scale; IPT = interpersonal psychotherapy; mg/day = milligram per day; N = number; NR = not reported; OR = odds ratio; SGA = second-generation antidepressant; vs. = versus.

Patients ranged between 18 and 66 years of age, and the samples comprised at least 72 percent females. Trial enrollment ranged from 60 to 318 patients. Treatment duration ranged from 8 to 24 weeks. None of the trials reported posttreatment followup results. The two trials rated high risk of bias provided few details about trial methods.^{80,83} Full risk of bias assessments for included trials are found in Appendix D.

Second-Generation Antidepressants Versus Interpersonal Psychotherapy: Monotherapy Comparisons

Of the four trials that met eligibility criteria, two trials, one medium risk of bias⁹² and one high risk of bias,⁸⁰ reported rates of remission. In the medium risk of bias trial, remission at 2 months was significantly lower in the SGA group (45 percent) than in the IPT group (59 percent; p=0.021). This trial reported no other main outcomes.

We did not find enough trials to pool data for any depression outcomes. Our network meta-analysis yielded a relative risk of response that indicated similar treatment effects between SGAs and IPT (RR, 1.01; 95% CI, 0.63 to 1.6).

Second-Generation Antidepressants Versus Interpersonal Psychotherapy: Combination Comparisons

In the sole trial that compared SGAs with a combination of SGA and IPT (N=97), rated medium risk of bias, nefazodone alone was associated with a significantly lower odds ratio (OR) of remission than the combination of nefazodone and IPT at 8 weeks, although the 95% CI was very wide (low SOE, small sample size, very wide CI).⁸⁴ The combination was also associated with a greater decrease in the HAM-D-17 at 12 weeks than either therapy alone (presumably not significant, p not reported); also, the reported result does not meet the minimum clinically meaningful difference of 3 points advocated by the National Institute of Health and Care Excellence.³⁶

Second-Generation Antidepressants Compared With Psychodynamic Therapies: Monotherapies or Combinations

Table 13 describes the four included trials (five articles) of an SGA compared with PSYD of various sorts.^{79,81,90,101,128} Of these four trials, one included an additional treatment arm that combined fluoxetine and PSYD.⁹⁰ One trial took place in the United States;⁷⁹ three were conducted in outpatient psychiatry clinics,^{79,90,101} and one was conducted in a primary care setting.^{81,128} Three trials were funded in part by a government agency.^{79,81,90}

Subjects ranged in age between 18 and 66 years of age; the samples comprised at least 72 percent females. Trial enrollment ranged from 51 to 272 patients. Treatment duration ranged from 8 weeks to 24 months. All four trials were rated medium risk of bias.

One trial reported rate of remission as measured by either the HAM-D-17 or criteria specified by the *Diagnostic and Statistical Manual*, fourth edition (DSM-IV);⁸¹ treatment groups did not differ significantly. In the one trial that reported response rate,⁷⁹ 62 percent of sertraline patients responded to treatment at 8 weeks, but the response rate for PSYD patients was not reported. (In that trial, nonresponders to sertraline were switched to a different medication at week eight, but no such switch in treatment was made in the psychotherapy arm.) Therefore, we are unable to report results for second medication in the latter.

Two trials reported changes in HAM-D-17.^{81,101} In both, HAM-D-17 scores decreased more for SGA patients than for PSYD patients; the difference, however, was statistically significant in only one (-4.2 versus -2.0; p=0.04).¹⁰¹ A third trial measured depressive symptoms with the Beck Depression Inventory (BDI), but results had not been published at the time of this draft report.⁹⁰

Two trials^{81,90} reported measures of functional and/or neuropsychological capacity. In one,⁸¹ both the fluoxetine and PSYD groups improved significantly on the Social and Occupational Functioning Assessment Scale, but the between-group difference was not significant. In the same trial, the proportion of patients on sick leave at 16 weeks was higher in the SGA group than in the PSYD group (12 percent versus 4 percent), although the difference was not statistically significant. One study measured several domains of the Wechsler Adult Intelligence Scale, third edition (WAIS-III) at time points between 6 and 24 months.⁹⁰ Few statistically significant between-group differences were reported, all of which favored PSYD.

In a comparison between fluoxetine monotherapy and fluoxetine plus long-term PSYD,⁹⁰ the combination group improved significantly from baseline in the following subtests: digit span, letter-number sequencing, digit-symbol coding, matrix reasoning, and picture arrangement. Effects on WAIS-III measures were similar for SGA and the combination of SGA and PSYD.

Table 13. Second-generation antidepressants versus psychodynamic therapies: Trial characteristics, main outcomes, and risk of bias ratings

Trial	N	Total Sample Mean Baseline Severity	SGA Type: (mg/day) Type of Psychotherapy: Number of Sessions	Response ^a and Significance Level	Remission ^a and Significance Level	Mean Change in HAM-D Score from Baseline and Significance Level	Risk of Bias Rating
Barber et al., 2012 ⁷⁹	106 8 ^b	HAM-D-17: 19.4	Sertraline: 50 to 200 Supportive-expressive therapy: 20	At 8 weeks: 61.8% vs. NR p=NR	NR	NR	Medium
Bastos et al., 2013 ⁹⁰	272 96	BDI: 26.8	Fluoxetine: 20 to 60 Long-term psychodynamic psychotherapy: weekly Fluoxetine: 20 to 60 + long-term psychodynamic psychotherapy: weekly	NR	NR	NR	Medium
Dekker et al., 2008 ¹⁰¹	141 8	HAM-D-17: 20.1	Venlafaxine: 75 to 225 Short-term psychodynamic supportive psychotherapy: 16	NR	NR	-4.21 vs. -2.01 p=0.039	Medium
Salminen et al., 2008 ⁸¹	51 16	HAM-D-17: 18.6	Fluoxetine: 20 to 40 Short-term psychodynamic supportive psychotherapy: 16	NR	48% vs. 46% p=NR	-11.2 vs. -11.0 p=0.87	Medium

^a Response and remission (as defined by authors of individual trials) are measured on the Hamilton Depression Rating Scale (HAM-D.) the BDI.

^b Treatment duration was 16 weeks, but only the week 8 results are relevant for this key question.

BDI = Beck Depression Inventory; HAM-D = Hamilton Depression Rating Scale; mg/day = milligram per day; N = number; NR = not reported; SGA = second generation antidepressant; vs. = versus.

Second-Generation Antidepressants Compared With Third-Wave Cognitive Behavioral Therapy

One high risk of bias trial compared an SGA (sertraline) with a third-wave CBT (namely, 16 sessions of behavioral activation).¹¹⁰ It took place in an outpatient psychiatry clinic in Iran over 49 weeks and received funding from two academic institutions. The sample of 100 patients was 85 percent female (see Table 14).

Among study completers at 13 weeks, over 90 percent of patients in both treatment groups reported response (between-group p=0.42). Significantly fewer patients taking sertraline were in remission at 13 weeks, compared with patients receiving BA CBT (69 percent versus 91 percent; p<0.01). However, if one assumes that trial dropouts failed to respond, then rates of response are 66 percent for SGA and 88 percent for BA CBT. With the same assumption for remission, the

rates are 48 percent and 82 percent, respectively. Among treatment completers at 13 weeks, the treatment groups differed significantly on remission rates and change in HAM-D-17 score outcomes (mean between-group difference=3.1; $p<0.01$).

At the 49-week followup, roughly half as many SGA patients as behavioral activation (BA) CBT patients reported at least a 50 percent reduction in symptoms (47 percent versus 89 percent; $p<0.01$). Similarly, fewer than half the number of SGA patients than BA CBT patients were in remission at 49 weeks (28 percent versus 66 percent; $p<0.01$). If one assumes that trial dropouts failed to remit, then rates of remission are 24 percent for SGA and 58 percent for BA CBT. With the same assumption for response, the rates are 40 percent and 78 percent, respectively. Among patients who were in remission at 13 weeks, more SGA patients relapsed during 49 weeks of followup than BA CBT patients (36 percent versus 23 percent; $p=0.02$).

Table 14. Second-generation antidepressants versus third-wave cognitive behavioral therapy: Trial characteristics, main outcomes, and risk of bias ratings of trials

Trial	N	Total Sample Mean Baseline Severity	SGA Type: mg/day and Type of Psychotherapy: Number of Sessions	Response ^a and Significance Level	Remission ^a and Significance Level	Mean Change in HAM-D Score from Baseline and Significance Level	Risk of Bias Rating
Moradveisi, 2013 ¹¹⁰	100	HAM-D-17: 21.4	Sertraline 100	At 13 weeks 98% vs. 94% $p=0.42$	At 13 weeks 69% vs. 91% $p<0.01$	At 13 weeks -14.2 vs. -17.3 $p<0.01$	High ^b
	49		BA 16				

^a Response was defined as at least a 50 percent reduction from baseline on both the HAM-D and the BDI-II. Remission was defined as scores of less than 8 on the HAM-D and less than 11 on the BDI.

^b High attrition.

BA = behavioral activation; BDI = Beck Depression Inventory; HAM-D = Hamilton Depression Rating Scale; N = number; SGA = second-generation antidepressant; vs. = versus.

KQ 1a. Second-Generation Antidepressants Compared with Complementary and Alternative Medicine Interventions

Description of Included Trials

We evaluated four CAM therapies: acupuncture, omega-3 fatty acids, SAMe, and Saint John's wort. All involved a comparison of an SGA with the CAM therapy of interest as monotherapy. When data were available, we also included an evaluation of an SGA with a combination of a CAM therapy plus an SGA. For all reports, the SGA was an SSRI; however, the term *SGA* has been used throughout for consistency. We defined acupuncture broadly to include techniques provided by trained practitioners that provide stimulation to meridian points using traditional needles. We elected to group trials of manual and electroacupuncture together because of the paucity of publications in this area and the uncertainty surrounding any meaningful differences between the two techniques for treating patients with depression.

We identified 20 primary RCTs (22 articles) comparing an SGA with a CAM therapy for treating patients with MDD.^{93-97,103-108,112,114,115,117-123,129} Five trials (six articles) evaluated acupuncture (503 participants), two trials evaluated omega-3 fatty acids (102 participants), one trial evaluated SAMe (189 participants), and 12 trials (13 articles) evaluated St. John's wort (1,855 participants). About one-half of the trials (11 of 20) compared fluoxetine with a CAM therapy. Other SGAs involved sertraline (3 trials), paroxetine (2), citalopram (2), and escitalopram (1). Trials enrolled participants according to a criteria-based diagnosis of MDD

based on the DSM-IV or the DSM revised third edition (DSM-III-R) and a predefined cutoff point of the HAM-D. Most participants had moderate to severe depression as measured by the HAM-D. All trials excluded patients who had additional Axis I disorders, high suicidal risk, or progressive medical diseases or who used psychotherapy, electroconvulsive therapy, or psychotropic medications.

Second-Generation Antidepressants Compared with Acupuncture

Table 15 describes the five trials (six articles, two reporting on substantially the same participants) that compared patients treated with an SGA to those treated with acupuncture monotherapy or with acupuncture plus an SGA. All trials took place in China. Four sets of analyses were funded by Chinese government agencies;^{94,96,114,115} the other two did not report their funding sources.^{117,118} Trial enrollment ranged from 75 to 157 participants. All trials performed primary outcome evaluations at 6 weeks.

Table 15. Second-generation antidepressant versus acupuncture: Study characteristics, main outcomes, and risk of bias ratings

Trial	N	Mean Baseline HAM-D Score	SGA Dose (mg/day) Type of Acupuncture (Number of Sessions)	Response ^a (%) and Significance Level	Remission ^a (%) and Significance Level	Risk of Bias Rating
Huang et al., 2005 ¹¹⁴	98	24.1	Fluoxetine (20–40) Scalp EA (36)	65 vs.56 p=NR	NR	Medium
	6					
Qu et al., 2013 ⁹⁴	157	24.4	Paroxetine (10–40) Paroxetine + MA (18) Paroxetine + EA (18)	42 vs. 70 (MA) vs.70 (EA) p=0.004 for SGA vs. MA or EA	22.9 vs.22.6 (MA) vs 28.6 (EA) p=0.72	Medium
Chen et al., 2014 ¹¹⁵ ^b	6					
Song et al., 2007 ¹¹⁸	90	25.3	Fluoxetine (20) EA (30)	NR	NR	High ^c
	6					
Sun et al., 2013 ⁹⁶ ^e	75	23.3	Fluoxetine (20) EA #1 (30) EA #2 (30)	60 vs.75 vs. 75 p=0.16	NR	High ^d
	6					
Zhang et al., 2009 ¹¹⁷	80	24.1	Fluoxetine (20–30) + sham MA (30) Fluoxetine (10) + MA (30)	80 vs. 78 p=0.79	NR	Medium
	6					

^a Response and remission are measured on the Hamilton Depression Rating Scale (HAM-D).

^b The Chen et al. trial had substantial overlap of participants (n=105) with the Qu et al. trial.

^c Very little information provided on randomization procedures and analytic methods.

^d High differential attrition; completers analysis.

^e Trial included two active electroacupuncture groups, with different sets of points, designed to treat depression.

EA = electroacupuncture; HAM-D = Hamilton Depression Rating Scale; MA = manual acupuncture; mg/day = milligram per day; N = number; NR = not reported; SGA = second generation antidepressant; vs. = versus.

Four trials used fluoxetine; the Qu et al. and Chen et al. trials used paroxetine. Trials employed a variety of experimental designs—including a variety of types of acupuncture, points used, and frequency of treatment—making this a heterogeneous set of trials. Three trials used the HAM-D-24^{96,114,118} and two used the HAM-D-17.^{94,117} Chen et al.¹¹⁵ reported on essentially the same dataset as the Qu et al. trial;⁹⁴ also, it described outcomes for only the SCL-90 (Symptom Checklist 90), so we excluded it from meta-analyses.

Second-Generation Antidepressants Versus Acupuncture: Monotherapy Comparisons

One medium risk of bias trial compared fluoxetine (20-40mg/day) with scalp electroacupuncture (36 sessions).¹¹⁴ This trial recruited participants from four university-based hospitals. After 6 weeks, participants treated with fluoxetine or scalp electroacupuncture reported similar response rates (65 percent versus 56 percent, p-value not reported). A second trial, which we rated high risk of bias, reported fewer treatment responses with fluoxetine (20 mg/day) than electroacupuncture (30 sessions) (60 percent versus 75 percent, p=0.16).⁹⁶

Results from network meta-analyses indicated no difference in response rates between patients treated with acupuncture and those treated with SSRIs (RR, 0.80; 95% CI, 0.46 to 1.4), those treated with SNRIs (RR, 0.80; 95% CI, 0.46 to 1.40), or other antidepressants (RR, 1.05; 95% CI, 0.59 to 1.87).

Second-Generation Antidepressants Versus Acupuncture: Combination Comparisons

Two medium risk of bias RCTs compared SGA monotherapy with a combination of acupuncture and an SGA.^{94,117} Qu et al. compared paroxetine (10–40 mg/d) with manual acupuncture (18 sessions) plus paroxetine and also with electroacupuncture (18 sessions) plus paroxetine. Response to treatment was significantly lower for paroxetine than for both combination acupuncture arms (42 percent versus 70 percent or 70 percent, $X^2=11.04$, p=0.004); the trial found no differences in remission among the three treatment arms (22.9 percent versus 22.6 percent or 28.6 percent, $X^2=0.65$, p=0.72). Zhang et al. compared fluoxetine (20–30 mg/d) plus sham acupuncture (30 sessions) with fluoxetine (10 mg/d) plus acupuncture (30 sessions). Response to treatment did not differ between the trial arms (80 percent versus 78 percent, p=0.79).

Second-Generation Antidepressants Compared With Omega-3 Fatty Acids

Second-Generation Antidepressants Versus Omega-3 Fatty Acids: Monotherapy Comparisons

One high risk of bias RCT compared fluoxetine with either EPA (eicosapentaenoic acid, 1,000 mg/day) or DHA (docosahexaenoic acid) monotherapy or a combination of EPA (1,000 mg/day) and fluoxetine (20 mg/day) (n=60).¹¹² This trial took place in Iran, recruited participants from a psychiatric hospital, and received funding from its local academic institution. After 8 weeks, patients treated with fluoxetine or omega-3 fatty acid supplements reported similar response rates (50 percent versus 56 percent, p=0.43).

Results from network meta-analyses indicated substantially higher response rates for patients treated with SSRIs (RR, 2.08; 95% CI, 0.99 to 4.36), SNRIs (RR, 2.08; 95% CI, 0.99 to 4.35), or other antidepressants (RR, 2.73; 95% CI, 1.29 to 5.79) than for patients treated with omega-3 fatty acids.

Second-Generation Antidepressants Versus Omega-3 Fatty Acids: Combination Comparisons

Two trials compared patients treated with either fluoxetine or citalopram with patients treated with combinations of omega-3 fatty acids plus an SGA; we rated both these trials as high risk of bias (Table 16). One trial took place in the United States (funded by the National Institutes of Health) and recruited participants from outpatient referrals and local advertisements.⁹⁵ The other

trial was from Iran (described above).¹¹² Combined, the trials evaluated 72 participants receiving either SGA monotherapy or the combination intervention; patient ages ranged from 18 to 65 years, and about 70 percent were female; the interventions took place over an 8-week period. Omega-3 fatty acid supplements consisted of either 1,000 mg daily of pure EPA¹¹² or a combination of 1,800 mg EPA, 400 mg DHA, and 200 mg other omega-3 fatty acids daily.⁹⁵ Primary outcome evaluations were based on the HAM-D.

Table 16. Second-generation antidepressants versus omega-3 fatty acids: Study characteristics, main outcomes, and risk of bias ratings

Trial	N	Mean Baseline HAM-D Score	SGA Dose (mg/day) Fatty Acid Dose (mg/day)	Response ^a (%) and Significance Level	Remission ^a (%) and Significance Level	Risk of Bias Rating
Gertsik et al., 2012 ⁹⁵	42	25.3	Citalopram 20–40	14 vs. 17	18 vs. 44	High ^b
	8		EPA 1,800 + DHA 400 + other 200 + citalopram 20-40	NR	NR	
Jazayeri et al., 2008 ¹¹²	60	30.0	Fluoxetine 20	50 vs. 56 vs. 81 ^c	NR	High ^d
	8		EPA 1,000 Fluoxetine 20 + EPA 1,000	p=0.43, p=0.005, p=0.009		

^a Response and remission are measured on the Hamilton Depression Rating Scale (HAM-D).

^b Unclear randomization methods; high attrition; completers analysis.

^c Fluoxetine vs. EPA vs. fluoxetine + EPA. p values are for fluoxetine vs. EPA, fluoxetine vs. combination, and EPA vs. combination, respectively.

^d Unclear randomization methods; high attrition; completers analysis.

DHA = docosahexaenoic acid; EPA = eicosapentaenoic acid; HAM-D = Hamilton Depression Rating Scale; mg/day = milligram per day; N = number; NR = not reported; vs. = versus.

In the U.S. trial, at 8 weeks, changes in HAM-D favored the combination of omega-3 fatty acid supplement plus citalopram over citalopram monotherapy (data not reported, $p < 0.05$).⁹⁵ The Iran trial reported superior 8-week treatment response rates for combination treatment with EPA plus fluoxetine (81 percent) over rates for either fluoxetine (50 percent) or EPA (56 percent) alone ($p = 0.005$).¹¹² Similarly, the combination treatment produced greater reductions in HAM-D over 8 weeks than either monotherapy (data not reported, $p = 0.005$). In summary, participants treated with a combination of omega-3 fatty acids plus SGA were more likely to benefit than participants treated with either SGA or omega-3 monotherapy.

Second-Generation Antidepressants Compared With S-Adenosyl methionine

One trial (high risk of bias) compared escitalopram (10–20 mg/day) to SAMe (1,600–3,200 mg/d).⁹³ The National Institutes of Health supplied funding. The trial recruited participants from outpatient referrals and local advertisements to academic hospitals in two U.S. locations. Patients ranged in age from 17 to 79 years. The sample was 50 percent female. The trial evaluated outcomes, based on the HAM-D, after 12 weeks of treatment.

Second-Generation Antidepressants Versus S-Adenosyl Methionine: Monotherapy Comparisons

Treatment groups did not differ significantly in treatment response (34 percent versus 36 percent, $p > 0.05$), remission (28 percent versus 28 percent, $p > 0.05$), or reduction in HAM-D scores over time (6.3 versus 6.1, p-value not reported) (see Table 17). Results of our network

meta-analyses also reported similar response rates for patients treated with SSRIs (RR, 1.24; 95% CI, 0.67 to 2.30), SNRIs (RR, 1.24; 95% CI, 0.67 to 2.30), or other SGAs (RR, 1.63; 95% CI, 0.87 to 3.07) and patients treated with SAME.

Table 17. Second-generation antidepressants versus SAME: Study characteristics, main outcomes, and risk of bias ratings

Trial	N	Mean Baseline HAM-D Score	SGA Dose (mg/day) and SAME Dose (units)	Response ^a (%) and Significance Level	Remission ^a (%) and Significance Level	Risk of Bias Rating
Miscoulon et al., 2014 ⁹³	189	19.2	Escitalopram 10-20	34% vs 36%	28% vs 28%	High ^b
	12		SAME 1600-3200	p>0.05	p>0.05	

^a Response and remission are measured on the Hamilton Depression Rating Scale (HAM-D).

^b High attrition.

Second-Generation Antidepressants Versus S-Adenosyl Methionine: Combination Comparisons

We did not find any trials comparing SGA monotherapy with a combination therapy of SGAs and SAME. Data were insufficient to estimate the comparative benefits of SGA monotherapy with combination SAME plus SGA using network meta-analyses.

Second-Generation Antidepressants Compared With St. John's Wort (*Hypericum perforatum*)

Overall, 12 trials (13 articles) compared an SGA with St. John's wort (Table 18). Trials used a variety of commercially available standardized extracts (LI-160, WS5570, Ze117, STW3, Calmigen, Iperisan, Swiss herbal remedies), most often standardized to 0.12 to 0.28 percent hypericin; doses ranged from 300 mg to 1,800 mg of the standardized extract daily. Nine trials included 900 mg within their dosing range. Six trials used fluoxetine for comparison,^{103,106,108,121-123} four used sertraline,^{97,105,107,120} one used paroxetine,¹⁰⁴ and one used citalopram¹¹⁹ (see Table 16 for dosages). In all, these trials provided data on 1,855 participants, predominantly with severe depression. Three trials took place in outpatient psychiatry clinics,^{103,104,106} six trials in outpatient primary care clinics,^{105,108,119,120,122,123} and three trials did not report the source of patients beyond outpatient communities.^{97,107,121} Five trials were conducted in Germany;^{104,108,119,120,122} three in the United States;^{97,103,107} and one each in Brazil,¹⁰⁶ Canada,¹⁰⁵ Denmark,¹²¹ and Sweden.¹²³ The maker of the supplement sponsored seven trials;^{103-108,123} the U.S. government sponsored one.⁹⁷ Treatment duration ranged from 4 to 12 weeks. Most trials had a medium risk of bias, although we rated three trials as high^{103,106,107} and two trials as low risk of bias.^{119,123} In two cases, we gave a medium risk of bias rating to high-risk trials when evaluating response and remission.^{103,107}

Table 18. Second-generation antidepressants versus St. John's wort: Trial characteristics, main outcomes, and risk of bias ratings

Trial	N	Mean Baseline HAM-D Score	SGA Dose (mg/day) and St. John's Wort Formulation (mg/day)	Response ^a (%) and Significance Level	Remission ^a (%) and Significance Level	Risk of Bias Rating
Behnke et al., 2002 ¹²¹	61	20.4	Fluoxetine 40 Calmigen 300	66 vs. 55 p=0.41	NR	Medium
	6					
Bjerkstedt et al., 2005 ¹²³	108	24.7	Fluoxetine 20 LI160 900	37 vs. 38 NS	28 vs. 24 NR	Low
	4					
Brenner et al., 2000 ¹⁰⁷	30	21.5	Sertraline 50–75 LI160 600–900	40 vs. 47 NS	NR	High ^{b, c}
	7					
Davidson et al., 2002 ⁹⁷	222	22.7	Sertraline 50–100 LI160 900–1,500	24 vs. 14 NR	25 vs. 24 NR	Medium
	8					
Fava et al., 2005 ¹⁰³	92	19.6	Fluoxetine 20 LI160 900	NR	30 vs. 38 NS	High ^{b, e}
Papakostas et al., 2007 ^{129d}	12					
Gastpar et al., 2005 ¹²⁰	200	22.1	Sertraline 50 STW3 612	69 vs. 74 NS	NR	Medium
	12					
Gastpar et al., 2006 ¹¹⁹	258	21.9	Citalopram 20 STW3-VI 900	56 vs. 54 p=0.63	NR	Low
	6					
Harrer et al., 1999 ^{122f}	149	NR	Fluoxetine 10 LoHyp-57 400	72 vs. 71 NR	NR	Medium
	6					
Moreno et al., 2006 ¹⁰⁶	72	NR	Fluoxetine 20 Iperisan 900	NR p=0.021	12 vs. 35 NR	High ^g
	8					
Schrader et al., 2000 ¹⁰⁸	238	19.6	Fluoxetine 20 Ze117 500	40 vs. 60 p=0.05	NR	Medium
	6					
Szegedi et al., 2005 ¹⁰⁴	244	25.5	Paroxetine 20–40 WS5570 900– 1,800	73 vs. 86 p=0.08	43 vs. 61 p=0.02	Medium
	6					
van Gurp et al., 2002 ¹⁰⁵	90	19.3	Sertraline 50–100 Swiss herbal remedies 900–1,800	NR	NR	Medium
	12					

^a Response and remission are measured on the HAM-D.

^b For dichotomous outcomes (e.g., response and remission), we rated the risk of bias for these trials medium because dropouts were counted as remission failures.

^c High attrition, unclear randomization methods.

^d Not included in meta-analyses because it is a reanalysis of Fava et al., 2005.¹⁰³

^e High attrition, unclear randomization methods.

^f Not included in response and remission meta-analyses because of the age of trial population (60 to 80 years).

^g Completers analysis.

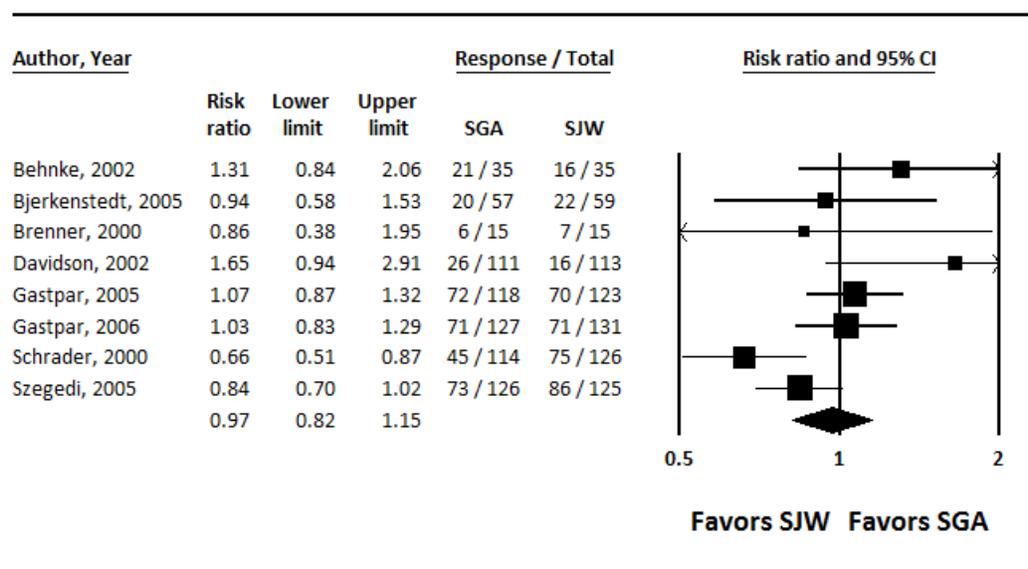
HAM-D = Hamilton Depression Rating Scale; mg/day = milligrams per day; N = number; NR = not reported; NS = reported as not significant; SGA = second-generation antidepressants; vs. = versus.

Second-Generation Antidepressants Versus St. John's Wort: Monotherapy Comparisons

Overall, treatment effects with respect to response to treatment, remission, and magnitude of change on the HAM-D scale were similar between patients treated with SGAs or St. John's wort. We did not find any evidence with respect to other outcomes of interest such as quality of life or functional capacity.

We conducted random-effects meta-analyses of eight low or medium risk of bias trials that reported data on response (1,179 participants), typically defined as ≥ 50 percent decrease in HAM-D.^{97,104,107,108,119-121,123} Patients treated with SGAs and those receiving St. John's wort had similar response rates (45.2 percent versus 46.0 percent; RR, 0.97; 95% CI, 0.82 to 1.15) after 4 to 12 weeks of treatment (Figure 10).

Figure 10. SGAs versus St. John's Wort: Response



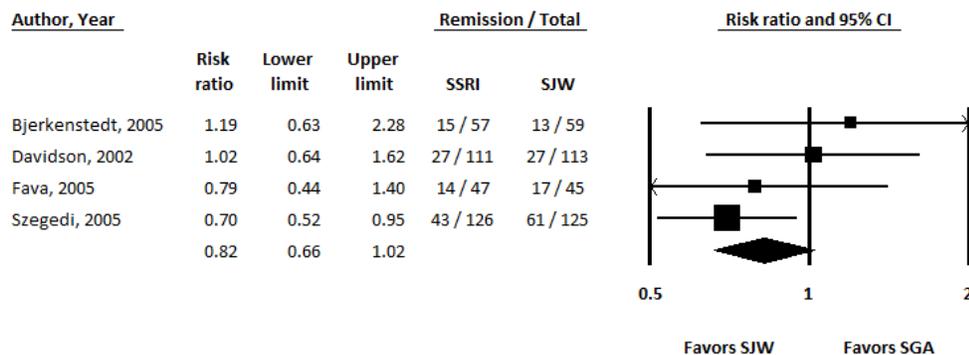
Random effects meta-analysis; I-squared 56%

CI = confidence interval; SGA = second generation antidepressants; SJW = St. John's wort.

Results from network meta-analyses indicated significantly higher response rates for patients treated with St. John's wort than for patients treated with SSRIs (RR, 0.82; 95% CI, 0.69 to 0.98) or SNRIs (RR, 0.82; 95% CI, 0.68 to 0.98). This was not true for patients given other antidepressants (RR, 1.08; 95% CI, 0.86 to 1.35).

Likewise, random-effects meta-analyses of four low or medium risk of bias trials (683 participants) showed similar remission rates (typically defined as HAM-D ≤ 7) for participants on SGAs or St. John's wort (29.0 percent versus 34.5 percent; RR, 0.83; 95% CI, 0.66 to 1.04, four trials) after 4 to 12 weeks of treatment (Figure 11).^{97,103,104,123} Sensitivity analysis including one high risk of bias trial¹⁰⁶ (40 participants) produced similar findings (29.4 percent versus 33.2 percent; RR, 0.92; 95% CI, 0.67 to 1.28; five trials).

Figure 11. SGA versus St. John's Wort: Remission

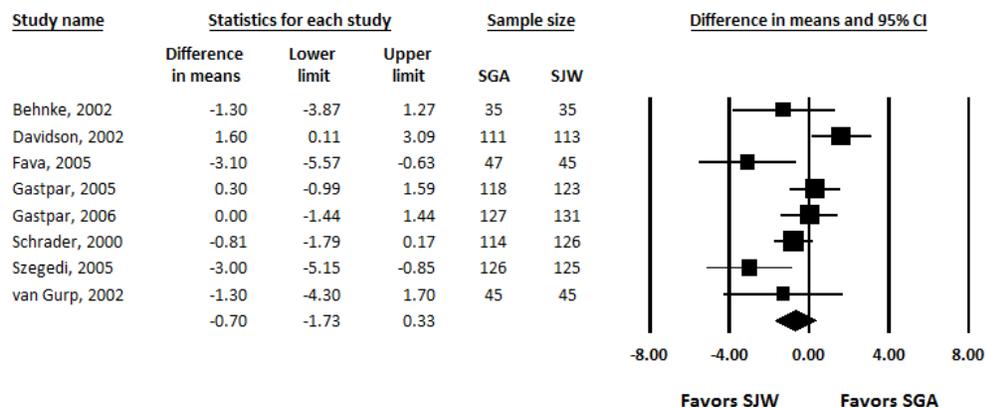


Random effects meta-analysis; I-squared 7%

CI = confidence interval; SGA = second generation antidepressants; SJW = St. John's wort;

Eight trials reported data on change in HAM-D scores (1,405 participants).^{97,103-105,108,119-121} We found similar HAM-D reductions for patients treated with an SGA and those treated with St. John's wort (mean difference -0.70; 95% CI, -1.73 to 0.33, eight trials; Figure 12). Sensitivity analysis including one high risk of bias trial indicated no difference in conclusions (mean difference -0.63; 95% CI, -1.63 to 0.37, nine trials).

Figure 12. SGA versus St. John's Wort: Change in HAM-D-17



Random effects meta-analysis; I-squared 65%

CI = confidence interval; HAM-D = Hamilton Depression Scale for Depression; SGA = second generation antidepressants; SJW = St. John's wort.

Second-Generation Antidepressants Versus St. John's Wort: Combination Comparisons

We did not find any trials comparing SGA monotherapy with a combination therapy of St. John's wort and SGAs. Data were insufficient to estimate the comparative benefits with network meta-analyses.

Second-Generation Antidepressants Compared With Yoga

We found no eligible trials that compared an SGA with yoga.

Second-Generation Antidepressants Compared With Meditation

We found no eligible trials that compared an SGA with meditation therapy.

KQ 1a. Second-Generation Antidepressants Compared With Exercise Interventions

We identified two primary RCTs (four articles) comparing an SGA with an aerobic exercise intervention for treating patients with MDD.^{98,99,130,131} The same group of researchers conducted both trials; we rated both as medium risk of bias (Table 19). Both trials evaluated sertraline compared with aerobic exercise; the earlier trial also evaluated the efficacy of sertraline alone compared with sertraline plus aerobic exercise.⁹⁸ The trials enrolled patients according to a criteria-based diagnosis of MDD based on DSM-IV;⁶³ they excluded patients who had additional Axis I disorders, high suicidal risk, or progressive medical diseases; who were sedentary; or who were not undergoing psychiatric treatment. Both trials used the HAM-D 17 to assess MDD severity at baseline and at 16 weeks. Participants had depression of moderate severity at baseline as measured by the HAM-D. Grants from the National Institutes of Health and Pfizer Pharmaceuticals funded both studies.

Table 19. Second-generation antidepressants versus exercise: Study characteristics, main outcomes, and risk of bias ratings

Trial	N	Mean Baseline HAM-D Score	SGA Dose (mg/day) Type of Exercise (Frequency of Sessions)	Response ^a (Percent) and Significance Level	Remission ^a (Percent) and Significance Level	Quality of Life, Functional Capacity and Significance Level	Risk of Bias Rating
Blumenthal et al., 1999 ⁹⁸	156	NR	Sertraline 50–200 mg/day Aerobic exercise (3 times per/week)	At 16 weeks: 68.6 vs. NR	60.4 vs. 65.5	Life satisfaction p=NS	Medium
Babiyak et al., 2000 ¹³²	16-week treatment; 6-month followup	Per group: range from 17 to 19	Sertraline + aerobic exercise (3 times per week)		p=0.67		
Blumenthal et al., 2007 ⁹⁹	202	NR	Sertraline 50–200 mg/day Supervised aerobic exercise (3 times per week)	At 16 weeks: 47 vs. NR	45 vs. 40	Neurocognitive tests battery p=NS	Medium
Hoffman et al., 2008 ¹³¹	16-week treatment	Per group: range from 16 to 17	Home-based aerobic exercise (3 times per week)		p=0.66		

^a Response (≥ 50 percent decrease in depressive severity) and remission (as defined by authors of individual trials) are measured using the HAM-D unless indicated otherwise.

HAM-D = Hamilton Depression Rating Scale; mg/day = milligrams per day; N = number; NR = not reported; NS = reported as not significant, SGA = second-generation antidepressants; vs. = versus.

The trials included 358 participants total, recruited from the community into an outpatient facility at an academic medical center. In the earlier trial, participants ranged from 50 to 77 years of age (mean, 57 years); 57 percent of the sample was female. In the 2007 trial, participants' mean age was 52 years, and 51 percent were female.

Both trials compared a 50–200 mg daily sertraline dose with a supervised aerobic exercise program of 45 minutes three times weekly over 16 weeks. The aerobic exercise program consisted of a 10-minute warm-up exercise period followed by 30 minutes of continuous walking

or jogging at an intensity that would maintain heart rate at 70 percent to 85 percent of heart rate reserve, followed by 5 minutes of cool-down exercises. In addition, the Blumenthal et al., 1999⁹⁸ trial compared the sertraline and supervised exercise arms, individually, with an arm combining sertraline with supervised exercise. In contrast, the Blumenthal et al., 2007⁹⁹ trial used a four-armed design—adding a home-based exercise program arm and a placebo pill arm. The primary outcome for both trials was the remission rate at 16 weeks (no longer meeting MDD criteria and HAM-D <8). At baseline and 16 weeks of treatment, participants also underwent a graded exercise treadmill test to measure exercise capacity and tolerance. The trial reported additional secondary outcomes: anxiety, self-esteem, life satisfaction, and dysfunctional attitudes in the 1999 trial and neurocognitive improvement in the 2007 trial.

Second-Generation Antidepressants Versus Aerobic Exercise: Monotherapy Comparisons

Neither trial found a statistically significant difference in remission rates between sertraline alone and aerobic exercise alone: 68.8 percent (sertraline) versus 60.4 percent (exercise) in the 1999 trial and 47 percent (sertraline) versus 45 percent (supervised exercise) versus 40 percent (home-based exercise) in the 2007 trial. All three active groups in the 2007 trial tended to have higher remission rates than the placebo control group (31 percent) ($p=0.057$). The crude pooled risk ratio comparing sertraline treatment with the exercise conditions (pooling data from the two exercise groups in the 2007 trial with the one exercise-only group in the 1999 trial, a total of three arms) was 1.10 (95% CI, 0.87 to 1.39). All active treatment groups in the 2007 trial showed a clinically and statistically significant decline ($p<0.0001$) in HAM-D scores from baseline to 16 weeks; the sertraline group decreased by 6.1, supervised exercise by 7.2, home-based exercise by 7.1, and placebo by 6.1 points. There were no between-group differences in this decline ($p=0.321$). The 1999 trial found the magnitude of the decline in HAM-D to be comparable across groups; it did not provide specifics. Neither trial reported response rates. Based on network meta-analyses, patients in the SGA and exercise groups had similar response rates (RR 1.87, 95% CI 0.81- 4.33).

In both trials, patients receiving sertraline showed significantly lower levels of aerobic capacity (peak $\dot{V}O_2$), as well as shorter treadmill times, than patients in the exercise groups ($p<0.001$). In the 2007 trial, patients in the sertraline group showed a mean 0.8 percent decrease in peak $\dot{V}O_2$ and a 3.9 percent improvement in treadmill time; those in the placebo group declined by 4 percent in peak $\dot{V}O_2$ and 2.3 percent in treadmill time. In comparison, participants in the supervised exercise group improved by 8.3 percent in peak $\dot{V}O_2$ and 18.8 percent in treadmill time; home-based exercise participants improved by 3.5 percent in peak $\dot{V}O_2$ and 7.5 percent in treadmill time. Similarly, in the Blumenthal et al., 1999 trial, patients in the sertraline group showed minimal (<3 percent) improvement in aerobic capacity, and those in the exercise group improved by 11 percent. This trial also assessed anxiety, self-esteem, life satisfaction, and dysfunctional attitudes. Although both the sertraline and the exercise groups improved, the groups did not differ on these measures. The companion report to the 2007 trial¹³¹ found little evidence of between-group differences in neurocognitive measures; exercise participants performed better than those on sertraline on tests of executive function (Trail-making Test, $p=0.02$; Ruff 2 & 7 test, $p=0.03$) but not on measures of verbal memory or verbal fluency/working memory.

In a sensitivity analysis, the magnitudes of the RRs are slightly attenuated with inclusion of trials with a high risk of bias, but the interpretations do not change.

Second-Generation Antidepressants Versus Aerobic Exercise: Combination Comparisons

A single trial, Blumenthal et al., 1999,⁹⁸ included an arm comparing sertraline alone to a combination of sertraline plus exercise; it had 48 participants in the sertraline-alone group and 55 in the combined sertraline plus exercise group. Data were insufficient to estimate comparative benefits of SGA monotherapy versus combination therapy with SGA and exercise using network analysis. Patients in the sertraline-only group showed minimal (<3 percent) improvement in aerobic capacity; those in the combined group improved by 9 percent. The two groups did not differ in improvements in anxiety, self-esteem, life satisfaction, or dysfunctional attitudes scores.

KQ 1b. Effect of Severity: Second-Generation Antidepressants Compared With Psychological Interventions

Description of Trials

In all, four RCTs compared SGA with a psychological treatment and provided data for KQ 1b (Table 20).^{79,86,92,110} We rated two trials medium risk of bias^{79,92} and two trials as high risk of bias.^{86,110} One high risk of bias trial compared SGA with either of two psychological treatments.⁸⁶ Included trials compared an SGA with a CBT (cognitive therapy),^{86,110} a third-wave CBT (behavioral activation),⁸⁶ a PSYD,⁷⁹ and an integrative therapy.⁹² We found no trials eligible for KQ 1b that compared a SGA with behavior therapy/behavior modification or with a humanistic therapy.

Table 20. SGAs versus psychological interventions by depression severity: Trial characteristics, main outcomes, and risk of bias ratings of trials

Author, Year	N	MDD Severity Definition and Mean Baseline Severity	SGA Dose (mg/day) Psychotherapy (Number of Sessions)	Response ^a (Percent) and Significance Level	Remission ^a (Percent) and Significance Level	Change (SD) in HAM-D and Significance Level	Risk-of Bias Rating
Dimidjian 2006 ⁸⁶	145 16	Low: HAM-D-17 ≤19	Paroxetine 50 CT 24 ^b	Low: Paroxetine 47 CT 60 p=0.60	Low: Paroxetine 33 CT 50 p=0.60	Low: Paroxetine: -8.53 (NR) CT: -9.46 (NR) p=NR	High ^c
Cognitive therapy		Mean: SGA 16.98 CT 16.65 High: HAM-D ≥20 Mean: SGA 23.79 CT 23.16		High: Paroxetine 40 CT 56 p=0.16	High: Paroxetine 23 CT 36 p=0.002	High: Paroxetine: -15.16 (NR) CT: -12.39 (NR) p=NR	

Table 20. SGAs versus psychological interventions by depression severity: trial characteristics, main outcomes, and risk of bias ratings of trials (continued)

Author, Year	N Duration (Weeks)	MDD Severity Definition and Mean Baseline Severity	SGA Dose (mg/day) Psychotherapy (Number of Sessions)	Response ^a (Percent) and Significance Level	Remission ^a (Percent) and Significance Level	Change (SD) in HAM-D and Significance Level	Risk of Bias Rating
Menchetti et al., 2014 ⁹²	287 8	HAM-D-21 <18 vs. HAM-D-21 ≥18 Mean: SGA 17.5 IPT 17.1	Citalopram 10–60 or sertraline 25–200 IPT 6 to 8	NR	HAM-D <18: SGA 56 IPT 75 p=NR but is statistically significant (SRD=0.19; 95% CI, 0.04 to 0.34) HAM-D ≥18: SGA 46 IPT 40 p=NR but is not statistically significant (SRD = -0.06; 95% CI, -0.24 to 0.12)	NR	Medium
Barber et al., 2012 ⁷⁹	106 49	Baseline HAM-D-17 score <19 vs. HAM-D-17 >20 Mean: SGA 19.0 SET 19.9	Sertraline 50-100 SET 20	NR	NR	Limiting the analysis to patients with high depression severity revealed no differences in rate of change of HAM-D	Medium
Moradveisi et al., 2013 ¹¹⁰	100 49	Baseline HAM-D-17 score included in regression model. Mean: SGA 21.62 BA 21.12	Sertraline 100 BA 16	NR	NR	β (95% CI): -2.03 (-3.01 to -1.05) p<0.001	High ^d

Table 20. SGAs versus psychological interventions by depression severity: trial characteristics, main outcomes, and risk of bias ratings of trials (continued)

Author, Year	N Duration (Weeks)	MDD Severity Definition and Mean Baseline Severity	SGA Dose (mg/day) Psychotherapy (Number of Sessions)	Response ^a (Percent) and Significance Level	Remission ^a (Percent) and Significance Level	Change (SD) in HAM-D and Significance Level	Risk of Bias Rating
Dimidjian et al., 2006 ⁸⁶	143 16	Low: HAM-D-17 ≤19	Paroxetine 50 BA 24 ^c	Low: Paroxetine 47 BA 39 p=0.60	Low: Paroxetine 33 BA 39 p=0.45	Low: Paroxetine: -8.53 (NR) BA: -9.36 (NR) p=NR	High ^c
Third-wave CBT (Behavioral activation)		Mean: SGA 16.98 BA 17.28		High: Paroxetine 40 BA 60 p=NR	High: Paroxetine 23 BA 56 p=0.002	High: Paroxetine: -15.16 (NR) BA: -15.60 (NR) p=NR	
		High: HAM-D-17 ≥20					
		Mean: SGA 23.79 BA 23.16					

^a Response and remission (as defined by authors of individual trials) are measured on the Hamilton Depression Rating Scale (HAM-D) or indicated otherwise.

^b This trial contained a fourth placebo control arm.

^c High attrition.

^d High attrition.

BA = behavioral activation; CBT = cognitive behavioral therapy; CI = confidence interval; CT = cognitive therapy; HAM-D = Hamilton Rating Scale for Depression; IPT = interpersonal psychotherapy; MDD = major depressive disorder; mg/day = milligrams per day; N = number; NR = not reported; SD = standard deviation; SET = supportive expressive therapy; SGA = second generation antidepressant; SRD = standardised rate difference; vs. = versus.

Two of the trials were conducted in the United States,^{79,86} and two were conducted in other countries: one in Iran¹¹⁰ and one in Italy.⁹² Two of the trials took place in outpatient primary care settings;^{86,92} two were conducted in outpatient psychiatry clinics.^{79,110} Three of the trials were funded entirely or in part by the government.^{79,86,92} Three trials did not provide any information on treatment fidelity,^{79,92,110} and only one trial reported adequate treatment fidelity.⁸⁶ None of the trials reported on functional capacity, quality of life, reduction of suicidality, relapse, or hospitalization. None of the trials excluded individuals with any comorbid anxiety disorder, although one trial reported that they did not include subjects with a primary diagnosis of panic disorder or obsessive-compulsive disorder.⁸⁶

None of the trials was designed to answer the primary question of whether depressive severity was a modifier of the comparative effectiveness of SGAs versus psychotherapy. However, two trials prespecified their plan to use depressive severity as a moderator.^{86,92} The methods to analyze whether outcome measured by depressive severity varied. One trial stratified its sample into a high- and low-severity subgroup and assessed the comparative benefits of the SGAs versus psychotherapy within each subgroup.⁸⁶ Another trial examined potential moderators of remission with logistic modeling, including stratification of high versus low severity as one possible predictor.⁹² A third trial used a mixed regression analysis model that tested whether the baseline depressive severity score moderated outcomes.¹¹⁰ Finally, one trial used hierarchical linear modeling to determine whether depressive severity had a moderating effect, considering both the full sample as well as the subgroup with higher depressive severity.⁷⁹

Generally, patient age ranged between 18 and 50 years old; trials reported a mean age between 31.4¹¹⁰ and 44.9 years.⁹² In all trials, the majority of the patients were female. Two trials reported minority status (18.3 percent⁸⁶ and 48.1 percent⁷⁹).

Impact of Severity on Various Outcomes

One high risk of bias trial (n=145), with one arm comparing paroxetine and CT, conducted subgroup analyses in patients with low- and high-severity MDD.⁸⁶ For the subgroup with high-severity MDD (i.e., those with a HAM-D-17 ≥ 20), those receiving paroxetine were less likely to achieve remission of MDD than those receiving CT (23 percent versus 36 percent, $p=0.002$).⁸⁶ For the subgroup with low-severity MDD (i.e., those with a HAM-D-17 ≤ 19); remission rates did not differ significantly for patients treated with paroxetine or CT. Efficacy did not differ significantly between treatments in either subgroup when measured by treatment response or change in HAM-D-17. Because of the small sample size and the fact that authors conducted multiple parallel comparisons of subgroups and not a test of interaction, findings might be attributable to chance and need to be viewed cautiously.

One medium risk of bias trial (n=287) reported subgroup analyses of patients with low- or high-severity MDD at baseline who were treated with either an SGA or IPT.⁹² From regression analyses, Menchetti and colleagues⁹² reported that the likelihood of remission varied as a function of depression severity; only those with less severe depression saw a worse outcome from SGA than from IPT. For patients with baseline HAM-D-21 < 18 , those receiving 2 months of citalopram or sertraline were 19 percent less likely to achieve remission than those receiving IPT (Standardised Rate Difference [SRD], 0.19; 95% CI, 0.04 to 0.34), consistent with a small-to-moderate effect size (ES = 0.25).⁹² However, for patients with high-severity MDD (HAM-D-20 ≥ 18), the likelihood of remission did not differ between the two treatment groups [SRD, -0.06; 95% CI, -0.24 to 0.12].⁹² The trial did not report treatment response or change in HAM-D score.

Two high risk of bias trials^{86,110} compared an SGA with behavioral activation and provided subgroup analyses that considered the effects of depressive severity on treatment outcome. In one trial (n=100), Moradveisi and colleagues used regression modeling to assess the effect of baseline severity on change in depressive severity.¹¹⁰ The difference in treatment effects between the two types of interventions increased as a function of severity. In patients with less severe MDD at baseline, the difference in treatment effects at weeks 4, 13, and 49 were minimal. However, as baseline severity increased, patients receiving sertraline had less improvement in depressive severity as measured by both HAM-D and BDI at each follow-up point.¹¹⁰

The other high risk of bias trial (n=143) reported on the effect of baseline depressive severity on all three main outcomes.⁸⁶ In this trial, the authors reported that for subjects with high-severity MDD (defined as HAM-D-17 ≥ 20), those receiving paroxetine were less likely to remit than those receiving BA. In those with low-severity MDD, remission rates did not differ to a statistically significant degree between the two treatment groups. For the other two outcomes, treatment response or change in HAM-D-17 score, having either high- or low-severity MDD did not produce different outcomes for the two interventions.

One medium risk of bias trial (n=106) that compared supportive–expressive psychotherapy conducted subgroup analyses in high- and low-severity patients.⁷⁹ The trial did not report on either response to treatment or remission. Although the authors did not report specific changes in HAM-D scores stratified by subgroup, they did analyze depression severity as a potential moderator of change in HAM-D scores. Limiting the analysis to patients with high depression

severity revealed no differences in rate of change of HAM-D. We contacted trial authors for additional data but did not receive any supplementary information.

Comparative Efficacy for Critical Efficacy Outcomes by Baseline Severity for Psychological Interventions and Second-Generation Antidepressants

We further investigated the role of depressive severity on outcomes by considering all trials from KQ1a that both directly compared psychological interventions to SGAs and reported on key effectiveness outcomes (response, remission, and/or functional capacity). These studies did not directly assess depressive severity as a moderator; however, one might observe whether there is evidence of a relation between mean baseline depressive severity and the comparative effectiveness of the interventions (Table 21). We were not able to stratify by whether the depressive severity of the populations was specifically “moderate” or “severe”, because most populations were mixed (i.e., they had both moderate and severely depressed populations mixed together). Rather, for each comparison we list the range of mean baseline depressive severity and the findings. Of note, we found no differences in the comparative effectiveness between SGAs and psychological treatments in patients with moderate to severe MDD, which is consistent with findings of the few studies that we have for KQ1b. However, as with our earlier KQ1b findings, the evidence was very limited.

Table 21. Comparative efficacy for critical efficacy outcomes by baseline severity for psychological interventions and second-generation antidepressants

Comparisons	Baseline MDD severity	Comparative effectiveness for critical efficacy outcomes	Strength of Evidence
SGA vs. CBT	Moderate to severe	<u>Response</u> : No statistically significant differences <u>Remission</u> : No statistically significant differences <u>Functional capacity</u> : No statistically significant differences	Moderate Low Low
SGA vs. CBT + SGA	Moderate to severe	<u>Response</u> : No statistically significant differences <u>Remission</u> : No statistically significant differences <u>Functional capacity</u> : Favours CBT + SGA combination	Low Low Low
SGA vs. IT	Moderate to severe	<u>Response</u> : No statistically significant differences <u>Remission</u> : No statistically significant differences	Low Low
SGA vs. IT + SGA	Moderate to severe	<u>Remission</u> : Favours SGA	Low
SGA vs. PSYD	Moderate	<u>Remission</u> : No statistically significant differences <u>Functional capacity</u> : No statistically significant differences	Low Low

KQ 1b. Effect of Severity: Second-Generation Antidepressants Compared With Complementary and Alternative Medicine Interventions

One trial compared SGA with a CAM therapy for treating patients with MDD.⁹³ Participants were enrolled according to a criteria-based diagnosis of MDD based on either the DSM-IV or the DMS-III-R and a predefined cutoff point for the HAM-D. Most participants had moderate to severe depression as measured by the HAM-D. Patients were excluded who had additional Axis I disorders, high suicidal risk, or progressive medical diseases or who used psychotherapy, electroconvulsive therapy, or psychotropic medications.

Second-Generation Antidepressants Compared With S-Adenosylmethionine

One trial (Table 22), rated medium risk of bias, compared escitalopram (10–20mg/d) with SAME (1,600–3,200 mg/d). The National Institutes of Health supplied funding. The trial recruited participants (N = 129) from outpatient referrals and local advertisements to academic hospitals in two U.S. locations. Participant age ranged from 17 to 79 years; the sample was 50 percent female. The trial evaluated outcomes, based on the HAM-D, after 12 weeks of treatment. Mean (SD) baseline HAM-D score was 19.2 (4.7) with a range from 4 to 32. No statistically significant interaction appeared between baseline HAM-D score and treatment groups for reduction in HAM-D scores over time (p=0.87).

Table 22. SGAs versus SAME by depression severity: Trial characteristics, main outcomes, and risk of bias ratings of trials

Trial	N Duratio n (Weeks)	Mean Baseline HAM-D score	SGA Dose (mg/day) and SAME Dose (mg/day)	Response ^a (Percent) and Significance Level	Remission ^a (Percent) and Significance Level	Reduction in HAM-D by baseline score ^b	Risk-of Bias Rating
Miscoulon et al., 2014 ⁹³	189 12	19.2	Escitalopram 10-20 SAME 1600- 3200	NR	NR	P = 0.87	High ^c

^a Response and remission are measured on the Hamilton Depression Rating Scale (HAM-D).

^b Interaction between baseline HAM-D score and overall reduction in HAM-D over 12 weeks.

^c High attrition.

Comparative Efficacy for Critical Efficacy Outcomes by Baseline Severity for Complementary and Alternative Interventions, Exercise, and Second-Generation Antidepressants

As with our psychological intervention comparison, we further investigated the role of depressive severity on outcomes by considering all trials from KQ1a that both directly compared CAM interventions and exercise to SGAs and reported on key effectiveness outcomes (response, remission, and/or functional capacity) (Table 23). Again, we found no differences in patients with moderate to severe MDD, which is consistent with findings of the few studies that we have for KQ1b. This evidence, too, was extremely limited.

Table 23. Comparative efficacy for critical efficacy outcomes by baseline severity for complementary and alternative interventions, exercise, and second-generation antidepressants

Comparisons	Baseline MDD severity	Comparative effectiveness for critical efficacy outcomes	Strength of Evidence
SGA vs. Acupuncture	Severe	<u>Response</u> : No statistically significant differences	Low
SGA vs. Acupuncture + SGA	Severe	<u>Response</u> : Favors Acupuncture + SGA combination <u>Remission</u> : No statistically significant differences	Low
SGA vs. Omega-3 Fatty Acids	Severe	<u>Response</u> : Favors SGA	Low
SGA vs. SAME	Moderate	<u>Response</u> : No statistically significant differences	Low
SGA vs. St. John's wort	Moderate to severe	<u>Response</u> : No statistically significant differences <u>Remission</u> : No statistically significant differences	Moderate
SGA vs. Exercise	Moderate	<u>Response</u> : No statistically significant differences <u>Remission</u> : No statistically significant differences	Low
SGA vs. Exercise + SGA	Moderate	<u>Remission</u> : No statistically significant differences	Low

KQ 2: Second-Line Therapy—Switching or Augmenting Strategies Involving a Second-Generation Antidepressant

KQ 2a addresses adult patients with acute-phase MDD who fail to recover after an initial treatment with an SGA (also referred to as second-line therapy). It examines the effectiveness of any eligible intervention (whether as a monotherapy or an augmentation therapy) that has been compared with one involving an SGA. The comparison can involve either switching to different treatment (pharmacologic or nonpharmacologic) or augmenting the initial SGA with a second treatment (pharmacologic or nonpharmacologic).

As with KQ 1, the nonpharmacologic interventions for this KQ include psychological interventions, CAM interventions, and exercise. For augmentation, however, the pharmacologic options increase; augmentation of the initial SGA can involve adding either a second SGA or an eligible non-SGA medication (e.g., buspirone). KQ 2b examines whether treatment effectiveness varies by MDD severity.

In all, two trials provided data that compared eligible second-line treatment strategies. Both used the HAM-D to measure outcome; neither reported quality of life or functional status outcomes. One trial compared switching to one SGA versus switching to a different SGA.¹⁰⁹ The other trial, the STAR*D study, provided data for multiple comparisons that were reported in three articles. These analyses allowed the comparison of four eligible second-line treatment strategies: switching to one SGA versus switching to a different SGA,¹³³ switching to CBT versus switching to any one of three SGAs,¹⁰⁰ augmenting with a second medication versus augmenting with CBT,¹⁰⁰ and augmenting with one non-SGA medication versus augmenting with an SGA.¹³⁴

We found no eligible switch trials directly comparing SGAs with either CAM or exercise, nor did we find any eligible augmentation trials comparing SGAs with CAM or exercise. Moreover, we found no direct comparison of switching strategies versus augmentation strategies.

Because of an insufficient number of eligible studies, we could not perform a network meta-analysis on response to treatment for second-line therapies that compared eligible second-line therapies with placebo.

Key Points: Switching Strategies

- When switching to a different SGA as a second-line therapy, various SGAs produce similar response rates (two RCTs, moderate SOE), similar remission rates (one RCT, low SOE), and a similar decrease in depressive severity (one RCT, low SOE).
- Switching to cognitive therapy does not produce statistically different rates of response (one RCT, low SOE) or remission (one RCT, low SOE) compared with switching to a different SGA.
- We did not find any eligible switch evidence comparing an SGA strategy with either CAM or exercise.

Key Points: Augmentation Strategies

- When augmenting with a second medication as a second-line therapy, adding a non-SGA augmenting medication does not lead to statistically different rates of response (one RCT, low SOE) or remission rate (one RCT, low SOE) compared with augmenting with a second SGA; augmentation with bupropion leads to a greater decrease in depressive severity than with buspirone (one RCT, low SOE).
- Augmenting with cognitive therapy does not produce statistically different rates of response (one RCT, low SOE), remission (one RCT, low SOE), or decrease in depressive severity compared with augmenting with an SGA.
- We did not find any eligible augmentation evidence comparing adding a second medication with adding either CAM or exercise.

Key Points: Severity as a Moderator of Treatment Effectiveness of Second-line Therapies

- For second-line therapies, the evidence is insufficient to draw conclusions about the effect of severity of disease on the comparative effectiveness of switching to different SGAs as measured by remission rates (secondary analyses of two RCTs, insufficient SOE).
- For second-line therapies, we did not find any eligible evidence about the effect of severity of disease on the comparative effectiveness of switching to a different SGA versus switching to any nonpharmacologic treatment.
- For second-line therapies, we did not find any eligible evidence about the effect of severity of disease on the comparative effectiveness of any augmentation strategies.

Detailed Synthesis: KQ 2

This section presents findings for both KQs 2a and 2b. KQ 2a concerns comparisons of “next step” treatment options. These can include comparisons of switch strategies against each other, augmentation strategies against each other, or switch versus augmentation strategies, as long as at least one arm involved an SGA. Eligible switch or augmentation strategies can involve eligible

psychotherapies, CAM, or exercise interventions. KQ 2b examines the question of whether the comparative effectiveness of these strategies differs by the severity of MDD.

Table 24 provides the number of included trials by eligible comparison. The evidence base for KQ 2a provided limited data (two trials reported in three articles) that addressed four comparisons—two switch and two augmentation—and involved only medications and psychotherapy. In the analyses comparing medications, specific medications were assessed head-to-head (e.g., sertraline versus bupropion); in the studies comparing medications to psychotherapy, however, the analyses grouped all medications into a single medication variable. No eligible studies involved CAM treatments or exercise. Further, the number of relevant placebo-controlled studies was insufficient to allow a network meta-analysis. In Appendix E, we present “summary of findings” tables of important outcomes. These tables are intended mainly for readers involved in developing clinical practice guidelines; they give basic information on the available evidence, show absolute and relative effect measures, and present SOE grades for each outcome on which we had evidence.

Table 24. Number of included trials by type of comparison

Comparison Category	Comparisons for KQ 2	Number of Trials and Citations
Switch	SGA switch ^a vs. SGA switch ^a	2 ^{109,133}
	SGA switch ^a vs. nonpharmacologic switch	1 ¹⁰⁰
Augmentation	SGA augmentation ^b vs. SGA augmentation ^b	1 ¹³⁴
	SGA augmentation ^b vs. nonpharmacologic switch	1 ¹⁰⁰

^aSwitching to another SGA.

^bAugmenting with a second SGA, an additional non-SGA medication, or a nonpharmacologic treatment.

KQ = Key Question; SGA = second-generation antidepressant; vs. = versus.

KQ 2a: Switching or Augmenting Strategies

Description of Included Trials

In all, two trials provided four comparison studies reported in four articles. All four comparisons reported in three of the articles^{100,133,134} involved data from the STAR*D study, which had multiple arms allowing several comparisons following a treatment failure. A different independent study reported data comparing various SGA switches.¹⁰⁹

The Lenox-Smith and Jiang trial was conducted in a single outpatient psychiatry setting in Great Britain and was funded by the pharmaceutical industry. The STAR*D comparison involved outpatients from 41 psychiatric (60 percent) and primary care (40 percent) settings in the United States and was government funded.

Generally, patients were between 18 and 65 years of age (mean ages between 41 and 43 years). In both, the majority of patients were female. Mean baseline depressive severity was at least moderate. STAR*D comparisons involved mean baseline HAM-D scores between 15.8 and 17.8; the Lenox-Smith and Jiang trial had greater depression severity, with a mean HAM-D score of approximately 26 (severe). The total daily dose of each SGA medication reached or exceeded the minimum recommended dose for that medication as prescribed for adults.

Whereas the Lenox-Smith and Jiang trial was a relatively standard RCT, the STAR*D study employed an equipoise randomization scheme that allowed some degree of patient preference. STAR*D was designed to allow multiple randomized comparisons of second-line therapies; the three relevant comparisons reported here^{100,133,134} all involved patients who did not remit following 3 months of treatment with citalopram. Patients could not refuse a specific medication

choice, but patients did have the option of refusing any of the available treatment strategies (switch to another SGA, switch to cognitive therapy, augment with a second medication, or augment with cognitive therapy), as long as at least two treatment options remained to allow randomization.

Second-Generation Antidepressant Switch Compared With Second-Generation Antidepressant Switch

Table 25 describes the trial characteristics, main outcomes, and risk of bias ratings for these analyses. The Lenox-Smith and Jiang trial lasted 12 weeks, with 396 patients randomized to one of two treatment arms.¹⁰⁹ The trial compared venlafaxine ER (doses ranged from 75 to 300 mg daily; mean daily dose was 191 mg) to citalopram (20 to 60 mg; mean daily dose, 51 mg). The investigators measured response with the HAM-D; they did not report response rate for the two study arms but instead stated that response did not differ (reported as $p=0.953$). They did not report remission rate or time to remission for any outcome. The decrease in depressive severity, whether measured by HAM-D, MADRS ($p=0.5002$), or CGI-S ($p=0.3014$), did not differ by groups. We rated the risk of bias as low.

The Rush et al. study lasted an average of 14 weeks; it randomized 727 patients into one of three treatment arms.¹³³ The switch comparison randomized patients to either bupropion SR (150 to 400 mg; mean daily dose at end of study was 282 mg), sertraline (50 to 200 mg; mean daily dose at end of study was 136 mg), or venlafaxine XR (37.5 mg to 375 mg; mean daily dose at end of study was 194 mg). Response rates did not differ by treatment arm; as reported for the QIDS-SR, they ranged from 26.1 percent to 28.2 percent (p -value not reported). Similarly, remission rates did not differ between treatment arms, whether reported for either the HAM-D ($p=0.16$) or the QIDS-SR (p -value not reported). The mean change in HAM-D score was not reported; however, the percentage decrease in QIDS-SR was reported and did not differ among the three groups. Neither the time to response (ranging from 5.5 to 7.0 weeks) nor the time to remission (5.4 to 6.2 weeks) differed among the three options.

We rated this study as medium risk of bias, as we did for all the STAR*D studies described below, for two reasons: less than 80 percent of the sample provided outcomes at study completion and because the mean medication doses ultimately prescribed indicated that some medications did not reach the maximal dose recommended in the protocol so that comparable adequate doses may not have been achieved among the various arms. Appendix C documents the full risk of bias assessments for included trials.

Table 25. Second-generation antidepressant switch versus another second-generation switch strategy: Trial characteristics, main outcomes, and risk of bias ratings

Trial	N	Total Sample Mean Baseline Severity (SD)	SGA Type: mg/day	Response ^a and Significance Level	Remission ^a and Significance Level	Mean Change in HAM-D Score from Baseline and Significance Level	Risk of Bias Rating
Lenox-Smith and Jiang, 2008 ¹⁰⁹	396 12	HAM-D: Venlafaxine ER: 28.6 (5.7) Citalopram: 28.8 (5.4) MADRS: Venlafaxine ER: 30.8 (5.7) Citalopram: 30.9 (6.1)	Venlafaxine ER: 75 to 300 Citalopram: 20 to 60	Response rate NR; text stated no difference in HAM-D response, p=0.953	NR	HAM-D -17.0 vs. -16.5, p=0.4778	Low
Rush et al., 2006 ¹³³	727 14	HAM-D: 18.9 (7.3)	Bupropion SR: 150 to 400 Sertraline: 50 to 200 Venlafaxine XR: 37.5 to 375	QIDS-SR: 26.1% vs. 26.7% vs. 28.2%, p=NR	HAM-D: 21.3% vs. 24.8%, p=0.16 QIDS-SR: 25.5% vs. 25.0%, p=NR	HAM-D NR, although % decrease in QIDS-SR is presented as 16.4% vs. 21.9% vs. 16.9%, p=NR	Medium
Thase et al., 2007 ¹⁰⁰	122 14	HAM-D Medication: 17.7(6.6) CT: 16.4 (6.2)	Medication: Sertraline: 50 to 200 Bupropion SR: 150 to 400 Venlafaxine XR: 37.5 to 375 CT: 16 sessions	QIDS-SR (Medication vs. CT): 26.7% vs. 22.2% P=0.84	HAM-D (Medication vs. CT): 27.9% vs. 25.0%, p=0.69 QIDS-SR: 26.7% vs. 30.6%, p=0.90	HAM-D (Medication vs. CT): NR % decrease in QIDS-SR is presented as 46.2% vs. 40.7%, p=0.90	Medium

^a Response (≥50 percent decrease in depressive severity) and remission (as defined by authors of individual trials) are measured using the HAM-D unless indicated otherwise.

CT = cognitive therapy; ER = extended release; HAM-D = Hamilton Depression Rating Scale; MADRS = Montgomery-Åsberg Depression Rating Scale; mg = milligram; N = number; NR = not reported; QIDS-SR = Quick Inventory of Depressive Symptoms – Self Report-16; SD = standard deviation; SGA = second-generation antidepressant; SR = sustained release; STAR*D = Sequenced Treatment Alternatives to Relieve Depression Study; vs. = versus; XR = extended release.

Second-Generation Antidepressant Switch Compared With a Nonpharmacologic Treatment Switch (Psychotherapy)

Thase et al. reported a STAR*D-based comparison (rated medium risk of bias) of switching to an SGA versus switching to a nonpharmacologic strategy, namely CT¹⁰⁰ (Table 25). Randomization to a different SGA could assign patients to receive sertraline (50 mg to 200 mg; mean daily dose at end of study was 137 mg), bupropion SR (150 mg to 400 mg, mean daily dose at end of study was 270 mg), or venlafaxine XR (37.5 mg to 375 mg, mean daily dose at end of study was 221 mg); however, the comparisons of SGA with CT consolidated the medications into a single SGA group variable. Response rates assessed on the QIDS-SR showed

no difference between SGA and CT (26.7 percent versus 22.2 percent $p=0.84$). Similarly, remission rates did not differ by treatment arm, whether measured by the HAM-D (27.9 percent versus 25.0 percent, $p=0.69$) or by the QIDS-SR (26.7 percent versus 30.6 percent, $p=0.90$). HAM-D change in depressive severity was not reported, but the percentage decrease in QIDS-SR-16 did not differ between the groups (46.2 percent versus 40.7 percent, $p=0.90$). Neither the time to response nor remission differed for these two switch strategies.

Second-Generation Antidepressant Augmentation Compared With Second-Generation Antidepressant Augmentation

One eligible trial (another from the STAR*D series, also rated medium risk of bias) compared an SGA augmentation strategy with another SGA augmentation strategy (Table 25).¹³⁴

This augmentation comparison randomized patients to the addition of either bupropion SR (150 mg to 400 mg, mean daily dose at end of study was, 268 mg) or buspirone, a nonbenzodiazepine anxiolytic (15 mg/60 mg; mean daily dose at end of study was 41 mg). Response rates did not differ by treatment arm; as reported for the QIDS-SR (31.8 percent versus 26.9 percent, $p=0.21$). Remission rates also did not differ (29.7 percent versus 30.1 percent, $p=0.93$, on HAM-D; 39.0 percent versus 32.9 percent, $p=0.13$, on QIDS-SR). The investigators did not report the mean change in HAM-D score; they did report the percentage decrease in QIDS-SR as favoring bupropion over buspirone (decrease of 25.3 percent versus 17.1 percent, $p<0.04$). Neither the time to response (ranging from 6.3 to 6.8 weeks) nor the time to remission (ranging from 5.4 to 6.3 weeks) differed between the two augmentation options.

Second-Generation Antidepressant Augmentation With Pharmacologic Treatment Compared With Second-Generation Antidepressant Augmentation With Nonpharmacologic Treatment Switch (Psychotherapy)

One eligible trial compared an SGA augmentation with a nonpharmacologic SGA augmentation strategy, CT (Table 26).¹⁰⁰

This augmentation comparison randomized patients to the addition of either a medication (bupropion SR, an antidepressant [150 to 400 mg, mean daily dose at end of study was 283 mg], or buspirone [15 to 60 mg, mean daily dose at end of study was 45.1 mg]) or CT (16 sessions). Response rates did not differ by treatment arm, as reported by the QIDS-SR (28.2 percent versus 35.4 percent, $p=0.25$). Remission rates also did not differ by HAM-D (33.3 percent versus 23.1 percent, $p=0.20$) or by QIDS-SR (33.3 percent versus 30.8 percent, $p=0.78$). Although the mean change in HAM-D score was not provided, the percentage decrease in QIDS-SR revealed no difference between the percentage decrease in depressive severity (39.6 percent versus 40.5 percent, $p=0.83$). Patients assigned to medication group did not differ from the CT group in terms of time to response; however, those receiving medication reached remission faster than those receiving CT (40.1 days versus 55.3 days, $p=0.022$).

Second-line Switch Strategy Compared With Any Augmentation Strategy

We found no eligible trials that directly compared a SGA switch strategy with an augmentation strategy.

Network Meta-analysis of Either Switch or Augment Comparisons Versus Placebo

We did not have enough eligible studies to conduct a network meta-analysis of the relevant treatment options compared with placebo.

Table 26. Second-generation antidepressant augmentation versus another second-generation augmentation strategy: Trial characteristics, main outcomes, and risk of bias ratings

Trial	N	Total Sample Mean Baseline Severity (SD)	Medication Type: mg/day or Psychotherapy: Number of Sessions	Response ^a and Significance Level	Remission ^a and Significance Level	Mean Change in HAM-D Score from Baseline and Significance Level	Risk of Bias Rating
Trivedi et al., 2006 ¹³⁴	565	HAM-D: 15.8 (7.1)	Bupropion SR: 150 to 400	QIDS-SR: 31.8% vs. 26.9% p=0.21	HAM-D: 29.7% vs. 30.1%, p=0.93	NR, although % decrease in QIDS-SR presented as 25.3% vs. 17.1%, p<0.04	Medium
STAR*D	14		Buspirone: 15 to 60 mg		QIDS-SR: 39.0 vs. 32.9%, p=0.13		
Thase et al., 2007 ¹⁰⁰	182	HAM-D: 16.0 (6.7)	Medication: Bupropion SR: 150 to 400	QIDS-SR (Medication vs. CT): 28.2% vs. 35.4% p=0.25	HAM-D (Medication vs. CT): 33.3% vs. 23.1%, p=0.20	HAM-D (Medication vs. CT): NR, Although % decrease in QIDS-SR presented as 39.6% vs. 40.5%, p=0.83	Medium
STAR*D	14		Buspirone: 15 to 60	CT: 16 sessions	QIDS-SR: 33.3% vs. 30.8%, p=0.78		

^a Response (≥ 50 percent decrease in depressive severity) and remission (as defined by authors of individual trials) are measured using the HAM-D unless indicated otherwise.

CT = cognitive therapy; HAM-D = Hamilton Depression Rating Scale; N = number; NR = not reported; QIDS-SR = Quick Inventory of Depressive Symptoms – Self Report; SD = standard deviation; SR = sustained release; vs. = versus; STAR*D = Sequenced Treatment Alternatives to Relieve Depression Study.

KQ 2b. Effect of Severity on the Comparative Effectiveness of Second-line Therapies

We identified two secondary analyses that addressed how depressive severity might moderate the comparative effectiveness of SGAs. Both involved trials described for KQ 2a,^{109,133} although the analysis in one case was published in a separate STAR*D article.¹³⁵

The Lenox-Smith and Jiang trial¹⁰⁹ performed secondary analyses to determine whether comparative effectiveness varied by the level of depression severity (severe versus moderate). In patients with moderate depression (HAM-D ≤ 31), depressive outcomes did not differ measured by either HAM-D or MADRS. However, in the group with HAM-D >31 , some clinical outcomes seemed better in patients receiving venlafaxine (produced by the trial sponsor) than in those receiving citalopram. Remission rates favored venlafaxine, although the difference was not statistically significant (31.6 percent versus 16.4 percent, p=0.08). Changes in depressive severity were better following venlafaxine treatment as measured by HAM-D (p=0.04) but not by MADRS (p=0.09).

A secondary analysis of the original 727 patients in the SGA switch analysis explored whether several variables, including depressive severity, might differentially moderate the effectiveness of the medications being compared.¹³³ The analysis assessed the effect of mild or moderate versus severe depression (defined as QIDS-SR ≥ 16) on remission rates. The odds of remission for patients with severe depression (relative to mild/moderate) were lower for all three

medications (bupropion SR 0.38, sertraline 0.38, venlafaxine XR 0.25), but the differences among the medications were not statistically significant (p=0.70).

KQ 3: Comparative Risks of Treatment Harms

In this section, we distinguish adverse events from serious adverse events based on the Food and Drug Administration (FDA) classification. The FDA defines adverse events as any medical occurrence associated with the use of an intervention, whether or not it is considered related to the intervention.¹³⁶ A serious adverse event is any medical occurrence that results in death, is life threatening, requires hospitalization, results in persistent or significant disability or incapacity, or is a congenital birth defect. We also report the findings of the one eligible trial providing information about how the risk of harms for our interventions of interest varies by baseline severity of MDD.⁸⁶ The trial’s authors collected data to address this issue but reported findings only qualitatively.

As we have done in previous sections, here we provide an overview of the articles, including the number of trials, for each comparison (Table 27); key points; and a detailed synthesis. All trials are of low or medium risk of bias except if noted otherwise. In Appendix E, we present summary of findings tables for the important outcomes. These tables describe basic information on the available evidence, summarize differences in risks of harms using absolute and relative effect measures, and present the SOE grades for each outcome.

Table 27. Number of trials for each comparison of interest

Comparison Category	Comparison Intervention	Number of Trials
SGA vs. Psychological Interventions	SGA vs. Behavior therapies/behavior modification	0
	SGA vs. CBT	12
	SGA vs. Humanistic therapies	0
	SGA vs. Integrative therapies	4
	SGA vs. Psychodynamic therapies	4
	SGA vs. Third-wave CBTs	2
SGA vs. Complementary and Alternative Medicine	SGA vs. Acupuncture	5
	SGA vs. Omega-3 fatty acids	2
	SGA vs. SAMe	1
	SGA vs. St. John’s wort	12
	SGA vs. Meditation	0
SGA vs. Yoga	0	
SGA vs. Exercise	SGA vs. Exercise	2
SGA Switch vs. SGA Switch	Switch to citalopram from different SSRI vs. Switch to venlafaxine from different SSRI	1

CBT = cognitive behavioral therapy; SGA = second generation antidepressant; SSRI = serotonin-specific reuptake inhibitor; vs. = versus.

Overview

We analyzed adverse events data from 43 head-to-head efficacy trials of 6,253 patients. Table 27 summarizes the number of trials that contributed some information to the assessment of the comparative risks of harms.

As described in more detail in the Methods section, we intended to include data from head-to-head trials *and* nonrandomized trials for assessing comparative risk of harms. However, we did not find any nonrandomized trials that met our eligibility criteria.

Few trials that examined the comparative effectiveness of SGAs with other eligible treatment options adequately determined differences in harms. Three trials, two of psychological

interventions^{89,111} and one of acupuncture,¹¹⁸ did not report any data on harms. None of the trials that reported harms data used objective scales such as the UKU-SES (Utvalg for Kliniske Undersogelser Side Effect Scale) or the SAFTEE-SI (Systematic Assessment for Treatment of Emergent Events-Specific Inquiry). Most trials combined spontaneous patient-reported adverse events with a regular clinical examination by an investigator. Determining whether assessment methods were unbiased and adequate was often difficult. Rarely did authors report whether adverse events were prespecified and defined. Short trial durations and small sample sizes also limited the validity of adverse event assessment in many trials.

No trials were designed to assess specific adverse events as primary outcomes. Detailed information on included trials can be found in KQ 1.

Key Points

Second-Generation Antidepressants Compared With Psychological Interventions

- Psychological interventions as a class and SGAs led to similar overall discontinuation rates (7 RCTs, moderate SOE). In contrast, discontinuations because of adverse events occurred less often after psychological interventions than SGAs (7 RCTs, moderate SOE). Patients receiving psychological interventions had a numerically but not statistically significant higher risk of suicidality than those given SGAs (7 RCTs, low SOE).
- The combination of psychological interventions as a class and SGAs did not produce statistically different discontinuation rates compared with patients treated with SGA monotherapy after 12 weeks of follow-up (3 RCTs, low SOE). In contrast, overall discontinuation rates were lower following SGA monotherapy than following combination treatment after 96 weeks of follow-up (1 RCT, low SOE). Adding psychological interventions to SGA treatment did not produce statistically different rates of discontinuation because of adverse events compared with SGA monotherapy after 12 weeks of follow-up (1 RCT, low SOE).
- We did not find any eligible trials comparing behavior therapies with SGAs (insufficient SOE).
- CBT and SGAs led to similar overall discontinuation rates (4 RCTs, moderate SOE). Rates of discontinuation because of adverse events (3 RCTs, low SOE) were numerically lower but not statistically significant for patients receiving CBT.
- Adding CBT to and SGA treatment did not lead to statistically different rates of overall discontinuation and discontinuation because of adverse events compared with SGA monotherapy (2 RCTs each, both low SOE).
- We did not find any eligible trials comparing humanistic therapies with SGAs (insufficient SOE).
- The evidence was insufficient to draw conclusions about any outcomes for integrative therapy alone or in combination with SGAs compared with SGA monotherapy.
- Psychodynamic therapies did not lead to statistically different rates of overall discontinuation compared with SGAs over the course of 48 weeks and 96 weeks of followup (1 RCT each, both low SOE). Psychodynamic therapy did not lead to

statistically different rates of suicidality compared with SGAs after 8 weeks and 96 weeks of followup (1 RCT each, both low SOE).

- Adding psychodynamic therapy to SGA treatment did not produce statistically different rates of discontinuation compared with patients receiving SGA monotherapy (1 RCT, low SOE). The addition of psychodynamic therapy to SGA treatment also did not lead to statistically different rates of suicidality compared with SGA monotherapy after 96 weeks of followup (1 RCT, low SOE).
- Third-wave CBT lead to lower overall discontinuation rates than SGAs (2 RCTs, low SOE).
- The evidence was insufficient to draw conclusions about the comparative overall risk of serious adverse events between psychological interventions in general and SGAs.

Second-Generation Antidepressants Compared With Complementary and Alternative Medicines

- Acupuncture led to a higher overall discontinuation rate than SGAs (1 RCT, low SOE). We were unable to draw conclusions about the overall comparative risk of harms of acupuncture and SGAs from direct evidence because of concerns about risk of bias and imprecision (1 RCT, insufficient SOE). However, indirect evidence from a systematic review that included depressive disorders other than MDD indicated that acupuncture had a lower overall risk of harms than SGAs (, 21 RCTs, moderate SOE).
- Adding acupuncture to SGA treatment led to an overall risk of adverse events (1 RCT, low SOE), overall discontinuation rates (3 RCTs, moderate SOE), and rates of discontinuation because of adverse events (2 RCTs, low SOE) that were similar to those among patients receiving SGA monotherapy.
- Omega-3-fatty acids did not lead to statistically different rates of overall discontinuation and discontinuation because of adverse events compared with SGAs (1 RCT each, both low SOE).
- Adding omega-3-fatty acids to SGA treatment also did not lead to statistically different rates of overall discontinuation and discontinuation because of adverse events (1 RCT each, both low SOE).
- SAME did not lead to statistically different overall discontinuation rates compared with patients treated with SGAs (1 RCT, low SOE,).
- St. John's wort led to lower rates of overall discontinuation (12 RCTs, moderate SOE) and discontinuation because of adverse events (11 RCTs, moderate SOE) than did SGAs. The overall risk of adverse events was also lower among patients receiving St. John's wort than those receiving SGAs, although this difference was statistically nonsignificant (8 RCTs, moderate SOE). In contrast, the risk of serious adverse events did not differ between patients receiving St. John's wort and those receiving SGAs (5 RCTs, low SOE).
- We did not find any eligible trials comparing meditation or yoga with SGAs (insufficient SOE).

Second-Generation Antidepressant Switches Compared With Other Second-Generation Antidepressant Switches

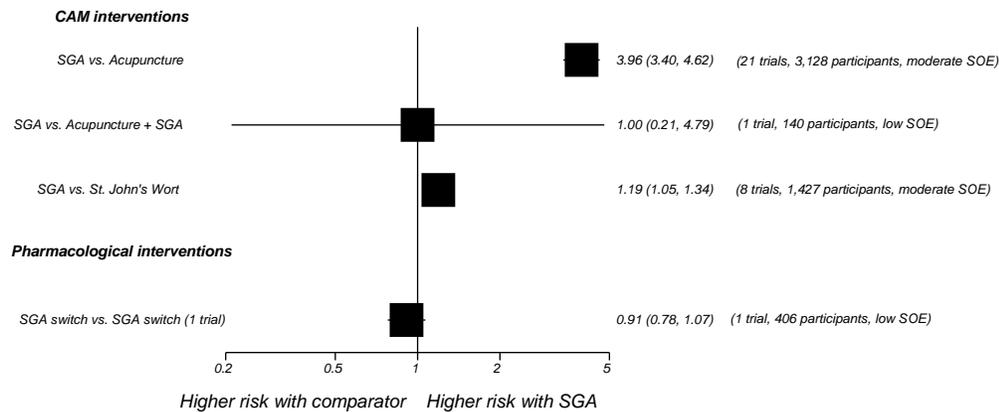
- Switching to citalopram and switching to venlafaxine led to similar risks of overall harms and overall discontinuation rates (1 RCT each, both low SOE).

Second-Generation Antidepressants Compared With Exercise

- Exercise and SGAs led to similar overall discontinuation rates (2 RCTs, moderate SOE). Discontinuation rates because of adverse events were lower for exercise than SGAs (2 RCTs, low SOE,).
- Adding exercise to SGA treatment led to overall discontinuation rates and discontinuation rates because of adverse events that were similar to those among patients receiving SGA monotherapy (1 RCT each, both low SOE,).

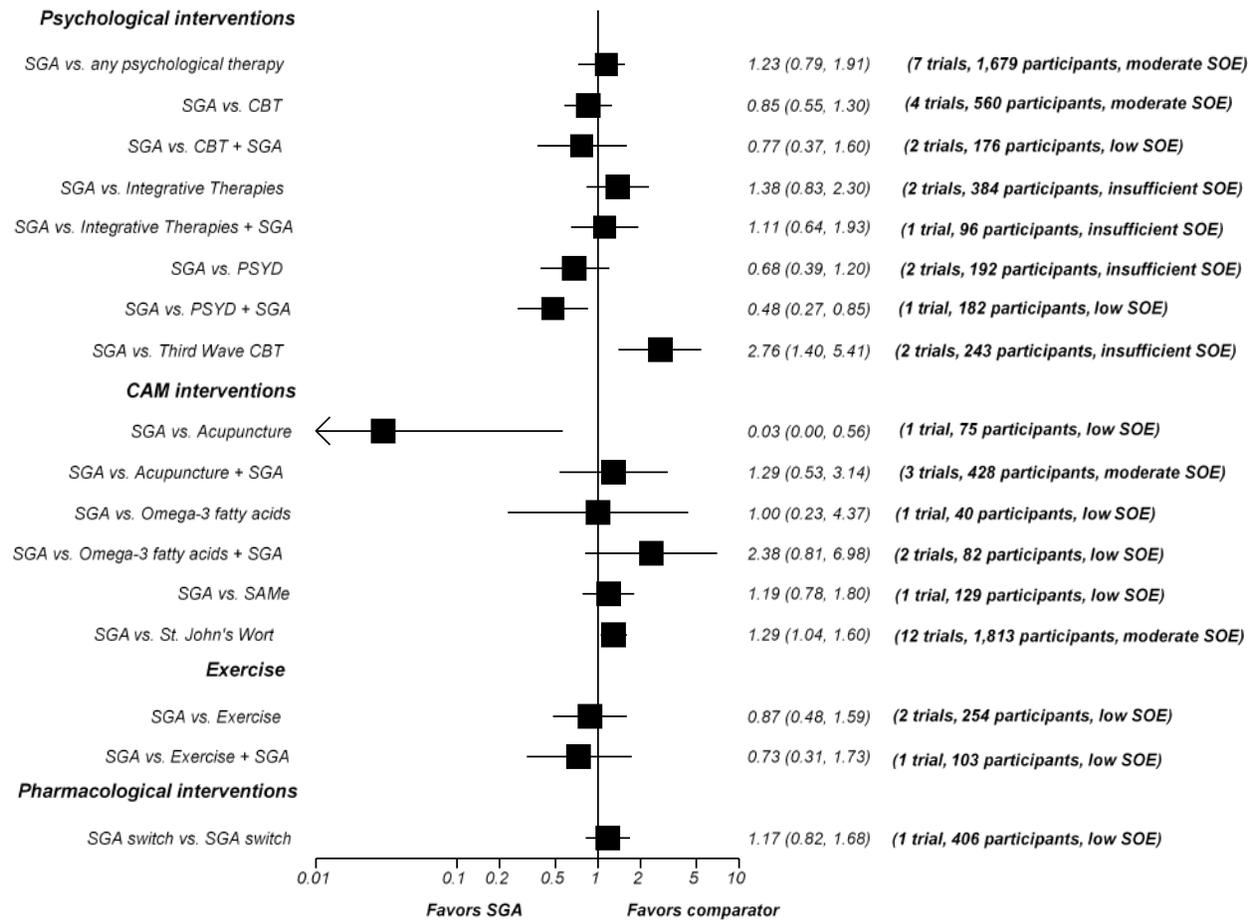
Figures 13, 14, and 15 graphically display relative risks of SGAs compared with other interventions for overall harms, overall discontinuation, and discontinuation because of adverse events.

Figure 13. Relative overall risk of harms of SGAs compared with other eligible interventions.



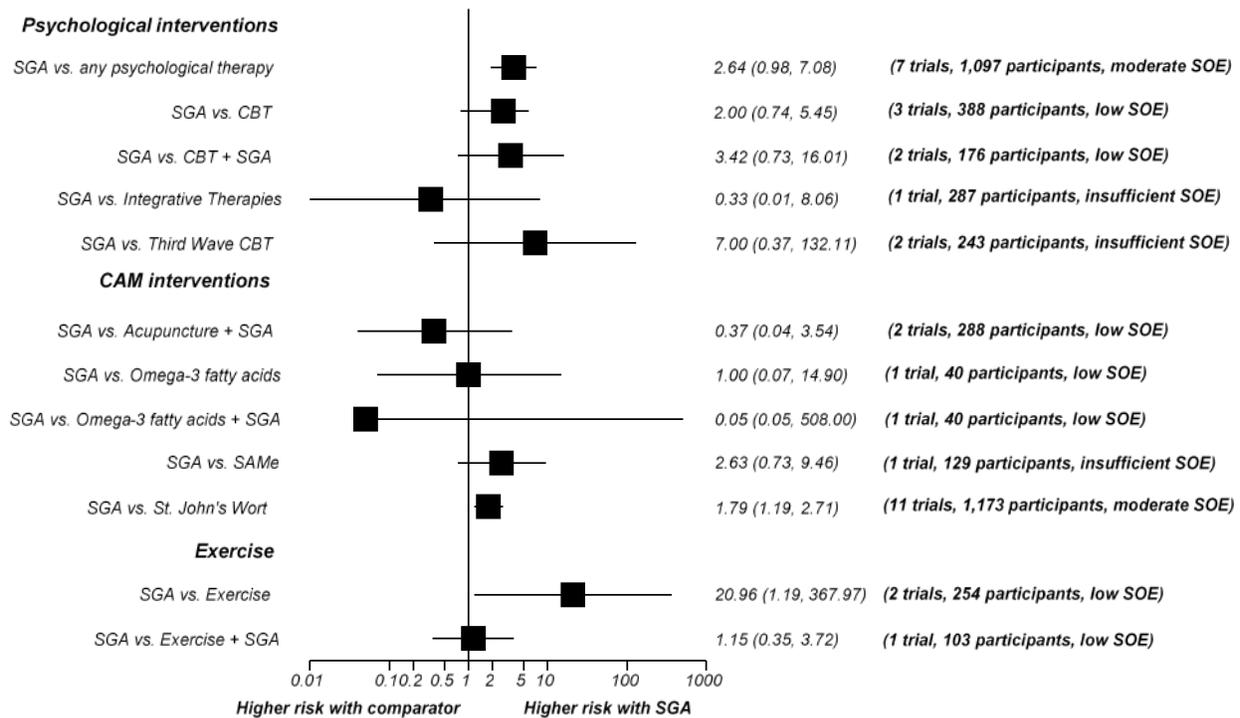
CAM = complementary and alternative medicine; SGA = second-generation antidepressant; SOE = strength of evidence; vs. = versus.

Figure 14. Relative risks of overall discontinuation of SGAs compared with other eligible interventions.



CAM = complementary and alternative medicine; CBT = cognitive behavioral therapy; KQ = Key Question; SAME = S-adenosyl-L-methionine; SGA = second-generation antidepressant; SOE = strength of evidence; vs. = versus.

Figure 15. Relative risks of discontinuation because of adverse events rates of SGAs compared with other eligible interventions



CAM = complementary and alternative medicine; CBT = cognitive behavioral therapy; KQ = Key Question; SAMe = S-adenosyl-L-methionine SGA = second-generation antidepressant; SOE = strength of evidence; vs. = versus.

Key Question 3a. Comparative Risks of Harms Between Pharmacological and Nonpharmacological Interventions

Detailed Synthesis: Overall Risk of Experiencing Harms and Discontinuation of Treatment

This section provides a detailed synthesis of the comparative risk of experiencing harms and discontinuing treatment. In general, reporting of adverse events was scarce, and we were able to draw only a few conclusions with certainty from the available evidence. Even common adverse events associated with SGAs, such as diarrhea, dizziness, dry mouth, headache, insomnia, nausea, vomiting, and weight gain, were rarely assessed or reported. Similarly, few trials addressed adverse events that are commonly associated with psychotherapies, such as worsening of symptoms or onset of new depression-associated symptoms.

Second-Generation Antidepressants Compared With Psychological Interventions

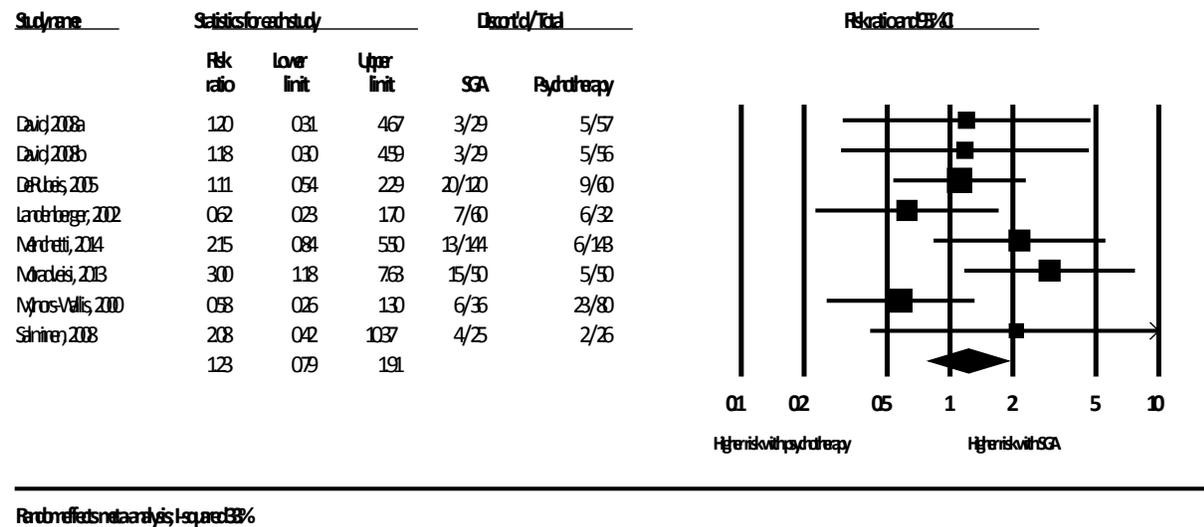
We first present the available evidence on the comparative risk of harms for SGAs and psychological treatments as a class. Next, we summarize the evidence for each included psychological intervention. As in KQ 1, we use classifications of the Cochrane Collaboration Depression, Anxiety and Neurosis Review Group.⁶³

Second-Generation Antidepressants Compared With Any Psychological Interventions

We conducted meta-analyses comparing overall discontinuation rates, discontinuation rates because of lack of efficacy, and discontinuation rates because of adverse events for patients treated with any SGA compared with those treated with any psychological intervention. Interventions for these comparisons were limited to fluoxetine, fluvoxamine, paroxetine, and sertraline (for the SGAs) and behavioral activation, cognitive therapy, problem solving therapy, rational emotive behavior therapy, and short-term psychodynamic supportive psychotherapy (for the psychological interventions).

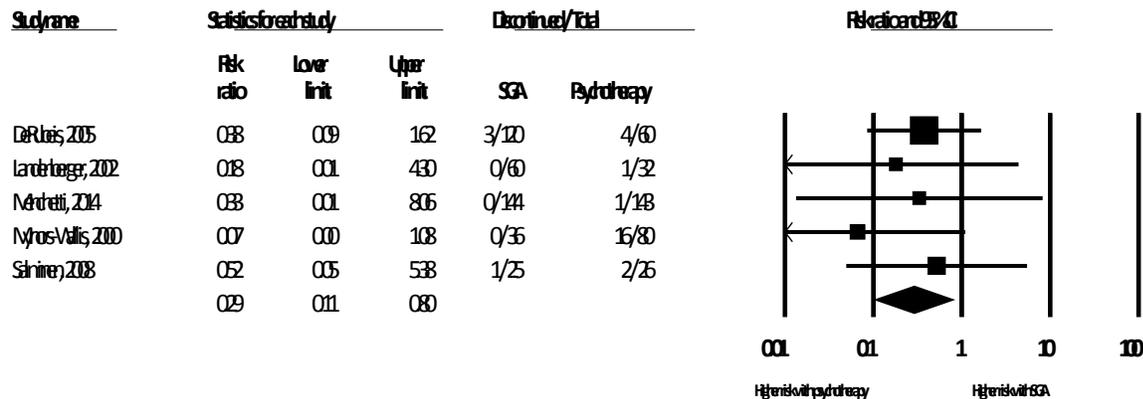
Overall discontinuation rates were similar between SGAs and psychological interventions (14 percent versus 12 percent; RR, 1.23; 95% CI, 0.79 to 1.91; Figure 16). Discontinuation rates because of lack of efficacy were statistically significantly lower for patients treated with SGAs than for patients treated with psychological interventions (1 percent versus 7 percent; RR, 0.29; 95% CI, 0.11 to 0.80; Figure 17). In contrast, discontinuation rates because of adverse events were more than twice as high for patients receiving SGAs than for those treated with psychological interventions (7 percent versus 2 percent; RR, 2.64; 95% CI, 0.98 to 7.08; Figure 18). The numbers of events of discontinuation because of lack of efficacy and discontinuation because of adverse events, however, were low; therefore, we urge caution in interpreting these results.

Figure 16. Overall discontinuation rates comparing second-generation antidepressants with psychological interventions as a class



CI = confidence interval; Discont'd = discontinued; SGA = second-generation antidepressant(s).

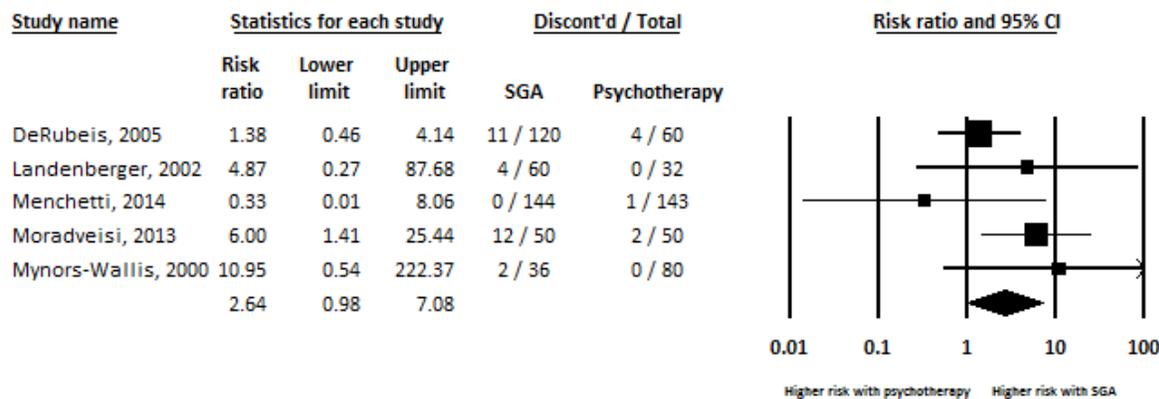
Figure 17. Discontinuation rates because of lack of efficacy comparing second-generation antidepressants with psychological interventions as a class



Random effects meta-analysis; I-squared 0%

CI = confidence interval; SGA = second-generation antidepressant(s).

Figure 18. Discontinuation rates because of adverse events comparing second-generation antidepressants with psychological interventions as a class



Random effects meta-analysis; I-squared 23%

CI = confidence interval; Discontin'd = discontinued; SGA = second-generation antidepressant(s).

For sensitivity analyses, we added high risk of bias trials to the meta-analytic models. The results of the analyses of overall discontinuation rates were consistent with the results presented above (RR, 1.17; 95% CI, 0.88 to 1.55). Differences in the rates of discontinuation because of lack of efficacy, however, were no longer statistically significantly different (RR, 1.18; 95% CI, 0.07 to 20.13) when we added a high risk of bias trial⁸⁶ that strongly favored psychological treatments over SGAs. Differences in the rates of discontinuation because of adverse events reached statistical significance favoring psychological interventions over SGAs when we included two high risk of- bias trials (RR, 3.21; 95% CI, 1.53 to 6.75).^{86,91}

Second-Generation Antidepressants Compared With Behavior Therapy/Behavior Modification

We found no eligible trials that compared an SGA with behavior therapy/behavior modification.

Second-Generation Antidepressants Compared With Cognitive Behavioral Therapy

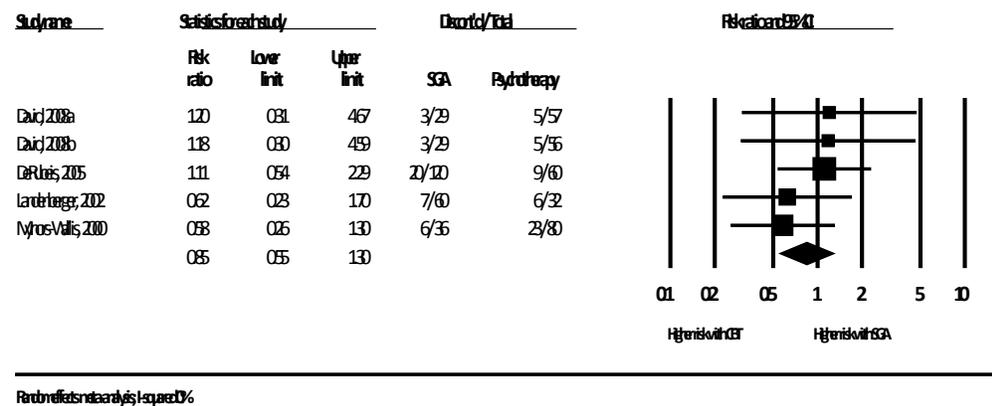
Of 11 trials included for KQ 1a, 7 reported limited data on adverse events (see KQ 1, Table 11 for more details on trial design and dosing).^{78,82,85,88,102,113,116,125} All but one trial compared SGA monotherapy with CBT alone; Lam and colleagues compared escitalopram monotherapy with escitalopram plus CBT.¹⁰² Treatment durations ranged from 8 weeks to 1 year.

Second-Generation Antidepressants Versus Cognitive Behavioral Therapy: Monotherapy Comparisons

Seven trials provided limited information on the comparative risk of harms of SGA monotherapy compared with CBT.^{78,82,85,87,88,113,116,125} In these trials, SGAs were limited to escitalopram, fluoxetine, fluvoxamine, paroxetine, and venlafaxine. None of the trials provided information on the comparative risk of specific adverse events, even common adverse events of SGAs. Only one trial provided data on the proportions of patients who experienced any adverse events.^{82,125} About 16 percent of patients treated with an SGA experienced adverse events as did 0 to 2 percent of patients treated with CBT. Particularly for SGAs, reported adverse event rates appear to underestimate substantially the actual risk. A comprehensive systematic assessment of the risk of harms for SGAs reported that an average of 60 percent of patients treated with SGAs experience at least one adverse event during treatment.³²

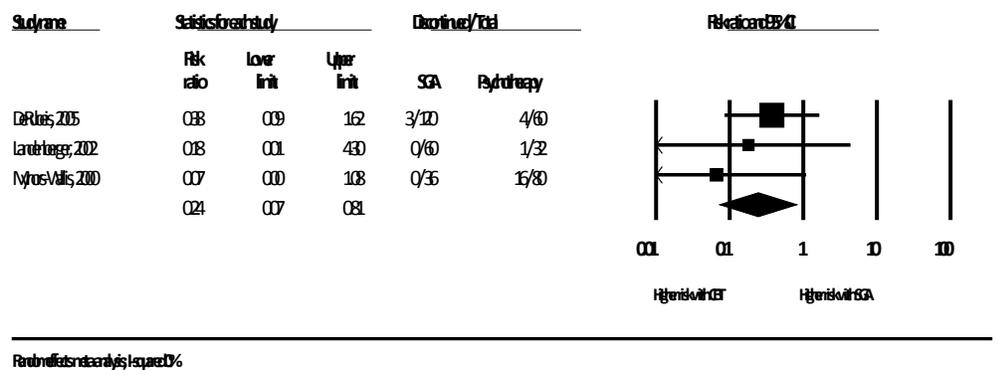
Overall discontinuation rates were similar for patients treated with SGAs or CBT (14 percent versus 17 percent; RR, 0.85; 95% CI, 0.55 to 1.30, Figure 19). Discontinuation rates because of lack of efficacy were statistically significantly lower for patients treated with SGAs than for those treated with CBT (1 percent versus 12 percent; RR, 0.24; 95% CI, 0.07 to 0.81, Figure 20). By contrast, discontinuation rates because of adverse events were numerically higher for patients on SGAs than for patients treated with CBT (8 percent versus 2 percent; RR, 2.00; 95% CI, 0.74 to 5.45, Figure 21), but the difference did not reach statistical significance. The numbers of events for discontinuation because of lack of efficacy and the numbers of events for discontinuation because of adverse events, however, were very low; therefore, we urge caution in interpreting these results.

Figure 19. Overall discontinuation rates comparing second-generation antidepressants with cognitive behavioral therapies



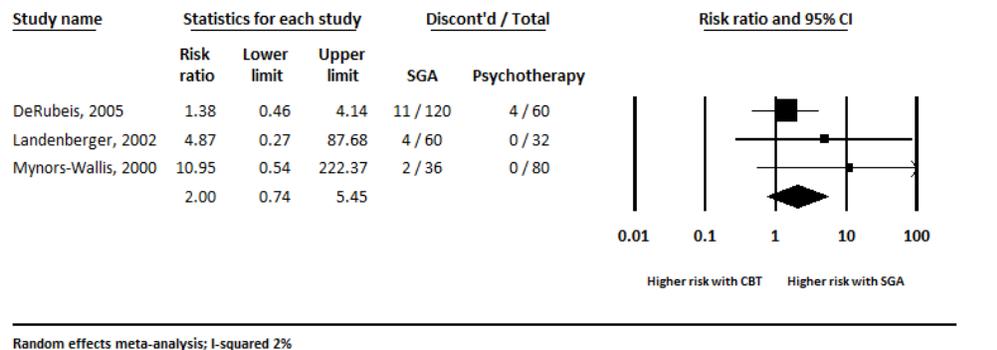
CBT = cognitive behavioral therapy; CI = confidence interval; Discont'd = discontinued; SGA = second-generation antidepressant(s).

Figure 20. Discontinuation rates because of lack of efficacy comparing second-generation antidepressants with cognitive behavioral therapies



CBT = cognitive behavioral therapy; CI = confidence interval; SGA = second-generation antidepressant(s).

Figure 21. Discontinuation rates because of adverse events comparing second-generation antidepressants with cognitive behavioral therapy



CBT = cognitive behavioral therapy; CI = confidence interval; Discont'd = discontinued; SGA = second-generation antidepressant(s).

In sensitivity analyses, we added high risk of bias trials to the meta-analytic models. The differences in overall discontinuation rates (RR, 1.00; 95% CI, 0.73 to 1.36) and rates of discontinuation because of lack of efficacy (RR, 0.47; 95% CI, 0.08 to 2.84) remained similar to the original analyses. Discontinuation rates because of adverse events were statistically significantly higher for patients receiving SGAs than for those receiving CBT when we included two high risk of bias trials (RR, 3.09; 95% CI, 1.14 to 8.35).^{86,91}

Second-Generation Antidepressants Versus Cognitive Behavioral Therapy: Combination Comparisons

The only trial that compared an SGA (escitalopram) with a combination of escitalopram and telephone CBT did not report information on specific adverse events.¹⁰² After 12 weeks, overall discontinuation rates (13 percent versus 23 percent) were numerically lower for patients treated with SGAs than for those treated with telephone CBT. Discontinuation rates because of adverse events were similar for the two treatment groups (6 percent versus 4 percent).

Second Generation Antidepressants Compared With Humanistic Therapies

We found no eligible trials that compared an SGA with humanistic therapies.

Second Generation Antidepressants Compared With Integrative Therapies

Of four trials of integrative therapies included in KQ 1, none provided information on the comparative risk of specific adverse events.^{80,84,92,127} Two trials provided limited data on discontinuation rates and overall rates of harms comparing patients receiving SGAs with patients receiving integrative therapies (see KQ 1, Table 12 for more details on trial design and dosing).^{80,84,92,127}

Second-Generation Antidepressants Versus Integrative Therapies: Monotherapy Comparisons

Neither of the two available trials comparing SGAs (nefazodone, citalopram, or sertraline) with integrative therapy reported on specific adverse events. Discontinuation rates ranged from 9 percent to 36 percent for patients treated with SGAs (citalopram, escitalopram, or nefazodone) and from 14 percent to 32 percent for patients receiving integrative therapies.^{84,92}

Second-Generation Antidepressants Versus Integrative Therapies: Combination Comparisons

One trial compared an SGA (nefazodone) with a combination of nefazodone and integrative therapy.⁸⁴ Authors did not report any data on adverse events except overall discontinuation rates, which were similar between the nefazodone monotherapy and combination treatment groups after 12 weeks of followup (36 percent versus 33 percent, respectively).

Second-Generation Antidepressants Compared With Psychodynamic Therapies

None of the four trials included for KQ 1 reported on the risk of specific adverse events (see KQ 1, Table 13 for more details on trial design and dosing).^{79,81,90,101,128}

Second-Generation Antidepressants Versus Psychodynamic Therapies: Monotherapy Comparisons

Four trials compared SGA monotherapies (fluoxetine, sertraline, venlafaxine) with psychodynamic therapies.^{79,81,90,101,128} One small trial (N=51) comparing fluoxetine monotherapy with psychodynamic therapy reported that overall rates of adverse events were similar for patients treated with fluoxetine or psychodynamic therapy (4 percent versus 8 percent) after 16 weeks of followup.⁸¹ Overall discontinuation rates ranged from 16 percent to 26 percent for patients treated with SGAs (fluoxetine, venlafaxine) and from 8 percent to 27 percent for patients who received psychodynamic therapy. None of the four trials reported any data on discontinuation because of adverse events.

Second-Generation Antidepressants Versus Psychodynamic Therapies: Combination Comparisons

The only trial that compared an SGA monotherapy (fluoxetine) with a combination of fluoxetine and long-term psychodynamic therapy did not report any data on differences in adverse events.⁹⁰ After 96 weeks, patients in both groups had similar overall discontinuation rates (about 26 percent versus 32 percent, respectively).

Second-Generation Antidepressants Compared With Third Wave Cognitive Behavioral Therapy

Second-Generation Antidepressants Versus Third Wave Cognitive Behavioral Therapy: Monotherapy Comparisons

Two trials,^{86,110} both high risk of bias, compared SGAs (paroxetine or sertraline) with third-wave CBT. Neither study reported overall risks of adverse events. Overall discontinuation rates ranged from 25 percent to 30 percent for patients treated with SGAs and from 9.3 percent to 10 percent for patients who received third-wave CBT. Similarly, rates of discontinuation because of adverse events were higher among patients treated with SGAs than those treated with third-wave CBT, ranging from 6 percent to 24 percent and from 2.3 percent to 4 percent, respectively.

Second-Generation Antidepressants Versus Third Wave Cognitive Behavioral Therapy: Combination Comparisons

We did not find any trials addressing this comparison.

Second Generation Antidepressants Compared With Complementary and Alternative Medicines

Second-Generation Antidepressants Compared With Acupuncture

For the comparison of SGAs with acupuncture, four efficacy trials reported data on harms or discontinuation rates.^{94,96,115,117} We rated one trial as high risk of bias.⁹⁶ Overall, the available data were sparse and prevented us from drawing any firm conclusions about the comparative risk of harms between SGAs and acupuncture. One trial reported overall rates of adverse events.¹¹⁵ Even adverse events that are specifically associated with acupuncture, such as fainting after needle insertion or needle-related pain, were not reported consistently. Likewise, typical SGA-associated adverse events, such as nausea, diarrhea, headache, and dizziness, were not reported adequately.

Second-Generation Antidepressants Versus Acupuncture: Monotherapy Comparisons

Two trials,^{96,114} one that had a medium risk of bias¹¹⁴ and the other a high risk of bias,⁹⁶ compared fluoxetine with acupuncture (electroacupuncture, see KQ 1, Table 14 for more details on trial design and dosing). The medium risk of bias study collected data on overall adverse events but not any type of discontinuation, finding that the rates of any adverse event were similar between patients treated with fluoxetine (4.2 percent) and those treated with acupuncture (6 percent).¹¹⁴ The high risk of bias study did report overall discontinuation data showing that rates (0 percent versus 36 percent, respectively) were numerically lower for patients treated with fluoxetine than those treated with acupuncture.⁹⁶

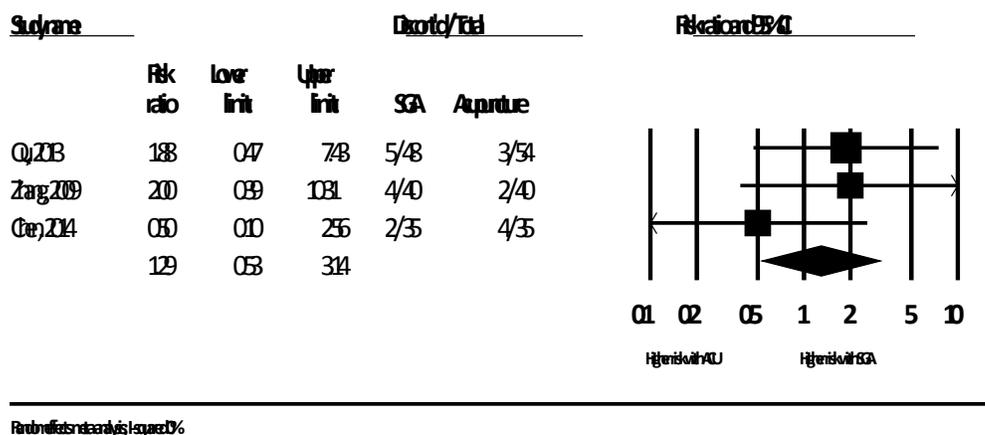
A systematic review that did not meet our eligibility criteria because it included depressive disorders other than MDD provided the most comprehensive assessment of the comparative risk of harms between SGAs and acupuncture.¹³⁷ Based on evidence from 25 RCTs, the authors reported that adverse event rates were statistically significantly higher in patients treated with SGAs than in those receiving active or sham acupuncture.¹³⁷ Overall, 40 percent of patients treated with SGAs reported adverse events compared with 10 percent of patients undergoing acupuncture ($p < 0.001$). The most commonly reported adverse events of patients treated with an SGA were headache, insomnia, and tiredness. Patients treated with acupuncture reported needling pain, dizziness, and nausea as the most common adverse events.

Second-Generation Antidepressants Versus Acupuncture: Combination Comparisons

Three trials compared SGA monotherapy (fluoxetine or paroxetine) with a combination of an SGA with acupuncture (see KQ 1, Table 14 for more details on trial design and dosing).^{94,115,117} One trial reported no statistically significant differences in specific adverse events, such as headache, dizziness, insomnia, and somnolence.¹¹⁵ The other two trials did not report any data on adverse events.

Data were available from all three trials for a meta-analysis of overall discontinuation rates.^{94,115,117} Overall discontinuation rates were similar for patients treated with SGAs or a combination of an SGA with acupuncture (8.9 percent versus 7 percent, respectively; RR, 1.29; 95% CI, 0.53 to 3.14; Figure 22). Rates of discontinuation because of adverse events ranged from 0 percent to 3.4 percent and did not differ significantly between treatment groups.^{94,117}

Figure 22. Meta-analysis of overall discontinuation rates comparing second-generation antidepressants with combination of acupuncture and SGA



ACU = acupuncture; CI = confidence interval; Discont'd = discontinued; SGA = second-generation antidepressant(s).

Second-Generation Antidepressants Compared With Omega-3 Fatty Acids

Second-Generation Antidepressants Versus Omega-3 Fatty Acids: Monotherapy Comparisons

The Iranian trial (high risk of bias) compared fluoxetine with omega-3 fatty acids (EPA, see KQ 1, Table 15 for more details on trial design and dosing).¹¹² The authors did not report whether the risks of specific adverse events differed in any statistically significant way between patients treated with fluoxetine and patients treated with EPA monotherapy. For the two treatment groups, rates of overall discontinuation (both 15 percent) and discontinuation because of adverse events (both 5 percent) were the same.

Second-Generation Antidepressants Versus Omega-3 Fatty Acids: Combination Comparisons

Two trials (both high risk of bias) compared SGAs (citalopram or fluoxetine) with combinations of SGAs and omega-3 fatty acids (see KQ 1, Table 15 for more details on trial design and dosing).^{95,112} Overall, the available data on harms were sparse and did not allow us to draw firm conclusions about the comparative risk of harms between SGA monotherapy and the combination of SGAs with omega-3 fatty acids. The trial comparing citalopram with a combination of citalopram and omega-3 fatty acids did not report on specific adverse events.⁹⁵ In the other trial, the authors did not report whether the risks of specific adverse events differed significantly between patients treated with fluoxetine or fluoxetine combined with EPA.¹¹² Fluoxetine monotherapy and combined fluoxetine and EPA treatment groups did not differ significantly in rates of overall discontinuation (20 percent versus 20 percent) or discontinuation because of adverse events (5 percent versus 10 percent, respectively).

Second-Generation Antidepressants Compared With S-Adenosyl-L-Methionine (SAME)

Second-Generation Antidepressants Versus SAME: Monotherapy Comparisons

The only trial that compared an SGA (escitalopram) with SAME (see KQ 1 for more details on trial design and dosing) reported two statistically significant differences in adverse events between treatment groups.⁹³ Significantly more patients treated with escitalopram than SAME experienced anorgasmia (18.2 percent versus 3.4 percent, respectively; $p=0.011$) or hot flashes (7.3 percent versus 0.0 percent, respectively; $p=0.017$) during 12 weeks of followup. Overall discontinuation rates (54 percent versus 44 percent, respectively) and discontinuation rates because of adverse events (12 percent versus 5 percent, respectively) were numerically higher for patients treated with escitalopram than for those treated with SAME. The differences, however, did not reach statistical significance.

Second-Generation Antidepressants Versus SAME: Combination Comparisons

We found no eligible trials that compared an SGA with a combination of SGA and SAME.

Second-Generation Antidepressants Compared With St. John's Wort

All 12 trials comparing SGAs with St. John's wort provided data on harms or discontinuation rates (see KQ 1, Table 16 for more details on trial design and dosing).^{97,103-108,119-123} Two were rated as high risk of bias.^{106,107}

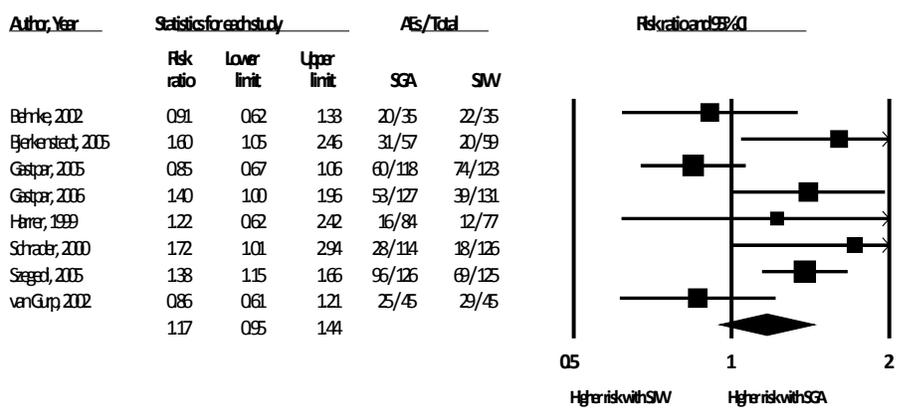
Second-Generation Antidepressants Versus St. John's Wort: Monotherapy Comparisons

Enough data were available to warrant meta-analyses of overall rates of adverse events and rates of overall discontinuation, discontinuation because of adverse events and because of lack of efficacy, and overall rates of serious adverse events.

Patients treated with SGAs experienced higher overall rates of adverse events, overall discontinuation, and discontinuation because of adverse events than patients treated with St. John's wort. Discontinuation rates because of lack of efficacy were similar between the treatment groups. In the following paragraphs, we describe the results of these meta-analyses in more detail.

Eight trials, all assigned a low or medium risk of bias rating, reported overall rates of adverse events.^{104,105,108,119-123} SGAs were limited to citalopram, fluoxetine, paroxetine, and sertraline. Our random-effects meta-analysis indicated a numerically but not statistically significantly higher overall risk of adverse events for patients treated with SGAs than those treated with St. John's wort (47 percent versus 39 percent, respectively; RR, 1.17; 95% CI, 0.95 to 1.44; Figure 23). However, when we conducted a sensitivity analysis using a fixed effect meta-analytic model in place of a random effects model, overall rates of adverse events were statistically significantly higher among patients given SGAs than those given St. John's wort (RR, 1.16; 95% CI, 1.04 to 1.29; forest plot not shown).

Figure 23. Meta-analysis of overall risk for adverse events comparing second-generation antidepressants with St. John’s wort



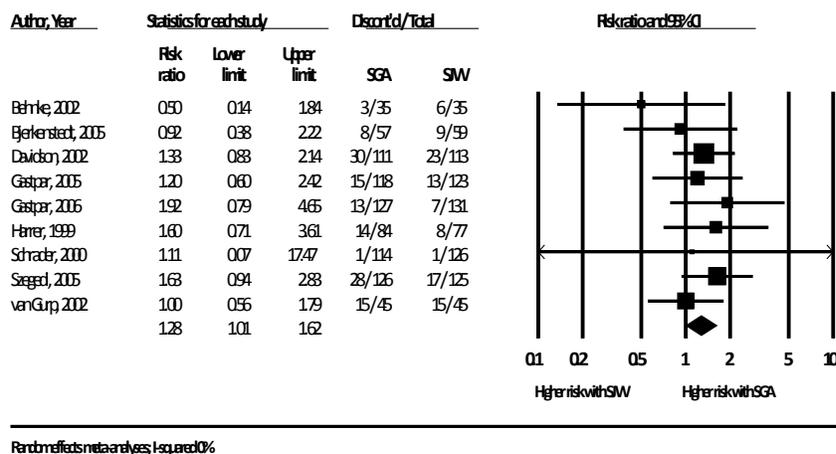
Random-effects meta-analysis; I-squared 68%

AEs = adverse events; CI = confidence interval; SGA = second-generation antidepressant(s); SJW = St. John’s wort.

Of note, a high degree of heterogeneity was present because three trials found a *higher* rate of overall adverse events for patients treated with St. John’s wort, although none of these trials’ risk ratios were statistically significant.^{105,120,121} An exploratory analysis to identify the cause of the heterogeneity did not reveal any systematic differences between these three trials and the five showing a higher rate with SGAs; we surmise that the between-trial differences can probably be attributed to chance.

All 12 trials comparing SGAs with St. John’s wort extracts, of which three had a high risk of bias rating,^{103,106,107} reported overall discontinuation rates. Random-effects meta-analysis findings based on low and medium risk of bias trials showed that patients treated with SGAs had a statistically significantly higher risk of overall discontinuation than those treated with St. John’s wort (16 percent versus 12 percent, respectively; RR, 1.28; 95% CI, 1.01 to 1.62; Figure 24). The results of our sensitivity analysis, which included the three high risk of bias trials, were similar and significant (18 percent versus 14 percent, respectively; RR, 1.25; 95% CI, 1.02 to 1.54; forest plot not shown). Using a fixed effect meta-analytic model instead of a random effects model did not change the findings.

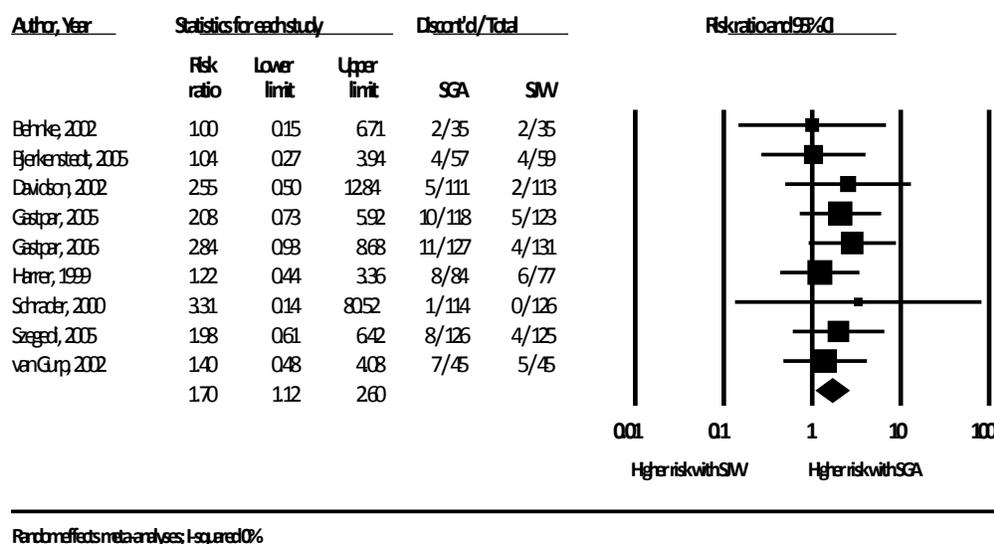
Figure 24. Meta-analysis of overall discontinuation rates comparing second-generation antidepressants with St. John's wort



CI = confidence interval; Discont'd = discontinued; SGA = second-generation antidepressant(s); SJW = St. John's wort.

Eleven of 12 trials, of which two were rated high risk of bias,^{103,107} reported rates of discontinuation because of adverse events.^{97,103-105,107,108,119-123} Our random-effects meta-analysis found a statistically significantly higher rate of discontinuation because of adverse events among patients treated with SGAs than those treated with St. John's wort (6.9 percent versus 3.8 percent, respectively; RR, 1.70; 95% CI, 1.12 to 2.60; Figure 25). Our sensitivity analysis, which included the same two trials mentioned above, found similar significant results (SGA: 6.8 percent versus St. John's wort: 3.8 percent; RR, 1.69; 95% CI, 1.12 to 2.54; data not shown). Using a fixed effect meta-analytic model instead of a random effects model did not change the results.

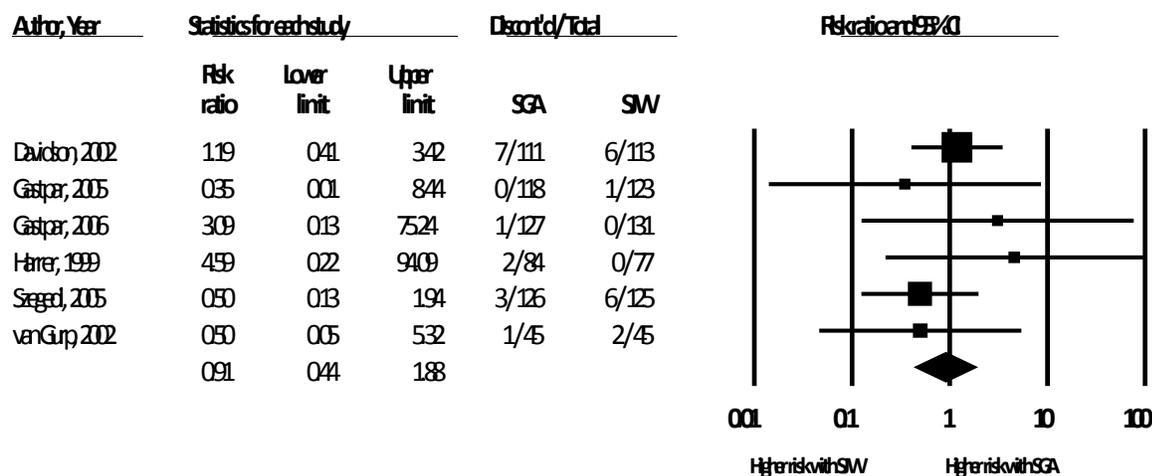
Figure 25. Meta-analysis of discontinuation because of adverse events comparing second-generation antidepressants with St. John's wort



AEs = adverse events; CI = confidence interval; Discont'd = discontinued; SGA = second-generation antidepressant(s); SJW = St. John's wort.

Six trials reported rates of discontinuation because of lack of efficacy.^{97,104,105,119,120,122} Our random-effects meta-analysis found similar rates of discontinuation because of a lack of efficacy between patients treated with SGAs and those treated with St. John's wort (2.3 percent versus 2.4 percent, respectively; RR, 0.91; 95% CI, 0.44 to 1.88; Figure 26). As with the other discontinuation outcomes discussed previously, a fixed effect meta-analytic model did not change these results.

Figure 26. Meta-analysis of discontinuation rates because of lack of efficacy comparing second-generation antidepressants with St. John's wort



Random-effects meta-analysis, I-squared 0%

CI = confidence interval; Discont'd = discontinued; SGA = second-generation antidepressant(s); SJW = St. John's wort.

Second-Generation Antidepressants Versus St. John's Wort: Combination Comparisons

We did not find any trials addressing this comparison.

Second-Generation Antidepressant Switches Compared With Other Second-Generation Antidepressant Switches

In this section, we present the available evidence on the comparative risk of harms from SGA switch strategies compared with other SGA switch strategies following failure of an adequate SGA trial.

Second-Generation Antidepressant Switches Versus Other Second-Generation Antidepressant Switches: Monotherapy Comparisons

A single trial compared the risks of harms from switching to citalopram following failure of a noncitalopram SSRI trial with switching to venlafaxine following failure of a nonvenlafaxine SSRI trial.¹⁰⁹ Overall discontinuation rates were similar regardless of whether treatment was switched to citalopram or venlafaxine (24.5 percent versus 20.9 percent, respectively; p=NR). Also similar between the groups were the overall risks of adverse events (57.5 percent versus 63.1 percent, respectively; p=NR).

Second-Generation Switches Antidepressant Versus Other Second-Generation Antidepressant Switches: Combination Comparisons

We did not find any trials addressing this comparison.

Second-Generation Antidepressants Compared With Exercise

As in previous sections, we first present the available evidence on the comparative risk of harms for SGAs compared with exercise, followed by the available evidence for SGAs compared with combination treatments of SGA and exercise.

Second-Generation Antidepressants Versus Exercise: Monotherapy Comparisons

Two trials comparing sertraline with exercise provided limited data about the comparative risk of harms (see KQ 1, Table 17, for more details on trial design and dosing).^{98,99} Neither trial adequately reported on specific adverse events. One trial reported that, of 36 adverse events that investigators assessed, only the difference in the rates of diarrhea reached statistical significance.⁹⁹ Significantly more patients treated with sertraline experienced diarrhea than those in the home-based and the supervised exercise groups (31 percent versus 21 percent versus 10 percent, respectively; $p=0.03$).⁹⁹ Overall discontinuation rates were similar between patients treated with sertraline and those enrolled in the exercise programs (10 percent versus 14 percent, respectively). Patients on sertraline, however, had statistically significantly higher rates of discontinuation because of adverse events than patients in the exercise programs (6 percent versus 0 percent, respectively; RR, 9.15; 95% CI, 1.09 to 77.06).

Second-Generation Antidepressants Versus Exercise: Combination Comparisons

One of these trials compared sertraline with a combination of sertraline and exercise.⁹⁸ Authors did not report information on specific adverse events. Patients treated with sertraline or a combination with exercise had similar rates of overall discontinuation (15 percent versus 20 percent, respectively) and discontinuation because of adverse events (10 percent versus 9 percent, respectively).⁹⁸

Detailed Synthesis: Risk of Experiencing Serious Adverse Events

Our included trials reported the incidence of serious adverse events even less frequently than more common adverse events. This could reflect the inherent rarity of serious problems, but the majority of our trials also failed to report whether any serious adverse events took place at all, and none indicated how they defined serious adverse events. Overall, 16 trials provided some data on these events.^{79,80,83,88,90,92,95,97,100-105,119,120,127}

Second-Generation Antidepressants Compared With Psychological Treatments

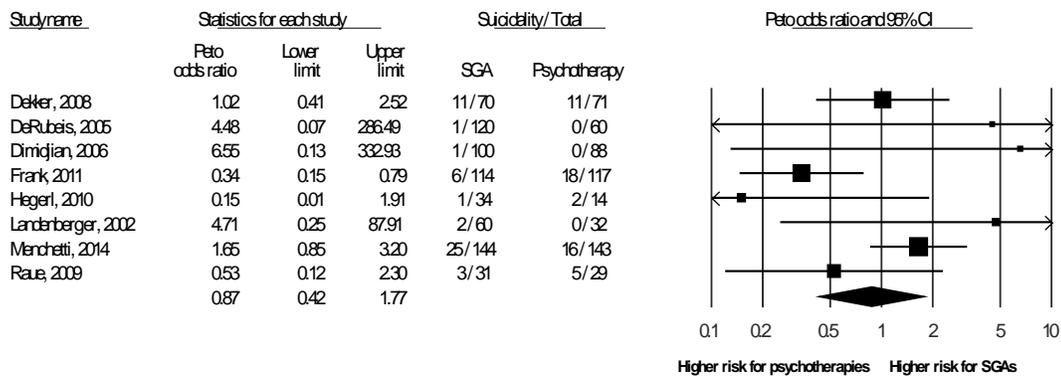
Eleven trials comparing SGA monotherapy with psychotherapy alone or in combination with SGAs provided explicit information about the occurrence or nonoccurrence of serious adverse events.^{79,80,83,86,88,90,92,100-102,116,127} None of these trials compared between-group differences in the rates of serious events.

Second-Generation Antidepressants Compared With Any Psychological Treatment

Nine trials reported the occurrence of one or more serious adverse events;^{80,83,86,88,91,92,100,101,116,127} of these, eight reported data on suicidality (i.e., suicidal ideation, suicide plans, suicide attempts, and/or completed suicides).^{80,83,86,88,91,92,101,116,127} Rates of suicidality ranged from 1 percent to 9 percent for patients treated with SGAs, and from 0 percent to 19 percent for patients receiving psychological treatments. One RCT conducted a comprehensive assessment of suicidality in patients treated with SGAs (escitalopram) or integrative therapy.^{80,127} We received data from the authors of one RCT that evaluated the presence of suicidality at all study timepoints, including baseline,⁸³ as well as data from the authors of three other RCTs that reported the incidence of suicidality at posttreatment followup.^{91,92,101}

We were able to conduct a meta-analysis of overall rates of suicidality using all eight of the above trials that reported suicidality data, four of which had a high risk of bias.^{80,83,86,91,127} These trials all compared patients receiving SGAs (paroxetine or sertraline) with those receiving different psychotherapies (CBT, integrative therapy, psychodynamic therapy, and third-wave CBT). Our analysis did not detect a statistically significant difference in the rate of suicidality events between the two groups, although patients receiving psychotherapy had a numerically higher rate than those receiving SGAs (9.4 percent versus 7.4 percent, respectively; RR, 0.87; 95% CI, 0.42 to 1.77; Figure 27).

Figure 27. Meta-analysis of suicidality data comparing second-generation antidepressants with any psychological treatment



Random effects metaanalysis; I-squared 46%

CI = confidence interval; SGA = second-generation antidepressant(s).

Second-Generation Antidepressants Compared With Behavior Therapy/Behavior Modification

We did not find any trials comparing an SGA with behavior therapy/behavior modification.

Second-Generation Antidepressants Compared With Cognitive Behavioral Therapy

Second-Generation Antidepressants Versus Cognitive Behavioral Therapy: Monotherapy Comparisons

Four trials comparing an SGA (paroxetine or sertraline) with CBT reported data on serious adverse events.^{86,88,91,116} A total of 10 patients experienced serious events, all but two of whom had received SGAs. Three committed suicide,^{86,88,116} one attempted but did not complete suicide,¹¹⁶ one exhibited an unspecified type of suicidality⁹¹, and two experienced severe allergic reactions or severe but unspecified adverse events.¹¹⁶ Both patients who were receiving CBT also exhibited an unspecified type of suicidality.⁹¹ The authors of one trial clearly indicated that they had reported all serious adverse events affecting their patients; therefore, we included this trial alone in our assessment of the SOE for serious adverse events.¹¹⁶

Second-Generation Antidepressants Versus Cognitive Behavioral Therapy: Combination Comparisons

Two trials comparing SGAs with a combination of SGAs and CBT reported data on serious adverse events.^{100,102} In one trial, patients did not experience any serious events, whether they were receiving escitalopram alone or escitalopram in combination with CBT.¹⁰² In the other trial, serious adverse events affected 10 patients.¹⁰⁰ The specific details of these adverse events were left unspecified, except that five were psychiatric in nature: two occurring among patients switching from citalopram to bupropion, sertraline, or venlafaxine after failure of an adequate treatment trial; four among patients augmenting citalopram with CT following a prior treatment failure (two being psychiatric in nature); and four among patients augmenting citalopram with bupropion or buspirone following a prior treatment failure.

Second-Generation Antidepressants Compared With Humanistic Therapies

We found no trials addressing this comparison.

Second-Generation Antidepressants Compared With Integrative Therapies

Second-Generation Antidepressants Versus Integrative Therapies: Monotherapy Comparisons

Three trials compared SGAs (citalopram, escitalopram, or sertraline) and integrative therapies and provided data about serious adverse events.^{80,83,92,127} In one trial, among patients who had no suicidal ideation at baseline but who developed it during the trial, 15.4 percent of patients were receiving integrative therapy and 5.2 percent were receiving SGAs at the onset of their suicidal ideation.^{80,127} No serious adverse events took place in another trial, which compared patients receiving escitalopram with those receiving integrative therapy.⁹² Unpublished data from the authors of the third, high risk of bias trial showed that a numerically greater proportion of patients treated with SGAs no longer endorsed suicidality than did patients treated with integrative therapy.⁸³

Second-Generation Antidepressants Versus Integrative Therapies: Combination Comparisons

We found no trials addressing this comparison.

Second-Generation Antidepressants Compared With Psychodynamic Therapies

Second-Generation Antidepressants Versus Psychodynamic Therapies: Monotherapy Comparisons

Three trials, two of which had a high risk of bias,^{90,101} comparing SGAs with psychodynamic therapies provided information about serious adverse events.^{79,90,101} Patients treated with SGAs (15.7%) and those treated with brief supportive psychodynamic therapy (15.5%) experienced suicidality at similar rates during 8 weeks of followup (p=NR).¹⁰¹ In the other high risk of bias trial, patients receiving fluoxetine (4.4 percent) and those receiving long-term psychodynamic therapy (3.3 percent) experienced similar rates of suicidality at the 96-week followup (p=NR).⁹⁰ In the third trial comparing sertraline or venlafaxine and psychodynamic therapy, no patients experienced serious adverse events.⁷⁹

Second-Generation Antidepressants Versus Psychodynamic Therapies: Combination Comparisons

One high risk of bias trial comparing SGAs with a combination of SGAs and psychodynamic therapy provided information about serious adverse events.⁹⁰ Patients receiving fluoxetine (4.4 percent) and those receiving long-term psychodynamic therapy (1.1 percent) experienced similar rates of suicidality at the 96-week followup (p=NR).⁹⁰

Second Generation Antidepressants Compared With Third-Wave CBT

We found no trials addressing this comparison.

Second-Generation Antidepressants Compared With Complementary and Alternative Medicines

Seven RCTs^{95,97,103-105,119,120} comparing SGA monotherapy with CAM interventions alone or in combination with SGAs provided information about serious adverse events.

Second-Generation Antidepressants Compared With Acupuncture

No information about serious adverse events was available from trials comparing SGAs with acupuncture.

Second-Generation Antidepressants Versus Omega-3 Fatty Acids

Second-Generation Antidepressants Versus Omega-3 Fatty Acids: Monotherapy Comparisons

We did not find any trials addressing this comparison.

Second-Generation Antidepressants Versus Omega-3 Fatty Acids: Combination Comparisons

No information about serious adverse events was available from trials comparing SGAs with omega-3 fatty acids except for one trial (high risk of bias) that compared citalopram with omega-3 fatty acids in combination with citalopram and DHA.⁹⁵ No patients experienced serious adverse events.

Second-Generation Antidepressants Compared With S-Adenosyl-L-Methionine

No information about serious adverse events was available from the sole trial comparing SGAs with SAME.⁹³

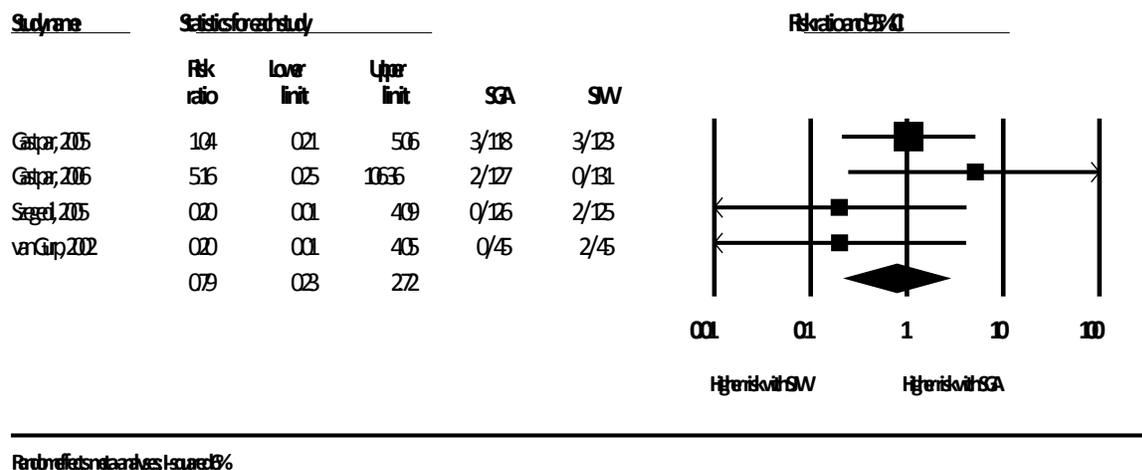
Second Generation Antidepressants Compared With St. John's Wort

Second-Generation Antidepressants Versus St. John's Wort: Monotherapy Comparisons

Seven trials comparing SGAs with various extracts of St. John's wort provided data on serious adverse events.^{97,103-105,119,120,123} A total of 13 patients, 5 receiving SGAs and 8 receiving St. John's wort, experienced serious events. For SGA patients, these included two hospitalizations (both citalopram patients) for serious depression with comorbid general anxiety disorder or for a lesion of the brachial plexus¹¹⁹ and three hospitalizations (all sertraline patients) for a latent suicidal tendency, lower arm fracture, or anaphylactic skin reaction with unknown origin.¹²⁰ Among St. John's wort patients, serious adverse events included single cases of psychic decompensation attributed to social problems (WS5570 extract),¹⁰⁴ hypertensive crisis (WS5570 extract),¹⁰⁴ suicidality (extract not specified),¹⁰⁵ mania,¹⁰⁵ shoulder blade contusion from a fall (STW3 extract),¹²⁰ hospitalization for somatic disorder and psychic decompensation (STW3 extract),¹²⁰ death from cerebral hemorrhage (STW3 extract),¹²⁰ and heroin overdose (LI160 extract).¹⁰³ In two trials, one comparing sertraline with the LI-160 extract of St. John's wort and the other comparing fluoxetine with the same LI-160 extract, no serious adverse events occurred.^{97,123} None of the six trials compared between-group differences in the rates of serious events.

Enough data were available to warrant meta-analyses of overall rates of serious adverse events. Specifically, we included five of the above trials in our analyses (one rated high risk of bias¹⁰³). These trials all compared patients receiving different SGAs (citalopram, fluoxetine, paroxetine, or sertraline) with those receiving St. John's wort.^{103-105,119,120} Our primary analysis with only low and medium risk of bias trials did not detect a statistically significant difference in the rate of serious adverse events between the two groups (1.2 percent versus 1.7 percent, respectively, for SGAs or St. John's wort; RR, 0.79; 95% CI, 0.23 to 2.72; Figure 28). Including the remaining trial (high risk of bias) in the sensitivity analysis did not affect the original findings (SGAs: 1.1 percent versus St. John's wort: 1.7 percent; RR, 0.71; 95% CI, 0.22 to 2.24; forest plot not shown). Because of the low number of events, we urge interpreting these findings with caution.

Figure 28. Meta-analysis of serious adverse events data comparing second-generation antidepressants with St. John’s wort



CI = confidence interval; SGA = second-generation antidepressant(s); SJW = St. John’s wort.

Second-Generation Antidepressants Versus St. John’s Wort: Combination Comparisons

We did not find any trials addressing this comparison.

Second-Generation Antidepressant Switching Strategies

A single trial comparing an SGA switch strategy with a different SGA switch strategy following failure of an adequate SGA trial provided information about serious adverse events.

Second-Generation Antidepressant Switches Versus Other Second-Generation Antidepressant Switches: Monotherapy Comparisons

A single trial comparing the risks of harms from switching to citalopram following failure of a different SSRI trial with switching to venlafaxine following failure of a different SSRI trial reported only that no “unexpected” serious adverse events occurred.¹⁰⁹

Second-Generation Antidepressant Switches Versus Other Second-Generation Antidepressant Switches: Combination Comparisons

We did not find any trials addressing this comparison.

Second-Generation Antidepressants Compared With Exercise

No information about serious adverse events was available from trials comparing SGAs with exercise.

KQ 3b: Variation in Risk of Harms by Severity of Major Depressive Disorder

Detailed Synthesis: Overall Risk of Experiencing Harms and Discontinuation of Treatment

A single trial (high risk of bias) comparing SGAs with CBT and third-wave CBT provided qualitative information about baseline MDD severity as a moderator of the risk of adverse events.⁸⁶ Specifically, the risk of adverse events in patients treated with SGAs did not differ by baseline severity except in two cases: higher-severity patients experienced more nausea but less diarrhea than lower-severity patients. Because of the methodological limitations and the small sample size of this trial, we urge interpreting these results with caution.

Detailed Synthesis: Risk of Experiencing Serious Adverse Events

We did not find any trials addressing the potential role of baseline MDD severity as a moderator of risk of experiencing serious adverse events.

KQ 4: Comparative Benefits and Risks of Harms for Selected Subgroups

Overview

In this section, we focus on the comparative benefits and harms of SGAs with psychotherapy, CAM, or exercise for treating MDD in selected subpopulations. Specific subgroups were defined by common accompanying psychiatric symptoms (coexisting anxiety, insomnia, low energy, or somatization) or by demographic characteristics (age, sex, or race or ethnicity).

As we have done in previous sections, here we provide an overview of the articles, including the number of trials for each comparison (listed in Table 28); key points; and a detailed synthesis. In Appendix E, we present “summary of findings” tables for a set of outcomes identified as especially important. These tables describe basic information on the available evidence and present the SOE grades for each outcome.

Table 28. Number of included trials for all subgroups by type of comparison

Comparison Category	Comparisons	Number of Trials
SGA vs. Psychological Interventions	SGA vs. Behavior therapies/behavior modification	0
	SGA vs. CBT	1 ⁸⁹
	SGA vs. Humanistic therapies	0
	SGA vs. Integrative therapies	1 ⁹²
	SGA vs. Psychodynamic therapies	0
	SGA vs. Third-wave CBTs	0
SGA vs. Complementary and Alternative Medicine	SGA vs. Acupuncture	0
	SGA vs. Omega-3 fatty acids	0
	SGA vs. SAMe	0
	SGA vs. St. John’s wort	1 ¹²²
	SGA vs. Meditation	0
	SGA vs. Yoga	0
SGA vs. Exercise	SGA vs. Exercise	0

CBT = cognitive behavioral therapy; SAMe = S-adenosyl-L-methionine; SGA = second-generation antidepressant.

No trials were specifically designed to assess differences in our specified subgroups. Overall, as documented in Table 28, only three trials addressing a subgroup of interest met the criteria for inclusion. We did not have sufficient data on any subgroup to conduct mixed treatment comparisons.

No trials at all addressed efficacy or harms in selected subgroups of patients who did not achieve remission following an initial adequate trial with one SGA.

Key Points: Common Accompanying Psychiatric Symptoms

- We did not find any evidence comparing SGAs with any other nonpharmacologic interventions in subgroups with comorbid anxiety (insufficient SOE).
- We did not identify any eligible trials for subgroups with accompanying insomnia, low energy, or somatization (insufficient SOE).

Key Points: Age

- St. John's wort did not lead to statistically different rates of rates compared with SGAs after 6 weeks of treatment in older adults with MDD (one RCT, low SOE for no differences); both groups developed adverse drug reactions, and discontinuation rates attributable to adverse events were similar (low SOE for no differences).
- We did not find any eligible evidence comparing SGAs with other CAM interventions by age (i.e., acupuncture, meditation, omega-3 fatty acids, SAME, or yoga) (all insufficient SOE).
- We did not find any eligible evidence comparing SGAs with psychological interventions by age (insufficient SOE).

Key Points: Sex

- We did not identify any trials assessing differences between men and women in efficacy or harms (insufficient SOE).
- SGAs and CBT showed similar reduction in depressive symptoms in a trial that included only minority women.

Key Points: Race or Ethnicity

- No trials directly compared the efficacy, effectiveness, or harms of SGAs with eligible psychotherapy, CAM, or exercise interventions among patients of different races or ethnicities (insufficient SOE).

Detailed Synthesis: Common Accompanying Psychiatric Symptoms

Second-Generation Antidepressant Compared With Psychotherapy Interventions

One trial comparing SGAs with IPT assessed differences in patients with and without comorbid anxiety disorders.⁹² The trial was conducted in primary care settings in New Zealand. The SGA produced higher remission rates than IPT in patients with a comorbid anxiety disorder;

it did not have that effect in patients without co-occurring anxiety. No evidence on risk of harms was reported.

We found no eligible trials in subgroups of MDD patients with accompanying insomnia, low energy, or somatization.

Table 29 provides detailed information on included trials for all subgroups.

Table 29. Second-generation antidepressants versus nonpharmacologic therapies in subgroups: Trial characteristics, main outcomes, and risk of bias ratings

Trial	N	Mean Baseline HAM-D Score	SGA: mg/day Comparator: mg/day or Number of Sessions	Remission ^a and Significance Level	Response ^a and Significance Level	Risks of Harms	Risk of Bias Rating
Menchetti et al., 2014 ⁹²	287	17.3	Citalopram: 10 to 60 or Sertraline: 25 to 200	Comorbid anxiety disorder: 70 vs. 65 SRD= -0.05; 95% CI, -0.33 to 0.23	NR	NR	Medium
Accompanying 8 psychiatric symptoms (anxiety)			Interpersonal psychotherapy: 6 to 8	No comorbid anxiety disorder: 46 vs. 67 SRD=0.21; 95% CI, 0.04 to 0.38			
Women Entering Care (WECare), 2003 ⁸⁹	178	16.9 ^c	Paroxetine: 10 to 50	NR	NR	NR	Medium
Minority women	8 ^b		Cognitive behavioral therapy: 8				
Harrer et al., 1999 ¹²²	149	NR	Fluoxetine: 10 to 400	NR	72 vs. 71 p=NR	Discontinued treatment because of adverse drug reactions: 8 vs. 6	Medium
Older adults	6						

^a Response and remission are measured on the HAM-D.

^b Results reported at 4 weeks.

^c Mean baseline score includes participants randomized to community referral intervention.

CI = confidence interval; HAM-D = Hamilton Depression Rating Scale; N = number; NR = not reported; SGA = second-generation antidepressant; SRD = standardized rate difference; vs. = versus.

Second-Generation Antidepressant Compared With Complementary and Alternative Medicine Interventions

We found no eligible trials in subgroups with accompanying psychiatric symptoms.

Second-Generation Antidepressant Compared With Exercise Interventions

We found no eligible trials in subgroups with accompanying psychiatric symptoms.

Detailed Synthesis: Age

No trials directly compared the efficacy, effectiveness, or harms of SGAs with eligible psychotherapy, CAM, or exercise interventions in older adults (55 years of age or older) and the general population. We identified one trial that exclusively enrolled older adults; it assessed response, remission, and harms for SGAs compared with St. John's wort. We did not find any evidence about other outcomes of interest such as quality of life or functional capacity.

Second-Generation Antidepressant Compared With Complementary and Alternative Medicine Interventions

One trial conducted in a primary care setting randomized older adults (60 to 80 years of age) to fluoxetine or St. John's wort for 6 weeks. Both treatments produced similar response rates and reductions in HAM-D scores.¹²² In addition, discontinuation because of harms was similar for both groups.

Second-Generation Antidepressant Compared With Exercise Interventions

No trials meeting our eligibility criteria compared SGAs with exercise. We identified post-hoc analysis from a trial in adults 55 years or older. Even though this analysis does not meet criteria for inclusion, we briefly describe it here because of the paucity of evidence on subgroups. This analysis found no significant difference between sertraline and exercise in neurocognitive function in older adults.¹³¹

Detailed Synthesis: Sex

Second-Generation Antidepressant Compared With Psychotherapy Interventions

We did not identify any trials assessing differences between men and women in efficacy or harms. One trial (described in KQ 1) randomized low-income minority women to SSRI or CBT for 8 weeks.⁸⁹ Both interventions improved patients' depressive symptoms. At month 6, SSRI-treated participants reported lower depressive symptoms and better instrumental role functioning than those treated with CBT.

Second-Generation Antidepressant Compared With Complementary and Alternative Medicine Interventions

We found no eligible evidence.

Second-Generation Antidepressant Compared With Exercise Interventions

We found no eligible evidence.

Detailed Synthesis: Race or Ethnicity

We did not identify any trials assessing benefits or harms of second-generation antidepressants with eligible psychotherapy, CAM, or exercise interventions across races or ethnic groups.

Discussion

This chapter summarizes the key findings and how they relate to published findings and current clinical practices and policies. We also briefly examine the applicability of our findings and their implications for decisionmaking. We comment on limitations of both the review process and the entire evidence base as a segue into our discussion of research gaps in this field.

Key Findings and Strength of Evidence

Pharmacotherapy (particularly second-generation antidepressants [SGAs]) is the primary intervention for treating patients with major depressive disorder (MDD) in primary care. Nonetheless, primary care patients and clinicians may prefer other options (or at least want to be able to consider them). These include psychotherapeutic interventions, complementary and alternative medicine (CAM) options, exercise, or a combination of these treatments. Our report provides a comprehensive summary of the available evidence on comparative effectiveness and risk of harms of commonly used pharmacological and nonpharmacological treatments for MDD.

In this review we focus on two key issues that primary care physicians commonly face:

1. How do different treatment options compare as an initial treatment choice, and how effective are SGAs compared with nonpharmacologic interventions?
2. For patients whose depression did not achieve remission following initial treatment with an SGA, what is the comparative effectiveness of alternative pharmacologic and nonpharmacologic options?

Overall, the available evidence was characterized by substantial methodological shortcomings and a lack of adequate assessment of harms. In addition, outcomes reporting bias often appeared to be an issue. For example, publications frequently did not report remission and adverse events, yet trials on treatment of patients with MDD are unlikely to fail to assess these outcomes.

The available evidence base has some clear limitations. Some nonpharmacological interventions have never been compared with any SGAs. Very limited evidence is available to address the comparative effectiveness of second-line therapies (i.e., treatment options for patients who did not achieve remission after an initial treatment trial). Further, the role of depression severity as a moderator of treatment effectiveness, whether for first- or second-line therapies, has received very little direct testing.

Nevertheless, we were able to draw some conclusions. Because reliable evidence supports similar effectiveness within the class of SGAs, our conclusions are likely valid for the entire class of SGAs.

Comparative Benefits and Harms of Treatment Options for Initial Treatment of Patients With Major Depressive Disorder

Across all interventions, we graded the strength of evidence as moderate for some outcomes of only two comparisons, namely SGAs compared with cognitive behavioral therapy (CBT) and St. John's wort. Results from trials of these comparisons indicate that CBT and St. John's wort have levels of effectiveness regarding symptomatic relief similar to those of SGAs. The overall risk for adverse events or discontinuation of treatment because of adverse events, however, is lower for these non-SGA therapies.

Our confidence in findings from the remaining comparisons of SGAs with other treatment options was low or insufficient, indicating that these bodies of evidence had major or unacceptable deficiencies. Nevertheless, for most comparisons the overall findings did not detect a statistically significant difference in effectiveness but did indicate a lower risk of adverse events for nonpharmacological treatment options. Notable exceptions are omega-3-fatty acids, which appear to have lower effectiveness than SGAs; and the combination of SGAs with acupuncture which appears to have greater effectiveness than SGA monotherapy. Our confidence in these findings, however, is low and results have to be interpreted cautiously. In addition, for many comparisons that are limited to single trials, determining whether similar treatment effects between SGAs and other interventions are based on similar effectiveness or high placebo response rates is impossible.

Table 30 summarizes our main findings and the respective certainty that we have about these findings, presented as SOE grades (high, moderate, low, or insufficient).⁷⁶ In this table, we do not present comparisons for which we found no studies whatsoever or for which we were unable to estimate the comparative effectiveness with network meta-analyses. We discuss the summary of findings in more detail below.

Table 30. Summary of findings with strength of evidence: Comparative benefits and harms of second-generation antidepressants and other treatment options as an initial choice for the treatment of patients with major depressive disorders

Comparison and Outcome of Interest	Strength of Evidence ^a	Findings
SGA vs. CBT monotherapy		
Remission	Low	Results from direct comparisons in 3 trials indicate that no substantial differences in remission exist between SGAs and CBT monotherapy.
Response	Moderate	Results from direct comparisons in 3 trials indicate that no substantial differences in response exist between SGAs and CBT monotherapy.
Functional capacity	Low	Results from 1 trial indicate that no substantial differences in functional capacity exist between SGAs and CBT monotherapy.
Overall risk of adverse events	Insufficient	Based on 1 trial with few events, the evidence is insufficient to draw conclusions.
Overall discontinuation of treatment	Moderate	Results from direct comparisons in 4 trials indicate that no significant differences exist in overall discontinuation between patients treated with SGAs and those treated with CBT.
Discontinuation of treatment because of adverse events	Low	Results from direct comparisons in 3 trials indicate that patients treated with SGAs experience a significantly higher rate of discontinuation because of adverse events than those treated with CBT.
Serious adverse events	Insufficient	Based on 2 trials with very few events, the evidence is insufficient to draw conclusions.
Suicidality	Insufficient	Based on 4 trials with very few events, the evidence is insufficient to draw conclusions.
SGA vs. SGA + CBT		
Remission	Low	Results from direct comparisons in 2 trials indicate that no substantial differences in remission exist between SGAs and SGAs combined with CBT.
Response	Low	Results from direct comparisons in 2 trials indicate that no substantial differences in response exist between SGAs and SGAs combined with CBT.
Functional capacity	Low	Results from 1 trial indicate that the combination of SGA with CBT results in greater improvement on 3 of 4 work functioning measures than SGA alone.
Overall discontinuation of treatment	Low	Results from direct comparisons in 1 head-to-head trial indicate that no significant differences exist in overall discontinuation between patients treated with SGAs and those treated with CBT.

Table 30. Summary of findings with strength of evidence: Comparative benefits and harms of second-generation antidepressants and other treatment options as an initial choice for the treatment of patients with major depressive disorders (continued)

Comparison and Outcome of Interest	Strength of Evidence^a	Findings
Discontinuation of treatment because of adverse events	Low	Results from direct comparisons in 1 head-to-head trial indicate that no significant differences exist in discontinuation because of adverse events between patients treated with SGAs and those treated with CBT.
SGA vs. IT monotherapy		
Remission	Low	Results from direct comparisons in 2 trials indicate that no substantial differences in remission exist between SGAs and interpersonal therapy monotherapy.
Response	Low	Results from 1 trial indicate that no substantial differences in response exist between SGAs and interpersonal therapy monotherapy.
Overall discontinuation of treatment	Insufficient	Based on 2 trials with very few events, the evidence is insufficient to draw conclusions.
Discontinuation of treatment because of adverse events	Insufficient	Based on 1 trial with few events, the evidence is insufficient to draw conclusions.
Suicidality	Insufficient	Based on 2 trials with very few events, the evidence is insufficient to draw conclusions.
SGA vs. SGA + IT		
Remission	Low	Results from 1 trial indicate that a substantial difference in remission favoring SGAs combined with interpersonal therapy exists, but the confidence interval is very wide.
Overall discontinuation of treatment	Insufficient	Based on 1 with very few events, the evidence is insufficient to draw conclusions.
Subgroup with anxiety	Insufficient	Based on 1 trial, the evidence is insufficient to draw conclusions for benefits or harms.
SGA vs. PSYD monotherapy		
Remission	Low	Results from 1 trial indicate that no substantial differences in remission exist between SGAs and PSYD monotherapy.
Functional capacity	Low	Results from direct comparisons based on 2 trials indicate that few substantial differences in functional capacity exist between SGAs and PSYD monotherapy.
Overall discontinuation of treatment (16 weeks)	Insufficient	Based on 2 trials with few events, the evidence is insufficient to draw conclusions.
Overall discontinuation of treatment (48 weeks)	Low	Results from direct comparisons in 1 head-to-head trial indicate that no significant differences exist in overall discontinuation after 48 weeks of followup between patients treated with SGAs and those treated with PSYD monotherapy.
Overall discontinuation of treatment (96 weeks)	Low	Results from direct comparisons in 1 head-to-head trial indicate that no significant differences exist in overall discontinuation after 96 weeks of followup between patients treated with SGAs and those treated with PSYD monotherapy.
Suicidality	Insufficient	Based on 1 trial with very few events, the evidence is insufficient to draw conclusions.
SGA vs. SGA + PSYD		
Overall discontinuation of treatment	Low	Results from direct comparisons in 1 head-to-head trial indicate that patients treated with SGAs experience a significantly higher rate of overall discontinuation than those treated with SGAs plus PSYD.
Suicidality	Low	Results from direct comparisons based on a single head-to-head trial indicate that no significant differences exist in suicidality between patients treated with SGAs and those treated with SGAs plus PSYD.

Table 30. Summary of findings with strength of evidence: Comparative benefits and harms of second-generation antidepressants and other treatment options as an initial choice for the treatment of patients with major depressive disorders (continued)

Comparison and Outcome of Interest	Strength of Evidence^a	Findings
SGA vs. Third wave CBT		
Remission	Insufficient	Results from 1 trial indicate that a substantial difference in remission favoring third wave CBT exists, but the sample size is very small.
Response	Insufficient	Results from 1 trial indicate that a substantial difference in response favoring third wave CBT exists, but the sample size is very small.
Overall discontinuation of treatment	Insufficient	Based on 2 trials with few events, the evidence is insufficient to draw conclusions.
Discontinuation of treatment because of adverse events	Insufficient	Based on 2 trials with few events, the evidence is insufficient to draw conclusions.
Suicidality	Insufficient	Based on 1 trial with very few events, the evidence is insufficient to draw conclusions.
SGA vs. Acupuncture monotherapy		
Response	Low	Results from direct comparisons based on 2 head-to-head trials, as well as network meta-analysis, indicate that no substantial differences in response exist between patients treated with SGA and those treated with acupuncture monotherapy.
Overall risk of adverse events: direct evidence	Insufficient	Based on 1 trial with few events, the evidence is insufficient to draw conclusions.
Overall risk of adverse events: indirect evidence	Moderate	Results from a single systematic review of 21 trials indicate that patients treated with SGAs experience a significantly higher overall risk of adverse events than those treated with acupuncture. However, this systematic review of 21 trials did not meet our eligibility criteria because some trials included depressive disorders other than MDD.
Overall discontinuation of treatment	Low	Results based on 1 head-to-head trial indicate that patients treated with SGAs experience significantly lower rates of overall discontinuation than those treated with acupuncture.
SGA vs. SGA + Acupuncture		
Remission	Low	Results from direct comparisons in 1 head-to-head trial indicate that no substantial differences in remission exist between patients treated with SGAs and those treated with acupuncture combination therapy.
Response	Low	Results from direct comparisons in 2 head-to-head trials indicate higher response rates for patients treated with SGAs plus acupuncture than patients treated with SGAs alone.
Overall risk of adverse events	Low	Results from direct comparisons based on 1 head-to-head trial indicate that no significant differences exist in overall risk of adverse events between patients treated with SGAs and those treated with acupuncture plus SGAs.
Overall discontinuation of treatment	Moderate	Results from direct comparisons in 3 head-to-head trials indicate that no significant differences exist in overall discontinuation between patients treated with SGAs and those treated with SGAs plus acupuncture.
Discontinuation of treatment because of adverse events	Low	Results from direct comparisons based on 2 head-to-head trials indicate that no significant differences exist in discontinuation because of adverse events between patients treated with SGAs and those treated with SGAs plus acupuncture.

Table 30. Summary of findings with strength of evidence: Comparative benefits and harms of second-generation antidepressants and other treatment options as an initial choice for the treatment of patients with major depressive disorders (continued)

Comparison and Outcome of Interest	Strength of Evidence ^a	Findings
SGA vs. Omega-3 fatty acids monotherapy		
Response	Low	Results from network meta-analysis indicate higher response rates for patients treated with SSRIs than for those receiving omega-3 fatty acids.
Overall discontinuation of treatment	Low	Results from direct comparisons in 1 head-to-head trial indicate that no substantial differences exist in overall discontinuation between patients treated with SGAs and those treated with omega-3 fatty acids.
Discontinuation of treatment because of adverse events	Low	Results from direct comparisons in 1 head-to-head trial indicate that no substantial differences exist in overall discontinuation between patients treated with SGAs and those treated with omega-3 fatty acids.
Suicidality	Insufficient	Based on 1 trial with few events, the evidence is insufficient to draw conclusions.
SGA vs. SGAs + Omega-3 fatty acids		
Remission	Insufficient	Based on 1 trial, the evidence is insufficient to draw conclusions.
Response	Insufficient	Based on 2 trials, the evidence is insufficient to draw conclusions.
Overall discontinuation of treatment	Low	Results from direct comparisons in 2 head-to-head trials indicate that no substantial differences in overall discontinuation between patients treated with SGAs and those with treated with SGAs plus omega-3 fatty acids.
Discontinuation of treatment because of adverse events	Low	Results from direct comparisons in 1 head-to-head trial indicate that no substantial differences in overall discontinuation between patients treated with SGAs and those with treated with SGAs plus omega-3 fatty acids.
SGAs vs. SAME monotherapy		
Remission	Insufficient	Based on 1 trial, the evidence is insufficient to draw conclusions.
Response	Low	Results from direct comparisons in 1 trial and our network meta-analysis indicate that no substantial differences in response exist between SGA and SAME monotherapy.
Overall discontinuation of treatment	Low	Results from direct comparisons in 1 head-to-head trial indicate that no significant differences exist in overall discontinuation between patients treated with SGAs and those treated with SAME.
Discontinuation of treatment because of adverse events	Insufficient	Based on 1 trial with few events, the evidence is insufficient to draw conclusions.
SGA vs. St. John's wort monotherapy		
Remission	Moderate	Results from direct comparisons based on 4 head-to-head trials indicate that no substantial differences in remission exist between patients treated with SGA and those treated with St. John's wort monotherapy.
Response	Moderate	Results from direct comparisons in 8 head-to-head trials indicate that no apparent differences in response exist between patients treated with SGAs and those treated with St. John's wort monotherapy.
Overall risk of adverse events	Moderate	Results from direct comparisons in 8 head-to-head trials indicate that patients treated with SGAs experience a significantly higher overall risk of adverse events than those treated with St. John's wort.
Overall discontinuation of treatment	Moderate	Results from direct comparisons in 12 head-to-head trials indicate that patients treated with SGAs experience significantly higher rates of overall discontinuation than those treated with St. John's wort.

Table 30. Summary of findings with strength of evidence: Comparative benefits and harms of second-generation antidepressants and other treatment options as an initial choice for the treatment of patients with major depressive disorders (continued)

Comparison and Outcome of Interest	Strength of Evidence ^a	Findings
Discontinuation of treatment because of adverse events	Moderate	Results from direct comparisons in 11 head-to-head trials indicate that patients treated with SGAs experience significantly higher rates of discontinuation because of adverse events than those treated with St. John's wort.
Serious adverse events	Low	Results from direct comparisons in 5 head-to-head trials indicate that no significant differences exist in the occurrence of serious adverse events between patients treated with SGAs and those treated with St. John's wort.
Suicidality	Insufficient	Based on 3 trials with few events, the evidence is insufficient to draw conclusions.
Subgroup based on older age	Low	Results from 1 trial in older adults indicate similar response rates and discontinuation rates because of adverse events for patients treated with SGAs and those treated with St. John's wort.
SGA vs. Exercise monotherapy		
Remission	Low	Results based on direct comparisons in 2 trials reveal no significant difference in remission between patients treated with SGAs and those treated with exercise therapy.
Response	Low	Estimates based on network meta-analysis reveal no significant difference in response between patients treated with SGAs and those treated with exercise therapy.
Overall discontinuation of treatment	Low	Results from direct comparisons in 1 head-to-head trial indicate that no significant differences exist in overall discontinuation between patients treated with SGAs and those treated with exercise.
Discontinuation of treatment because of adverse events	Low	Results from direct comparisons in 1 head-to-head trial indicate that patients treated with SGAs experience significantly higher rates of discontinuation because of adverse events than those treated with exercise.
SGA vs. Exercise + SGA		
Remission	Low	Results based on direct comparison from 1 trial reveal no significant difference in effectiveness between SGA and SGAs plus exercise.
Overall discontinuation of treatment	Low	Results from direct comparisons based on a single head-to-head trial indicate that no significant differences exist in overall discontinuation between patients treated with SGAs and those treated with SGAs plus exercise.
Discontinuation of treatment because of adverse events	Low	Results from direct comparisons based on a single head-to-head trial indicate that no significant differences exist in discontinuation because of adverse events between patients treated with SGAs and those treated with SGAs plus exercise.

^aStrength of evidence grades (high, moderate, low, or insufficient) are based on methods guidance for the AHRQ EPC program.⁷⁶

CBT = cognitive behavioral therapy; IT = integrative therapies; MMD = major depressive disorder; PSYD = psychodynamic therapies; SGA = second-generation antidepressant; SAME = S-adenosyl-L-methionine; SSRI = selective serotonin reuptake inhibitor; third wave CBT = third wave cognitive behavioral therapy; vs. = versus.

For psychotherapies, the available evidence based on 20 randomized controlled trials (RCTs) with 3,000 patients suggests similar beneficial treatment effects of SGAs and psychotherapies, either alone or in combination. Except for SGAs compared with CBT, however, the strength of evidence was low or insufficient, indicating a strong uncertainty associated with these findings.

With respect to risk of harms SGAs often had higher rates of adverse events or discontinuation rates because of adverse events than psychotherapies. For most of these comparisons, the SOE was also low or insufficient. For example, the evidence was insufficient to draw any conclusions about the comparative risk for serious adverse events. Reasons for low or

insufficient SOE grades reflected mainly levels of risk of bias for individual trials and lack of precision of results that encompassed substantial benefits for both interventions.

Many trials had methodological shortcomings such as high dropout rates or lack of blinding of outcome assessors that reduced our confidence in the results. In addition, few trials adequately determined or reported differences in harms. Some comparisons were based on single trials with small sample sizes, which led to indeterminate results because of wide confidence intervals that encompassed appreciable benefits for both comparators. The best available evidence for psychological interventions with moderate SOE was SGAs compared with CBT monotherapy. We found no statistically significant difference in treatment effects on response or remission in our analysis of trials that we rated as low or medium risk of bias trials, although a sensitivity analysis of remission that included three trials that we rated high risk of bias yielded a result that favored SGAs.

For the comparison of SGAs with CAM interventions, we identified 20 RCTs including 2,649 patients comparing an SGA with one of six CAM therapies for treating patients with MDD. Individual trials faced the same methodological issues as trials for psychological interventions. We rated nearly half of them as high risk of bias (nine trials). Few trials adequately assessed and reported the risk of harms. Because of the lack of evidence and the methodological limitations of many head-to-head trials, we relied on both direct evidence and network meta-analyses to draw conclusions. With the exception of omega-3-fatty acids, beneficial effects appeared to be similar between SGAs and CAM interventions; however, results for comparisons of SGAs with acupuncture and S-adenosyl-L-methionine (SAME) are limited to low SOE, indicating substantial uncertainty of findings. Network meta-analyses resulted in higher response rates for SGAs than omega-3-fatty acids.

Moderate SOE based on 12 trials found similar effectiveness for SGAs and St. John's wort; however, St. John's wort had lower risks of adverse events and discontinuation rates than SGAs. The evidence was insufficient to determine differences in risk of serious adverse events for any comparisons.

Based on two RCTs with low SOE, we found that the beneficial treatment effects of SGAs and exercise, either alone or in combination, were not significantly different. In one trial, patients in the exercise groups reported a slightly lower risk of side effects (diarrhea) than those treated with SGAs.

We did not find any trials on behavior therapy and behavior modification, meditation, and yoga that met our eligibility criteria.

Comparative Benefits and Harms as a Function of Baseline Depressive Severity

The evidence was insufficient to draw any firm conclusions about differences in benefits and harms among interventions of interest as a function of depressive severity. Table 31 summarizes our findings and the respective certainty that we have about these findings, presented as SOE grades (high, moderate, low, or insufficient).⁷⁶

Table 31. Summary of findings with strength of evidence: Variation in effectiveness by severity for second-generation antidepressants compared with other treatments for patients with major depressive disorder

Comparison and Outcome of Interest	Strength of Evidence ^a	Findings
SGA vs. CBT monotherapy		
Remission	Insufficient	Based on 1 trial, the evidence is insufficient to draw conclusions.
Response	Insufficient	Based on 1 trial, the evidence is insufficient to draw conclusions.
SGA vs. IT monotherapy		
Remission	Insufficient	Based on 1 trial, the evidence is insufficient to draw conclusions.
SGA vs. Third wave CBT		
Remission	Low	Results from 1 trial with a small sample size indicate that patients with high-severity MDD treated with behavioral activation experience a significantly higher rate of remission than those treated with SGAs, but results did not indicate a difference in remission for patients with low-severity MDD.
Response	Low	Results from 1 trial with a small sample size indicate that baseline severity exerts no significant difference on response between SGA and behavioral activation.
SGAs vs. SAmE		
Remission	Insufficient	Based on 1 trial, the evidence is insufficient to draw conclusions.
Response	Insufficient	Based on 1 trial, the evidence is insufficient to draw conclusions.

^aStrength of evidence grades (high, moderate, low, or insufficient) are based on methods guidance for the AHRQ EPC program.⁷⁶

AHRQ = Agency for Healthcare Research and Quality; CBT = cognitive behavioral therapy; EPC = Evidence-based Practice Center; IT = integrative therapies; MDD = major depressive disorder; SAmE = S-adenosyl-L-methionine; SGA = second-generation antidepressant; third wave CBT = third wave cognitive behavioral therapy.

Comparative Benefits and Harms of Alternative Pharmacologic and Nonpharmacologic Options for Patients Whose Depression Did Not Achieve Remission Following Initial Treatment With a Second-Generation Antidepressant

Table 32 summarizes our findings and the respective certainty that we have about these findings, presented as SOE grades (high, moderate, low, or insufficient).⁷⁶ Comparisons only involved medications and CT; no eligible trials involving CAM or exercise interventions were identified.

Two trials involved 1,992 patients and provided data for four comparisons. All findings suggested little difference in benefit for depression regardless of whether a switch or augmentation strategy was used or whether medications or cognitive therapy were involved. Both trials suffered from attrition rates over 20 percent, and all comparisons other than SGA switch compared with SGA switch were based on data from one study. For all the comparisons except one, the SOE was low, indicating limited confidence that the estimate of effect lies close to the true effect for these outcomes.

Table 32. Summary of findings with strength of evidence: Comparative benefits of second-generation antidepressants and other treatment options as an initial choice for the treatment of major depressive disorder (KQ 2a)

Comparison and Outcome of Interest	Strength of Evidence ^a	Findings
Switch Strategies: SGA switch vs. SGA switch		
Response	Moderate	Results from 2 direct comparisons involving 1,123 patients indicate no substantial differences in response rates between SGAs.
Remission	Low	Results from 1 direct comparison involving 727 patients indicate no substantial difference in remission rates between SGAs.
Decrease in depressive severity	Low	Results from 1 direct comparison involving 727 patients indicate no substantial differences in decrease in depressive severity between SGAs.
Switch Strategies: SGA switch vs. CT switch		
Response, remission, and change in depressive severity	Low	Results from 1 direct comparison of switching to a different SGA vs. switching to CT involving 122 patients indicate no substantial differences in rates of response or remission or in the decrease in depressive severity.
Augmentation Strategies: SGA augment vs. SGA augment		
Response and remission	Low	Results from 1 direct comparison involving 565 patients indicate no substantial differences in rates of response or remission between SGAs.
Decrease in depressive severity	Low	Results from 1 direct comparison involving 565 patients indicate a greater decrease in depressive severity after adding bupropion than buspirone.
Augmentation Strategies: SGA augment vs. CT augment		
Response, remission, and change in depressive severity	Low	Results from 1 direct comparison involving 182 patients of augmenting with a second medication vs. augmenting with CT indicate no substantial differences in rates of response or remission or in the decrease in depressive severity.

^aStrength of evidence grades (high, moderate, low, or insufficient) are based on methods guidance for the AHRQ EPC program.¹⁶

CT = cognitive therapy; KQ = Key Question; SGA = second-generation antidepressant; vs. = versus.

Comparative Benefits and Harms of Second-line Therapies as a Function of Baseline Depressive Severity

The evidence was insufficient to draw any conclusions about differences in benefits and harms among second-line interventions of interest as a function of depressive severity. Table 33 summarizes our findings.

Table 33. Summary of findings with strength of evidence: Second-generation antidepressants compared with other treatments for major depressive disorder: Does effectiveness vary by severity? (KQ 2b)

Comparison and Outcome of Interest	Strength of Evidence ^a	Findings
Switch strategies		
SGA switch vs. SGA switch		
Remission	Insufficient	One industry-supported secondary analysis found an insignificant trend toward difference in remission rates for those with severe depression, while a second government-funded secondary analysis found that having mild/moderate vs. severe depression did not modify responses to different SGAs.

^aStrength of evidence grades (high, moderate, low, or insufficient) are based on methods guidance for the AHRQ EPC program; outcomes for which we have no studies are designated no evidence.

KQ = Key Question; SGA = second-generation antidepressant; vs. = versus.

Findings in Relationship to What is Already Known

Our findings are consistent with several prior systematic reviews and meta-analyses that compared SGAs with nonpharmacological intervention. Most of these reviews, however, included populations that were not eligible for our review, such as patients with minor depression, bipolar disorder, or dysthymia.

For psychological treatments, one meta-analysis found that serotonin-specific reuptake inhibitors (SSRIs) were more effective than psychotherapy in treating patients with depressive disorders; however, this effect was small and potentially clinically insignificant.¹³⁸ Another meta-analysis found that SGAs and psychotherapy have equivalent efficacy in the short term after 6 to 26 weeks of treatment.⁴⁰ Our finding that SGA monotherapy, CBT, interpersonal therapy, and psychodynamic therapy may all have equivalent effects in the short-term treatment of depressed patients is consistent with those results.

Our results are also consistent with the recommendations of both the American Psychiatric Association¹⁹ and the U.S. Department of Veterans Affairs/Department of Defense.¹³⁹ These two groups consider both pharmacotherapy and psychotherapy to be appropriate individual first-line treatments for mild to moderate MDD. Furthermore, they state that pharmacotherapy plus psychotherapy may be a useful initial treatment for patients with moderate to severe MDD and for those with MDD and comorbid conditions.

Several reviews have been done of CAM therapies for treating MDD patients; these include an APA Task Force Report, Clinical Guidelines from the Canadian Psychiatric Association, and a systematic review from the U.S. Department of Veterans Affairs.^{48,49,140} Additionally, many reviews of individual CAM therapies have been published for the treatment of MDD,^{53,137,141,142} including reviews by the Cochrane Collaboration.^{46,143,144}

Although one systematic review of acupuncture concluded that it had efficacy comparable with that for antidepressant medications,¹³⁷ a Cochrane review⁴⁶ and reviews from the American Psychiatric Association, Canadian Network for Mood and Anxiety Treatments, and Department of Veteran Affairs agree evidence is insufficient to recommend acupuncture as monotherapy or combination therapy for treating MDD patients. Some reports recognize that risk of harms for acupuncture may be low. Nevertheless, most reports note that current trials often have high risk of bias. Similarly, we found few high-quality trials to support the use of acupuncture for MDD. Nevertheless, we found that a few RCTs, in addition to network meta-analysis, may indicate (a) similar effectiveness for acupuncture monotherapy compared with SGA and (b) better treatment

response for a combination of acupuncture with SGA compared with only SGA. However, we concluded that the SOE for these associations was low due to the relative paucity of trials and high risk of bias among those trials we identified.

Both the U.S. and Canadian reviews recommend omega-3 fatty acids as augmentation for treating patients with mild to moderate MDD, noting modest evidence of efficacy and low risks of harm. However, a well-done systematic review and meta-analysis comparing omega-3 fatty acids with placebo found only a small, nonsignificant benefit that was largely attributable to publication bias.⁵³ Currently, the Cochrane Collaboration is conducting a systematic review on its use for treatment of MDD.¹⁴⁴ Our network meta-analyses clearly favored treatment with SGAs over omega-3-fatty acids monotherapy.

Although the Canadian guidelines recommend the use of SAME as monotherapy for mild to moderate MDD, the U.S. report calls for more studies to determine its efficacy.^{48,49} Most studies of SAME are limited to parenteral administration of the supplement, which appears to have better efficacy than a placebo.¹⁴⁵ However, few studies evaluate oral preparations, and little is known about optimal SAME dosing.¹⁴¹ We found only one trial to evaluate comparative effectiveness and concluded evidence was insufficient to make a recommendation for (or against) use of SAME.

St. John's wort is perhaps the most commonly evaluated CAM therapy for MDD patients. A Cochrane review evaluated 18 RCTs comparing St. John's wort with placebo; it demonstrated superior efficacy for St. John's wort but noted high heterogeneity among trials. However, their analysis of 17 head-to-head RCTs comparing St. John's wort with both tricyclic antidepressants and SSRIs demonstrated similar treatment effectiveness for patients with mild to moderate MDD.¹⁴³ Both the U.S. and Canadian reviews recommend St. John's wort for first-line treatment of mild to moderate MDD, whereas there is less consensus on its use for severe MDD. Similarly, we found moderate SOE to support the use of St. John's wort for MDD. Interestingly, most trials we identified included participants with severe MDD. However, many trials did not use the full dose ranges of SGAs in their comparisons with St. John's wort.

Numerous systematic reviews and meta-analyses have been done on exercise for depression.^{47,57,58,146-150} These reviews have examined a variety of types of exercise, including walking, aerobic and nonaerobic forms of movement, and strength training, using randomized and nonrandomized designs and various comparison groups, including no treatment, wait-list controls, and active treatments. Overall, exercise has been found to have a small to moderate clinical benefit when compared with no treatment, wait-list, or placebo and comparable benefit when compared with other active treatments, including SGAs. Our findings are consistent with the recent Cochrane Review by Cooney et al.⁴⁷ that included a separate analysis of SGAs versus exercise and found that the SGA (sertraline) was no more effective than exercise for reducing depression. The Cooney et al. report included four studies—we included two in our review and excluded the other two; for the latter, one was excluded because the population was older adults with minor depression rather than MDD¹⁵¹ and the other because the population was patients with coronary artery disease.⁵⁹

Current literature suggests that depression severity is an important factor to consider when deciding to treat with an antidepressant. In particular, patients with higher severity MDD respond better to medication than those with lower severity depression.²¹ Based on trials that met the eligibility criteria for our report, we could not draw any firm conclusions about whether depression severity influences the comparative benefits and harms of SGAs and psychological interventions or CAM treatments.

Applicability

The scope of this review was limited to trials that enrolled adult patients with MDD. We did not attempt to review literature on interventions for MDD in children or for patients with subthreshold depression, dysthymia, or perinatal depression. Because of the serious methodological limitations of some trials, the degree of applicability of some of our findings to real-world settings might be compromised, grades of low or insufficient for SOE also reflect that problem.

The included trials covered populations with mild, moderate, and severe MDD. Most trial populations, however, excluded patients with medical comorbidities; few trials included elderly patients. Furthermore, most trials were conducted in clinical settings. Results from samples of patients attending a clinic might not apply to members of the general community who suffer from MDD of the same type. Similarly, we did not find evidence to confirm or refute whether treatments are more or less efficacious for various subgroups: patients characterized by sex, race, or ethnicity or individuals with coexisting psychiatric conditions. The samples in many trials had some subjects with the aforementioned subgroup characteristics, even if the main focus was on a different population. For instance, the trials may have included individuals with a history of psychiatric comorbidities but did not report whether interventions were similarly efficacious (or not) for such individuals. Finally, many trials, particularly for CAM interventions, were conducted outside the United States. Whether and how differences in ethnic or cultural backgrounds and health systems affect the applicability of results to U.S. populations remains uninvestigated and unanswered. For example, most of the acupuncture trials were conducted in China, where acupuncture is commonplace, and the effects of acupuncture treatment expectancy may differ substantially between such populations and Western populations.¹⁵²

With few exceptions, interventions in included trials were in line with clinical practice. Except for some CAM trials in which patients received SGA dosages at the lower end of the recommended range, prescribing patterns and doses in the SGA arms of our evidence base were consistent with clinical practice. Some newer SGAs such as desvenlafaxine, levomilnacipran, vilazodone, or vortioxetine, however, have never been compared with psychological or CAM treatments or exercise. Nevertheless, reliable evidence indicates that the comparative effectiveness of SGAs is similar.³² Consequently, we believe that our findings are applicable across the class of SGAs.

As noted above, detecting no statistically significant difference does not necessarily mean the treatment options are equivalent. The studies involved were designed to test whether an outcome for one intervention was different from another rather than to test equivalence, which would generally require a larger sample size. This point is especially relevant for those findings with a low SOE. Further, while comparative effectiveness at a group level did not detect a difference between SGA and CBT or St John's wort, how best to tailor this information to an individual patient is still not clear. Indeed, other potentially relevant indicators (e.g., depressive severity, comorbid psychiatric illness) may favor one over another, but the current evidence base (as indicated in the KQ 1b and 2b findings) is quite limited.

The number and length of sessions of the various psychological interventions were generally consistent with clinical practice and likely represent an adequate course of treatment. As is generally the case when comparing the effectiveness of psychological treatments with other psychological interventions or other types of treatment, heterogeneity of the content and delivery of the identified intervention is problematic. Many of the psychological interventions in our

evidence base provided broad descriptions of the type of intervention; others used a manualized protocol.

Further, variability among the trials was high with respect to the degree to which treatment fidelity was assessed and adhered to. Type, training, and experience of the providers of the various interventions were also quite heterogeneous. Although clinician characteristics may be less problematic than the content of the intervention for understanding comparative benefits or harms, unlike the case with SGAs that are broadly equivalent and have standardized dosing, the cumulative effect of the various sources of heterogeneity within and across psychological interventions may limit the applicability of our findings.

For acupuncture, treatment protocols were so varied as to preclude definitive conclusions about any single acupuncture intervention. For these reasons, we find it difficult to recommend any single type of acupuncture, or acupuncture more generally, as a substitute for treatment with antidepressant medications.

For St. John's wort, use of standardized extracts may be broadly applicable with certain caveats. Although several different St. John's wort preparations were represented among the trials we found, many of the trials used St. John's wort doses that were consistent with current recommendations (i.e., 900 mg daily, standardized to 0.1 percent to 0.3 percent hypericin).^{41,42,153} Furthermore, high quality, standardized St. John's wort extracts are now commonly available.

An important concern about the use of St. John's wort is its potential to interact with other medications. St. John's wort is well known to cause substantial changes in plasma concentrations of drugs metabolized by cytochrome P450 3A4, which includes SSRIs, tricyclic antidepressants, and many drugs used to treat common conditions such as heart disease, hypertension, hypercholesterolemia, HIV, and many cancers.¹⁵⁴ Therefore, St. John's wort should not be recommended to patients taking any pharmaceutical medications without the advice of a medical provider or pharmacist with expertise in evaluating herb-drug interactions.

Doses in the exercise arms were within the dose range suggested for exercise programs for middle-age to older adults. For example, the guideline for depression from the National Institute of Health and Care Excellence recommends structured, supervised exercise three times per week.³⁵ However, the small numbers of trials that have examined dose-response of exercise for depression indicate that higher intensity and frequency of exercise may be more helpful in alleviating depression.⁴⁷ Although the two Blumenthal et al. trials, reported reasonable compliance rates for both the SGA and exercise groups, in clinical practice depressed patients may well have more difficulty staying motivated to exercise, because depression is known to be associated with lower levels of physical activity.¹⁵⁵ Although our report had insufficient data to determine whether depression enhances quality of life, we did find that aerobic capacity increased significantly more among the exercise group. Because both of these trials targeted middle-age and older adults, the results cannot be generalized to younger age groups.

Most trials did not assess quality of life or functional capacity as outcomes. Conceivably, response to treatment and remission does also improve quality of life and functional capacity.

The lack of assessment of harms in many trials poses a serious threat to the applicability of findings to average settings or patient populations. The comparative balance of benefits and harms among treatment options is impossible to determine when harms are not assessed and reported reliably. In clinical trials of SGAs with close adverse events surveillance, up to 60 percent of patients experienced adverse events.³² For some patients, these adverse effects were tolerable; for others, they led to discontinuation of treatment. In the body of evidence for this

report, neither harms for SGAs nor harms for nonpharmacological treatments were assessed adequately. For that reason, we could not draw any conclusions about applicability.

Implications for Clinical and Policy Decisionmaking

Results and outcomes with moderate SOE detecting no statistically significant difference in effectiveness can serve as a reasonable starting place for providers and patients for starting a course of medication or psychotherapy to treat MDD. We caution, however, that whether this conclusion differs as a function of depression severity is still unknown. The specific psychotherapy interventions include cognitive behavioral therapy, interpersonal therapy, psychodynamic psychotherapy, and behavioral activation.

Health care reform around the world reflects a trend toward integrative care as a remedy for the current, fragmented delivery of health and social services common in many health care systems. Given that both SGAs and psychotherapies can have equal merit in treating MDD, locating clinicians who render mental health care in primary care settings needs to be part of this trend. Doing so would likely increase patient access to therapy and enhance coordination of care between primary care clinicians and therapists. Further, we know that approximately 20 percent of patients do not fill their prescriptions for antidepressant medication; even if they start a course of treatment, they may discontinue early before receiving an adequate course.¹⁵⁶ Having access to nonpharmacologic interventions in the primary care setting might enhance treatment adherence and improve treatment outcomes for patients with MDD. It may also have additional downstream effects in reducing the stigma associated with mental illness in general, empowering patients to address the symptoms and issues associated with not only depression but also other mental health–related concerns, and encouraging them to seek and maintain treatment more quickly at an earlier stage of their illness.

Related to this, access to psychotherapy should not be financially prohibitive. Some insurance plans in the United States consider psychotherapy a specialty and charge different copayment rates for those services than they do for generalized medical care. In as much as psychotherapy is a special type of care, decision- and policymakers need to make sure that fees associated with accessing these interventions do not make them unaffordable for patients that need and would benefit from these services the most.

Similarly, one great difficulty for CAM therapies, for both patients and providers, is how to pay for them. For most patients, their insurers do not cover CAM services. This is particularly vexing for patients and providers, especially when the weight of the evidence favors efficacy for the CAM treatment (e.g., ginger root for treatment of chemotherapy-induced nausea and vomiting, St. John's wort for treatment of depression). In many of these instances, patients need to pay for these treatments out of pocket, which creates disparities in care by limiting access to proven treatments for patients who cannot manage those out-of-pocket costs.

The moderate SOE regarding the lack of statistically significant differences in effectiveness of SGAs and exercise, combined with the low adverse effects generally found in exercise trials, can provide clinicians with reasonable choices as to how to guide their patients in clinical practice. In particular, those patients who strongly prefer one or the other therapy can be allowed freedom to choose a course of exercise or a course of antidepressants, while under a physician's supervision and monitoring. Moreover, those patients who would like to maintain or start an exercise regimen in addition to undergoing SGA therapy can be encouraged to do so. The enhanced potential for increasing physical well-being as well as expanding social interactions may be an added incentive to encourage an exercise regimen.

Limitations of the Comparative Effectiveness Review Process

To find relevant studies, we employed an intensive search process in multiple electronic databases; we also conducted searches for grey literature. Because of time and monetary limitations, however, we limited eligible studies to those published in English, German, and Italian. Methods research indicates that such an approach can introduce language bias; in general, however, it may also lead to overestimates of the effectiveness of interventions.

For KQ 2, we extended eligibility criteria after we realized that we would not find sufficient evidence to answer this KQ. Despite re-reviewing more than 6,000 abstracts, we could still not find reliable evidence to address the question about the best treatment option for patients who did not achieve remission during an initial treatment trial.

For harms, studies conducted in other patient populations (e.g., those with subthreshold depression or dysthymia) might have yielded useful information. Many studies using psychological or CAM therapies included populations suffering from any form of depression, not just MDD. In addition, studies with placebo or waiting list control groups could have provided important information about adverse effects of interventions. We lacked the resources to explore such a broad evidence base just to assess harms.

Because we dealt with study-level data, we could not reliably assess the impact of severity of MDD on the comparative benefits and harms of interventions. Such a question would best be addressed with individual patient data from trials and individual patient data meta-analyses.

If information in full-text articles was unclear or missing, we attempted to contact authors for clarification. The yield of this effort, however, was small. Despite multiple attempts to contact authors, few replied or were able to provide missing information.

Finally, publication bias and selective outcome reporting are potential limitations. Although we searched for grey and unpublished literature, the extent and impact of publication and reporting bias in this body of evidence is impossible to determine.

Limitations of the Evidence Base

Overall, several major limitations characterize this body of evidence. First, no reliable evidence was available assessing the effectiveness or risk of harms of many of our eligible interventions. Particularly for KQ 2 on populations who did not achieve response to an initial treatment attempt, we found no eligible switch trials directly comparing SGAs with CAM or exercise; neither did we find any eligible augmentation trials comparing SGAs with CAM or exercise. We also found no direct comparisons of switching strategies versus augmentation strategies. Likewise, the role of depressive severity as a moderator of the comparative effectiveness of both first- and second-line therapies has received very little planned, prospective study.

Second, even when evidence was available, the small number of trials and the small sample sizes posed considerable limitations. Much of the evidence base directly comparing treatments was powered to test whether one treatment was superior to the other. Failure to find such a difference is not equal to concluding that the interventions are equivalent. In addition, for some trials we had concerns about adequate dosing of SGAs. For example, three of eight trials compared St. John's wort to either fluoxetine 20 mg or sertraline 50 mg, the lowest recommended doses of these drugs. Considering that mean baseline depressive severity for most trials fell in the severe range (HAM-D scores 19 to 23), patients in the SGA arms were

undertreated. The extent to which this affects the comparative benefits between SGAs and St. John's wort remains unclear.

Third, available evidence was frequently fraught with methodological shortcomings. Of the 45 trials meeting our eligibility criteria, we rated 16 as high risk of bias and only 4 as low risk of bias. Trials assessed as high risk of bias have significant flaws of various types (e.g., stemming from serious errors in design, conduct, or analysis) that may invalidate their results. Consequently, the evidence base for most critical outcomes was insufficient to draw conclusions. The SOE could be rated as low or moderate for only a few outcomes; the latter indicates reasonable confidence in the effect estimates from those trials.

Fourth, even when trials assessing the comparative effectiveness of interventions were available, they often did not assess harms or did not assess harms adequately. Of the 45 included trials, only one trial used an objective scale to assess harms. Most trials combined spontaneous patient-reported adverse events with a regular clinical examination by an investigator. Determining whether assessment methods were unbiased and adequate was often difficult. Rarely did authors report whether adverse events were prespecified and defined. Short trial durations and small sample sizes also limited the validity of adverse event assessment in many trials. No trials were designed to assess specific adverse events as primary outcomes.

Fifth, of the limited body of evidence, most trials were explanatory rather than pragmatic trials; this factor may well compromise the applicability of findings.

Research Gaps

Across all comparisons of interventions, major research gaps pertain to information about the comparative risk of harms and patient-relevant outcomes such as functional capacity and quality of life. For patients and clinicians, balancing benefits and harms based on objective information is crucial. Lack of information about harms can lead to a biased knowledge base and the potential for decisions that cause more harm than good. Findings from the STAR*D study suggest that factors other than depression severity contribute significantly to the health-related quality of life of outpatients with MDD. A comprehensive assessment of quality of life outcomes is, therefore, paramount for informed decisions about treatment options.

We found no eligible studies that compared SGAs with behavior therapy or behavior modification, humanistic therapies, yoga, or mindfulness interventions. Given the wide use of these types of psychotherapies in clinical practice, further research into their comparative effectiveness with SGAs in treating MDD patients is desirable. For many psychotherapies and all CAM therapies that have been evaluated against an SGA, the data were insufficient because trials did not report important outcomes, most notably quality of life and functional capacity. Future studies should assess remission, response to treatment, quality of life, and functional capacity using standardized measures to allow for more direct comparisons across studies using the same or similar SGAs and psychological interventions.

These same deficiencies in the literature extend to the comparative effectiveness of SGAs and both psychological and CAM interventions for treating MDD as a function of depression severity. Only a single trial evaluating SAME and no trials assessing psychological interventions or other CAM therapies were designed to address the question of whether depression severity affects the comparative effectiveness of SGAs as compared with these interventions.

Research comparing an SGA with exercise, either alone or in combination with an SGA, is also limited. We found only two trials comparing SGAs with exercise that met our criteria, and these both used aerobic exercise, in which individuals were assigned continuous walking or

jogging that would maintain heart rate from 70 to 85 percent of their heart rate reserve. Missing from the literature were any studies meeting our criteria using other forms of exercise (e.g., strength training or mindful exercises such as yoga, tai chi, or qigong). Moreover, we found no studies in which an SGA was systematically compared with differing intensities and frequencies of exercise (this research could be helpful, because there is indication from non-SGA studies of better treatment outcomes with high-dose versus low-dose exercise regimens).¹⁵⁷ Changes in aerobic capacity were reported in both our included trials; more trials, however, should include standardized measures of quality of life and functional capacity. Having such data might then enable reviewers to compare results across trials. Trials that include a wider age range of participants would also be helpful in determining whether different types and intensities of exercise are more effective for patients of different ages; preferences and usefulness of various types and intensities of exercises may differ by sex or ethnic or cultural variables. Research should also investigate how baseline depression severity affects patient preferences, adherence, and outcomes of prescribing an SGA versus exercise or exercise-SGA combination.

One primary challenge for studies of CAM therapies is defining the proper dose of the therapy being tested. Although experts tend to agree about dosing of St. John's wort,^{41,42,153} only scant evidence informs dosing regimens of SAME, and dosing practices for omega-3 fatty acids differ widely. Future studies of natural products should be based on dosing regimens that are supported by investigations of their pharmacokinetic and dose-response properties. Similar problems exist for acupuncture dosing, but this particular issue is even more complex because of the heterogeneity of point selection, needle stimulation, session duration, and number of treatments for acupuncture interventions.

The limited amount of comparative intervention data addressing whether depressive severity moderates outcomes provides little guidance on how selection of treatment strategies might differ based on whether a depression is on the milder end of the spectrum compared with the more severe end. This question, raised by a number of systematic reviews,²¹⁻²³ remains without a clear answer.

Finally, beyond the two articles identified comparing switch and augmentation strategies employing a limited number of medication options or CT, the absence of relevant comparative data about which treatment options are most effective for those needing second-line treatment (about 70 percent of patients with MDD)^{25,26} was striking. Further, no second-line therapy data at all exist comparing SGAs with CAM or exercise treatments. This void in the evidence base is a major one that will perplex and confound clinicians, patients, policymakers, and guideline-developers alike.

Conclusions

Available evidence indicates that SGAs, CBT, and St. John's wort do not differ significantly in effectiveness as first-line treatments for adult outpatients with mild to severe MDD. The SOE for these findings is moderate which means that the body of evidence has some deficiencies, but we believe that the findings are likely to be stable as new studies emerge. Most comparisons of SGAs with other treatment options also did not detect statistically significant differences. Exceptions, however, are omega-3-fatty acids that appear to have lower effectiveness than SGAs and the combination of SGAs with acupuncture which appears to have greater effectiveness than SGA monotherapy. These findings, however, have to be interpreted cautiously because of methodological limitations. Our confidence in these results was low or evidence was simply insufficient. We believe that future studies will have a substantial impact on results. In addition,

populations with MDD are known to have high response rates to placebos. For many comparisons that are limited to single trials, determining whether similar treatment effects between SGAs and other interventions are based on similar effectiveness or high placebo response rates is impossible.¹⁵⁸

Interventions other than SGAs usually have a lower risk for harms. Some, however, require more personal engagement or costs than others, which could affect patient adherence.

The choice of the initial treatment of MDD should, therefore, be strongly based on patient preferences following a discussion of the advantages and disadvantages and the feasibility (e.g. costs, likely adherence) of each treatment option. Differences with respect to adverse events, personal engagement, and costs may be taken into consideration for the choice of a first-line treatment. Such shared and informed decisionmaking might enhance treatment adherence and improve treatment outcomes for patients with MDD, especially because treatment continuity is one of the main challenges in treating such patients.¹⁵⁹

For second-line therapies, although evidence is limited, no clear benefit emerges to suggest either switching to a particular SGA or to cognitive therapy or augmenting with a particular medication or cognitive therapy. The more important decision appears to be simply to try a different evidence-based approach.

References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed., Arlington, VA: American Psychiatric Publishing; 2013.
2. Kessler RC, Berglund P, Demler O, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA*. 2003 Jun 18;289(23):3095-105. Epub: 2003/06/19. PMID: 12813115.
3. Khan A, Sambunaris A, Edwards J, et al. Vilazodone in the treatment of major depressive disorder: efficacy across symptoms and severity of depression. *Int Clin Psychopharmacol*. 2014 Mar;29(2):86-92. Epub: 2013/11/20. PMID: 24247740.
4. Nemeroff CB. The burden of severe depression: a review of diagnostic challenges and treatment alternatives. *J Psychiatr Res*. 2007 Apr-Jun;41(3-4):189-206. Epub: 2006/07/28. PMID: 16870212.
5. Thase ME. Treatment of severe depression. *J Clin Psychiatry*. 2000;61 Suppl 1:17-25. Epub: 2000/03/07. PMID: 10703759.
6. The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries, and Risk Factors in 1990 and Projected to 2020. Geneva, Switzerland: Harvard School of Public Health on behalf of the World Health Organization and the World Bank 1996.
7. Fendrich M, Avci O, Johnson TP, et al. Depression, substance use and HIV risk in a probability sample of men who have sex with men. *Addict Behav*. 2013 Mar;38(3):1715-8. PMID: 23254224.
8. Silberbogen AK, Busby AK, Ulloa EW. Impact of Psychological Distress on Prostate Cancer Screening in U.S. Military Veterans. *Am J Mens Health*. 2013 Dec 20;8(5):399-408. Epub: 2013/12/24. PMID: 24362494.
9. DiMatteo MR, Lepper HS, Croghan TW. Depression is a risk factor for noncompliance with medical treatment: meta-analysis of the effects of anxiety and depression on patient adherence. *Arch Intern Med*. 2000 Jul 24;160(14):2101-7. Epub: 2000/07/25. PMID: 10904452.
10. McLaughlin KA. The public health impact of major depression: a call for interdisciplinary prevention efforts. *Prev Sci*. 2011 Dec;12(4):361-71. Epub: 2011/07/07. PMID: 21732121.
11. Farmer A, Korszun A, Owen MJ, et al. Medical disorders in people with recurrent depression. *Br J Psychiatry*. 2008 May;192(5):351-5. Epub: 2008/05/03. PMID: 18450658.
12. Chen YH, Lin HC. Increased risk of cancer subsequent to severe depression--a nationwide population-based study. *J Affect Disord*. 2011 Jun;131(1-3):200-6. Epub: 2011/01/19. PMID: 21242002.
13. Greenberg PE, Kessler RC, Birnbaum HG, et al. The economic burden of depression in the United States: how did it change between 1990 and 2000? *J Clin Psychiatry*. 2003 Dec;64(12):1465-75. Epub: 2004/01/20. PMID: 14728109.
14. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: The National Academies Press; 2001.
15. Mojtabai R, Olfson M. National patterns in antidepressant treatment by psychiatrists and general medical providers: results from the national comorbidity survey replication. *J Clin Psychiatry*. 2008 Jul;69(7):1064-74. Epub: 2008/04/11. PMID: 18399725.
16. Gaynes BN, Rush AJ, Trivedi M, et al. A direct comparison of presenting characteristics of depressed outpatients from primary vs. specialty care settings: preliminary findings from the STAR*D clinical trial. *Gen Hosp Psychiatry*. 2005 Mar-Apr;27(2):87-96. Epub: 2005/03/15. PMID: 15763119.
17. Gaynes BN, Rush AJ, Trivedi MH, et al. Major depression symptoms in primary care and psychiatric care settings: a cross-sectional analysis. *Ann Fam Med*. 2007 Mar-Apr;5(2):126-34. Epub: 2007/03/29. PMID: 17389536.
18. Wang PS, Lane M, Olfson M, et al. Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005 Jun;62(6):629-40. Epub: 2005/06/09. PMID: 15939840.
19. American Psychiatric Association (APA). Practice Guideline for the Treatment of Patients with Major Depressive Disorder.

- 3rd ed., Arlington, VA: American Psychiatric Association (APA); 2010.
20. Qaseem A, Snow V, Denberg TD, et al. ACP clinical guidelines. Using second-generation antidepressants to treat depressive disorders: a clinical practice guideline from the American College of Physicians. *Ann Intern Med.* 2008;149(10):725. PMID: 2010112496. Corporate Author: Clinical Efficacy Assessment Subcommittee of American College of Physicians. Language: English. Entry Date: 20090123. Revision Date: 20090123. Publication Type: journal article.
 21. Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA.* 2010 Jan 6;303(1):47-53. Epub: 2010/01/07. PMID: 20051569.
 22. Khan A, Leventhal RM, Khan SR, et al. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J Clin Psychopharmacol.* 2002 Feb;22(1):40-5. Epub: 2002/01/19. PMID: 11799341.
 23. Kirsch I, Deacon BJ, Huedo-Medina TB, et al. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med.* 2008 Feb;5(2):e45. Epub: 2008/02/29. PMID: 18303940.
 24. Mitchell AJ, Rao S, Vaze A. Can general practitioners identify people with distress and mild depression? A meta-analysis of clinical accuracy. *J Affect Disord.* 2011 Apr;130(1-2):26-36. Epub: 2010/08/17. PMID: 20708274.
 25. Gaynes BN, Rush AJ, Trivedi MH, et al. Primary versus specialty care outcomes for depressed outpatients managed with measurement-based care: results from STAR*D. *J Gen Intern Med.* 2008 May;23(5):551-60. Epub: 2008/02/06. PMID: 18247097.
 26. Gaynes BN, Lux LJ, Lloyd SW, et al. Nonpharmacologic Interventions for Treatment-Resistant Depression in Adults. Comparative Effectiveness Reviews, No. 33. (Prepared by RTI International-University of North Carolina (RTI-UNC) Evidence-based Practice Center under Contract No. 290-02-0016I.) AHRQ Publication No. 11-EHC056-EF. Rockville, MD: Agency for Healthcare Research and Quality; Sep 2011.
 27. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry.* 2006 Nov;163(11):1905-17. Epub: 2006/11/01. PMID: 17074942.
 28. Gaynes BN, Warden D, Trivedi MH, et al. What did STAR*D teach us? Results from a large-scale, practical, clinical trial for patients with depression. *Psychiatr Serv.* 2009 Nov;60(11):1439-45. Epub: 2009/11/03. PMID: 19880458.
 29. Mojtabai R, Olfson M. National trends in long-term use of antidepressant medications: results from the US National Health and Nutrition Examination Survey. *J Clin Psychiatry.* 2014 Feb;75(2):169-77. Epub: 2013/12/19. PMID: 24345349.
 30. Kupfer DJ, Perel JM, Pollock BG, et al. Fluvoxamine versus desipramine: comparative polysomnographic effects. *Biol Psychiatry.* 1991 Jan 1;29(1):23-40. Epub: 1991/01/01. PMID: 1900443.
 31. Linde K, Schumann I, Meissner K, et al. Treatment of depressive disorders in primary care--protocol of a multiple treatment systematic review of randomized controlled trials. *BMC Fam Pract.* 2011;12:127. Epub: 2011/11/17. PMID: 22085705.
 32. Gartlehner G, Hansen RA, Morgan LC, et al. Comparative benefits and harms of second-generation antidepressants for treating major depressive disorder: an updated meta-analysis. *Ann Intern Med.* 2011 Dec 6;155(11):772-85. Epub: 2011/12/08. PMID: 22147715.
 33. Gartlehner G, Hansen RA, Morgan LC, et al. Second-Generation Antidepressants in the Pharmacologic Treatment of Adult Depression: An Update of the 2007 Comparative Effectiveness Review. (Prepared by the RTI International--University of North Carolina Evidence-based Practice Center, Contract No. 290-2007-10056-I.) AHRQ Publication No. 12-EHC012-EF. Rockville, MD: Agency for Healthcare Research and Quality; December 2011. www.effectivehealthcare.ahrq.gov/reports/final.cfm.
 34. Gartlehner G, Thieda P, Hansen RA, et al. Comparative risk for harms of second-generation antidepressants : a systematic

- review and meta-analysis. *Drug Saf.* 2008;31(10):851-65. Epub: 2008/09/02. PMID: 18759509.
35. Health NCCfM. National Institute for Health and Clinical Excellence: Guidance. Depression: The Treatment and Management of Depression in Adults (Updated Edition). Leicester (UK): British Psychological Society. Copyright (c) The British Psychological Society & The Royal College of Psychiatrists; 2010.
 36. National Guideline Clearinghouse. Non-pharmaceutical management of depression in adults. A national clinical guideline. Rockville MD: Agency for Healthcare Research and Quality (AHRQ) <http://www.guideline.gov/content.aspx?id=15596>. Accessed 9/19/2014.
 37. Gaynes BN, Dusetzina SB, Ellis AR, et al. Treating depression after initial treatment failure: directly comparing switch and augmenting strategies in STAR*D. *J Clin Psychopharmacol.* 2012 Feb;32(1):114-9. Epub: 2011/12/27. PMID: 22198447.
 38. Cuijpers P, van Straten A, Andersson G, et al. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. *J Consult Clin Psychol.* 2008 Dec;76(6):909-22. Epub: 2008/12/03. PMID: 19045960.
 39. Nieuwsma JA, Trivedi RB, McDuffie J, et al. Brief psychotherapy for depression: A systematic review and meta-analysis. *Int J Psychiatry Med.* 2012;43(2):129-51. PMID: 2012-17351-003. PMID: 22849036. First Author & Affiliation: Nieuwsma, Jason A.
 40. Spielmanns GI, Berman MI, Usitalo AN. Psychotherapy versus second-generation antidepressants in the treatment of depression: A meta-analysis. *J Nerv Ment Dis.* 2011;199(3):142-9. PMID: 21346483.
 41. European Scientific Cooperative on Phytotherapy. E/S/C/O/P monographs : the scientific foundation for herbal medicinal products. Exeter, U.K.; Stuttgart, Germany; New York: European Scientific Cooperative on Phytotherapy; Thieme; 2003.
 42. Blumenthal M. Herbal Medicine: Expanded Commission E Monographs. Integrative Medicine Communications; 2000.
 43. Butterweck V. St. John's Wort: Quality Issues and Active Compounds. In: Cooper R, Kronenberg F, eds. *Botanical Medicine: From Bench to Bedside.* Mary Ann Liebert; 2009:69-91.
 44. Maratos AS, Gold C, Wang X, et al. Music therapy for depression. *Cochrane Database Syst Rev.* 2008(1):CD004517. Epub: 2008/02/07. PMID: 18254052.
 45. Jorm AF, Morgan AJ, Hetrick SE. Relaxation for depression. *Cochrane Database Syst Rev.* 2008(4):CD007142. Epub: 2008/10/10. PMID: 18843744.
 46. Smith CA, Hay PP, Macpherson H. Acupuncture for depression. *Cochrane Database Syst Rev.* 2010(1):CD004046. Epub: 2010/01/22. PMID: 20091556.
 47. Cooney GM, Dwan K, Greig CA, et al. Exercise for depression. *Cochrane Database Syst Rev.* 2013;9:CD004366. Epub: 2013/09/13. PMID: 24026850.
 48. Freeman MP, Fava M, Lake J, et al. Complementary and alternative medicine in major depressive disorder: the American Psychiatric Association Task Force report. *J Clin Psychiatry.* 2010 Jun;71(6):669-81. Epub: 2010/06/25. PMID: 20573326.
 49. Ravindran AV, Lam RW, Filteau MJ, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) Clinical guidelines for the management of major depressive disorder in adults. V. Complementary and alternative medicine treatments. *J Affect Disord.* 2009 Oct;117 Suppl 1:S54-64. Epub: 2009/08/12. PMID: 19666194.
 50. Linde K, Berner MM, Kriston L. St John's wort for major depression. *Cochrane Database Syst Rev.* 2008(4):CD000448. Epub: 2008/10/10. PMID: 18843608.
 51. Jorm AF, Morgan AJ, Hetrick SE. Relaxation for depression. *Cochrane Database Syst Rev.* 2008(4):CD007142. Epub: 2008/10/10. PMID: 18843744.
 52. Balasubramaniam M, Telles S, Doraiswamy PM. Yoga on our minds: a systematic review of yoga for neuropsychiatric disorders. *Front Psychiatry.* 2012;3:117. Epub: 2013/01/29. PMID: 23355825.
 53. Bloch MH, Hannestad J. Omega-3 fatty acids for the treatment of depression: systematic review and meta-analysis. *Mol Psychiatry.* 2012 Dec;17(12):1272-82. Epub: 2011/09/21. PMID: 21931319.
 54. Papakostas GI. Evidence for S-adenosyl-L-methionine (SAM-e) for the treatment of major depressive disorder. *J Clin Psychiatry.* 2009;70 Suppl 5:18-22. Epub: 2010/01/14. PMID: 19909689.
 55. Freeman MP. Complementary and Alternative Medicine (CAM): considerations for the treatment of major

- depressive disorder. *J Clin Psychiatry*. 2009;70 Suppl 5:4-6. Epub: 2010/01/14. PMID: 19909686.
56. Kessler RC, Soukup J, Davis RB, et al. The use of complementary and alternative therapies to treat anxiety and depression in the United States. *Am J Psychiatry*. 2001 Feb;158(2):289-94. Epub: 2001/02/07. PMID: 11156813.
 57. Rimer J, Dwan K, Lawlor DA, et al. Exercise for depression. *Cochrane Database Syst Rev*. 2012;7:CD004366. Epub: 2012/07/13. PMID: 22786489.
 58. Bridle C, Spanjers K, Patel S, et al. Effect of exercise on depression severity in older people: Systematic review and meta-analysis of randomised controlled trials. *Br J Psychiatry*. 2012;201(3):180-5. PMID: 22945926.
 59. Blumenthal JA, Sherwood A, Babyak MA, et al. Exercise and pharmacological treatment of depressive symptoms in patients with coronary heart disease: results from the UPBEAT (Understanding the Prognostic Benefits of Exercise and Antidepressant Therapy) study. *J Am Coll Cardiol*. 2012 Sep 18;60(12):1053-63. Epub: 2012/08/04. PMID: 22858387.
 60. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009;339:b2535. Epub: 2009/07/23. PMID: 19622551.
 61. AHRQ Methods for Effective Health Care. *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008.
 62. (QIDS) IoDSIaQIoDS. Inventory of Depressive Symptomatology (IDS) and Quick Inventory of Depressive Symptomatology (QIDS). University of Pittsburgh Epidemiology Data Center; 2010.
 63. Cochrane Depression, Anxiety, and Neurosis Group. CCDAN Topic List: Intervention - Psychological therapies. 2013 http://ccdan.cochrane.org/sites/ccdan.cochrane.org/files/uploads/CCDAN%20topics%20list_psychological%20therapies%20for%20website.pdf. Accessed October 17, 2014.
 64. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol*. 2011 Apr;64(4):395-400. Epub: 2011/01/05. PMID: 21194891.
 65. Wallace BC, Small K, Brodley CE, et al. Deploying an interactive machine learning system in an evidence-based practice center: abstract. *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*; 2012 Miami, Florida, USA. ACM; pp. 819-24.
 66. Ioannidis JP, Evans SJ, Gotzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med*. 2004 Nov 16;141(10):781-8. Epub: 2004/11/17. PMID: 15545678.
 67. Viswanathan M, Ansari MT, Berkman ND, et al. Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. Agency for Healthcare Research and Quality *Methods Guide for Comparative Effectiveness Reviews* AHRQ Publication No. 12-EHC047-EF. Rockville MD: Agency for Healthcare Research and Quality; March 2012. www.effectivehealthcare.ahrq.gov/.
 68. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928. Epub: 2011/10/20. PMID: 22008217.
 69. Deeks JJ, Dinnes J, D'Amico R, et al. Evaluating non-randomised intervention studies. *Health Technol Assess*. 2003;7(27):iii-x, 1-173. Epub: 2003/09/23. PMID: 14499048.
 70. Shea BJ, Hamel C, Wells GA, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol*. 2009 Oct;62(10):1013-20. Epub: 2009/02/24. PMID: 19230606.
 71. West SL, Gartlehner G, Mansfield AJ, et al. *Comparative Effectiveness Review Methods: Clinical Heterogeneity*. *Methods Research Report*. (Prepared by the RTI International–University of North Carolina Evidence-based Practice Center (EPC) under Contract No. 290-2007-10056-I.) AHRQ Publication No. 10-EHC070-EF. Rockville MD: Agency for Healthcare Research and Quality; September 2010. www.effectivehealthcare.ahrq.gov/.
 72. Hong H, Carlin BP, Shamliyan TA, et al. Comparing Bayesian and frequentist approaches for multiple outcome mixed

- treatment comparisons. *Med Decis Making*. 2013 Jul;33(5):702-14. Epub: 2013/04/04. PMID: 23549384.
73. Jones B, Roger J, Lane PW, et al. Statistical approaches for conducting network meta-analysis in drug development. *Pharm Stat*. 2011 Nov-Dec;10(6):523-31. Epub: 2012/01/04. PMID: 22213533.
74. Glenny AM, Altman DG, Song F, et al. Indirect comparisons of competing interventions. *Health Technol Assess*. 2005 Jul;9(26):1-134, iii-iv. Epub: 2005/07/15. PMID: 16014203.
75. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004 Oct 30;23(20):3105-24. Epub: 2004/09/28. PMID: 15449338.
76. Berkman ND, Lohr KN, Ansari M, et al. Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update. *Methods Guide for Effectiveness and Comparative Effectiveness Reviews (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290-2007-10056-I) AHRQ Publication No. 13(14)-EHC130-EF*. Rockville MD: Agency for Healthcare Research and Quality; November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.
77. Atkins D, Chang S, Gartlehner G, et al. Assessing the Applicability of Studies When Comparing Medical Interventions *Methods Guide for Comparative Effectiveness Reviews AHRQ Publication No. 11-EHC019-EF*. Rockville MD: Agency for Healthcare Research and Quality; January 2011.
78. McGrath CL, Kelley ME, Holtzheimer PE, et al. Toward a neuroimaging treatment selection biomarker for major depressive disorder. *JAMA Psychiatry*. 2013 Aug;70(8):821-9. Epub: 2013/06/14. PMID: 23760393.
79. Barber JP, Barrett MS, Gallop R, et al. Short-term dynamic psychotherapy versus pharmacotherapy for major depressive disorder: a randomized, placebo-controlled trial. *J Clin Psychiatry*. 2012 Jan;73(1):66-73. Epub: 2011/12/14. PMID: 22152401.
80. Frank E, Cassano GB, Rucci P, et al. Predictors and moderators of time to remission of major depression with interpersonal psychotherapy and SSRI pharmacotherapy. *Psychol Med*. 2011 Jan;41(1):151-62. Epub: 2010/04/13. PMID: 20380782.
81. Salminen JK, Karlsson H, Hietala J, et al. Short-term psychodynamic psychotherapy and fluoxetine in major depressive disorder: a randomized comparative study. *Psychother Psychosom*. 2008;77(6):351-7. Epub: 2008/08/15. PMID: 18701831.
82. David D, Szentagotai A, Lupu V, et al. Rational emotive behavior therapy, cognitive therapy, and medication in the treatment of major depressive disorder: a randomized clinical trial, posttreatment outcomes, and six-month follow-up. *J Clin Psychol*. 2008 Jun;64(6):728-46. Epub: 2008/05/14. PMID: 18473339.
83. Raue PJ, Schulberg HC, Heo M, et al. Patients' depression treatment preferences and initiation, adherence, and outcome: a randomized primary care study. *Psychiatr Serv*. 2009 Mar;60(3):337-43. Epub: 2009/03/03. PMID: 19252046.
84. Blom MB, Jonker K, Dusseldorp E, et al. Combination treatment for acute depression is superior only when psychotherapy is added to medication. *Psychother Psychosom*. 2007;76(5):289-97. Epub: 2007/08/19. PMID: 17700049.
85. Kennedy SH, Konarski JZ, Segal ZV, et al. Differences in brain glucose metabolism between responders to CBT and venlafaxine in a 16-week randomized controlled trial. *Am J Psychiatry*. 2007 May;164(5):778-88. Epub: 2007/05/04. PMID: 17475737.
86. Dimidjian S, Hollon SD, Dobson KS, et al. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *J Consult Clin Psychol*. 2006 Aug;74(4):658-70. Epub: 2006/08/03. PMID: 16881773.
87. Segal ZV, Kennedy S, Gemar M, et al. Cognitive reactivity to sad mood provocation and the prediction of depressive relapse. *Arch Gen Psychiatry*. 2006 Jul;63(7):749-55. Epub: 2006/07/05. PMID: 16818864.
88. DeRubeis RJ, Hollon SD, Amsterdam JD, et al. Cognitive therapy vs medications in the treatment of moderate to severe depression. *Arch Gen Psychiatry*. 2005 Apr;62(4):409-16. Epub: 2005/04/06. PMID: 15809408.
89. Miranda J, Chung JY, Green BL, et al. Treating depression in predominantly low-

- income young minority women: a randomized controlled trial. *JAMA*. 2003 Jul 2;290(1):57-65. Epub: 2003/07/03. PMID: 12837712.
90. Bastos AG, Guimaraes LS, Trentini CM. Neurocognitive changes in depressed patients in psychodynamic psychotherapy, therapy with fluoxetine and combination therapy. *J Affect Disord*. 2013 Dec;151(3):1066-75. PMID: 24103853.
 91. Hegerl U, Hautzinger M, Mergl R, et al. Effects of pharmacotherapy and psychotherapy in depressed primary-care patients: a randomized, controlled trial including a patients' choice arm. *Int J Neuropsychopharmacol*. 2010;13(1):31-44. PMID: 19341510.
 92. Menchetti M, Rucci P, Bortolotti B, et al. Moderators of remission with interpersonal counselling or drug treatment in primary care patients with depression: randomised controlled trial. *Br J Psychiatry*. 2014 Feb;204(2):144-50. PMID: 24311553.
 93. Mischoulon D, Price LH, Carpenter LL, et al. A double-blind, randomized, placebo-controlled clinical trial of S-adenosyl-L-methionine (SAME) versus escitalopram in major depressive disorder. *J Clin Psychiatry*. 2014 Dec 24. Epub: 2014/02/07. PMID: 24500245.
 94. Qu SS, Huang Y, Zhang ZJ, et al. A 6-week randomized controlled trial with 4-week follow-up of acupuncture combined with paroxetine in patients with major depressive disorder. *J Psychiatr Res*. 2013 Jun;47(6):726-32. Epub: 2013/03/19. PMID: 23498306.
 95. Gertsik L, Poland RE, Bresee C, et al. Omega-3 fatty acid augmentation of citalopram treatment for patients with major depressive disorder. *J Clin Psychopharmacol*. 2012 Feb;32(1):61-4. Epub: 2011/12/27. PMID: 22198441.
 96. Sun H, Zhao H, Ma C, et al. Effects of electroacupuncture on depression and the production of glial cell line-derived neurotrophic factor compared with fluoxetine: a randomized controlled pilot study. *J Altern Complement Med*. 2013 Sep;19(9):733-9. PMID: 23647408.
 97. Davidson JRT, Gadde KM, Fairbank JA, et al. Effect of *Hypericum perforatum* (St John's wort) in major depressive disorder: a randomized controlled trial. *J Am Med Assoc*. 2002;287(14):1807-14.
 98. Blumenthal JA, Babyak MA, Moore KA, et al. Effects of exercise training on older patients with major depression. *Arch Intern Med*. 1999 Oct 25;159(19):2349-56. Epub: 1999/11/05. PMID: 10547175.
 99. Blumenthal JA, Babyak MA, Doraiswamy PM, et al. Exercise and pharmacotherapy in the treatment of major depressive disorder. *Psychosom Med*. 2007 Sep-Oct;69(7):587-96. Epub: 2007/09/12. PMID: 17846259.
 100. Thase ME, Friedman ES, Biggs MM, et al. Cognitive therapy versus medication in augmentation and switch strategies as second-step treatments: a STAR*D report. *Am J Psychiatry*. 2007 May;164(5):739-52. Epub: 2007/05/04. PMID: 17475733.
 101. Dekker JJ, Koelen JA, Van HL, et al. Speed of action: the relative efficacy of short psychodynamic supportive psychotherapy and pharmacotherapy in the first 8 weeks of a treatment algorithm for depression. *J Affect Disord*. 2008 Jul;109(1-2):183-8. Epub: 2007/12/07. PMID: 18061276.
 102. Lam RW, Parikh SV, Ramasubbu R, et al. Effects of combined pharmacotherapy and psychotherapy for improving work functioning in major depressive disorder. *Br J Psychiatry*. 2013 Nov;203(5):358-65. PMID: 24029535.
 103. Fava M, Alpert J, Nierenberg AA, et al. A double-blind, randomized trial of St John's wort, fluoxetine, and placebo in major depressive disorder. *J Clin Psychopharmacol*. 2005 Oct;25(5):441-7. Epub: 2005/09/15. PMID: 16160619.
 104. Szegedi A, Kohnen R, Dienel A, et al. Acute treatment of moderate to severe depression with hypericum extract WS 5570 (St John's wort): randomised controlled double blind non-inferiority trial versus paroxetine. *BMJ*. 2005 Mar 5;330(7490):503. Epub: 2005/02/15. PMID: 15708844.
 105. van Gorp G, Meterissian GB, Haiek LN, et al. St John's wort or sertraline? Randomized controlled trial in primary care. *Can Fam Physician*. 2002 May;48:905-12. Epub: 2002/06/11. PMID: 12053635.
 106. Moreno RA, Teng CT, Almeida KM, et al. *Hypericum perforatum* versus fluoxetine in the treatment of mild to moderate depression: a randomized double-blind trial in a Brazilian sample. *Rev Bras Psiquiatr*. 2006 Mar;28(1):29-32. PMID: 16612487.
 107. Brenner R, Azbel V, Madhusoodanan S, et al. Comparison of an extract of hypericum (LI 160) and sertraline in the treatment of

- depression: a double-blind, randomized pilot study. *Clin Ther.* 2000 Apr;22(4):411-9. PMID: 10823363.
108. Schrader E. Equivalence of St John's wort extract (Ze 117) and fluoxetine: a randomized, controlled study in mild-moderate depression. *Int Clin Psychopharmacol.* 2000 Mar;15(2):61-8. PMID: 10759336.
109. Lenox-Smith AJ, Jiang Q. Venlafaxine extended release versus citalopram in patients with depression unresponsive to a selective serotonin reuptake inhibitor. *Int Clin Psychopharmacol.* 2008 May;23(3):113-9. Epub: 2008/04/15. PMID: 18408525.
110. Moradveisi L, Huibers MJ, Renner F, et al. Behavioural activation v. antidepressant medication for treating depression in Iran: randomised trial. *Br J Psychiatry.* 2013 Mar;202(3):204-11. Epub: 2013/02/09. PMID: 23391727.
111. Shamsaei F, Rahimi A, Zarabian MK, et al. Efficacy of pharmacotherapy and cognitive therapy, alone and in combination in major depressive disorder. *Hong Kong J Psychiatr.* 2008;18(2):76-80.
112. Jazayeri S, Tehrani-Doost M, Keshavarz SA, et al. Comparison of therapeutic effects of omega-3 fatty acid eicosapentaenoic acid and fluoxetine, separately and in combination, in major depressive disorder. *Aust N Z J Psychiatry.* 2008 Mar;42(3):192-8. Epub: 2008/02/06. PMID: 18247193.
113. Mynors-Wallis LM, Gath DH, Day A, et al. Randomised controlled trial of problem solving treatment, antidepressant medication, and combined treatment for major depression in primary care. *BMJ.* 2000 Jan 1;320(7226):26-30. Epub: 2000/01/05. PMID: 10617523.
114. Huang Y, Htut W, Li D, et al. Studies on the clinical observation and cerebral glucose metabolism in depression treated by electro-scalp acupuncture compared to fluoxetine. *Int J Clin Acupunct.* 2005;14(1):7-26. PMID: 0077160.
115. Chen JQ, Lin WR, Wang SX, et al. Acupuncture/electroacupuncture enhances antidepressant effect of seroxat: the symptom checklist-90 scores. *Neural Regen Res.* 2014;9(2):213-22.
116. Landenberger NAD. Self-concept and attributional style in the treatment of depression: ProQuest Information & Learning; 2002.
117. Zhang WJ, Yang XB, Zhong BL. Combination of acupuncture and fluoxetine for depression: a randomized, double-blind, sham-controlled trial. *J Altern Complement Med.* 2009 Aug;15(8):837-44. Epub: 2009/08/15. PMID: 19678773.
118. Song Y, Zhou D, Fan J, et al. Effects of electroacupuncture and fluoxetine on the density of GTP-binding-proteins in platelet membrane in patients with major depressive disorder. *J Affect Disord.* 2007 Mar;98(3):253-7. Epub: 2006/08/22. PMID: 16919758.
119. Gastpar M, Singer A, Zeller K. Comparative efficacy and safety of a once-daily dosage of hypericum extract STW3-VI and citalopram in patients with moderate depression: a double-blind, randomised, multicentre, placebo-controlled study. *Pharmacopsychiatry.* 2006 Mar;39(2):66-75. Epub: 2006/03/24. PMID: 16555167.
120. Gastpar M, Singer A, Zeller K. Efficacy and tolerability of hypericum extract STW3 in long-term treatment with a once-daily dosage in comparison with sertraline. *Pharmacopsychiatry.* 2005 Mar;38(2):78-86. PMID: 15744631.
121. Behnke K, Jensen GS, Graubaum HJ, et al. Hypericum perforatum versus fluoxetine in the treatment of mild to moderate depression. *Adv Ther.* 2002 Jan-Feb;19(1):43-52. Epub: 2002/05/15. PMID: 12008860.
122. Harrer G, Schmidt U, Kuhn U, et al. Comparison of equivalence between the St. John's wort extract LoHyp-57 and fluoxetine. *Arzneimittelforschung.* 1999 Apr;49(4):289-96. Epub: 1999/05/25. PMID: 10337446.
123. Bjerkenstedt L, Edman GV, Alken RG, et al. Hypericum extract LI 160 and fluoxetine in mild to moderate depression: a randomized, placebo-controlled multi-center study in outpatients. *Eur Arch Psychiatry Clin Neurosci.* 2005 Feb;255(1):40-7. Epub: 2004/11/13. PMID: 15538592.
124. Hollon SD, DeRubeis RJ, Fawcett J, et al. Effect of cognitive therapy with antidepressant medications vs antidepressants alone on the rate of recovery in major depressive disorder: a randomized clinical trial. *JAMA Psychiatry.* 2014 Oct 1;71(10):1157-64. Epub: 2014/08/22. PMID: 25142196.
125. Sava FA, Yates BT, Lupu V, et al. Cost-effectiveness and cost-utility of cognitive

- therapy, rational emotive behavioral therapy, and fluoxetine (Prozac) in treating depression: a randomized clinical trial. *J Clin Psychol*. 2009 Jan;65(1):36-52. Epub: 2008/12/04. PMID: 19051275.
126. Leykin Y, Amsterdam JD, DeRubeis RJ, et al. Progressive resistance to a selective serotonin reuptake inhibitor but not to cognitive therapy in the treatment of major depression. *J Consult Clin Psychol*. 2007 Apr;75(2):267-76. Epub: 2007/05/02. PMID: 17469884.
127. Rucci P, Frank E, Scocco P, et al. Treatment-emergent suicidal ideation during 4 months of acute management of unipolar major depression with SSRI pharmacotherapy or interpersonal psychotherapy in a randomized clinical trial. *Depress Anxiety*. 2011 Apr;28(4):303-9. Epub: 2011/02/11. PMID: 21308882.
128. Kronstrom K, Salminen JK, Hietala J, et al. Does defense style or psychological mindedness predict treatment response in major depression? *Depress Anxiety*. 2009;26(7):689-95. Epub: 2009/06/06. PMID: 19496102.
129. Papakostas GI, Crawford CM, Scalia MJ, et al. Timing of clinical improvement and symptom resolution in the treatment of major depressive disorder. A replication of findings with the use of a double-blind, placebo-controlled trial of *Hypericum perforatum* versus fluoxetine. *Neuropsychobiology*. 2007;56(2-3):132-7. Epub: 2008/02/09. PMID: 18259086.
130. Babyak M, Blumenthal JA, Herman S, et al. Exercise treatment for major depression: Maintenance of therapeutic benefit. In: Monat A, Lazarus RS, Reevy G, eds. *The Praeger handbook on stress and coping*. Vol. 2. Westport, CT: Praeger Publishers/Greenwood Publishing Group; 2007:529-40.
131. Hoffman BM, Blumenthal JA, Babyak MA, et al. Exercise fails to improve neurocognition in depressed middle-aged and older adults. *Med Sci Sports Exerc*. 2008 Jul;40(7):1344-52. Epub: 2008/06/27. PMID: 18580416.
132. Babyak M, Blumenthal JA, Herman S, et al. Exercise treatment for major depression: maintenance of therapeutic benefit at 10 months. *Psychosom Med*. 2000 Sep-Oct;62(5):633-8. Epub: 2000/10/06. PMID: 11020092.
133. Rush AJ, Trivedi MH, Wisniewski SR, et al. Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *N Engl J Med*. 2006 Mar 23;354(12):1231-42. Epub: 2006/03/24. PMID: 16554525.
134. Trivedi MH, Fava M, Wisniewski SR, et al. Medication augmentation after the failure of SSRIs for depression. *N Engl J Med*. 2006 Mar 23;354(12):1243-52. Epub: 2006/03/24. PMID: 16554526.
135. Rush AJ, Wisniewski SR, Warden D, et al. Selecting among second-step antidepressant medication monotherapies: predictive value of clinical, demographic, or first-step treatment features. *Arch Gen Psychiatry*. 2008 Aug;65(8):870-80. Epub: 2008/08/06. PMID: 18678792.
136. U.S. Department of Health and Human Services, Food and Drug Administration. *Guidance for Industry and Investigators. Safety Reporting Requirements for INDs and BA/BE Studies*. Draft Guidance. Rockville, MD: United States Department of Health and Human Services, Food and Drug Administration; 2010.
137. Zhang ZJ, Chen HY, Yip KC, et al. The effectiveness and safety of acupuncture therapy in depressive disorders: systematic review and meta-analysis. *J Affect Disord*. 2010 Jul;124(1-2):9-21. Epub: 2009/07/28. PMID: 19632725.
138. Cuijpers P, van Straten A, van Oppen P, et al. Are psychological and pharmacologic interventions equally effective in the treatment of adult depressive disorders? A meta-analysis of comparative studies. *J Clin Psychiatry*. 2008 Nov;69(11):1675-85; quiz 839-41. Epub: 2008/10/24. PMID: 18945396.
139. Department of Veteran Affairs, Department of Defense. *VA/DoD clinical practice guideline for management of major depressive disorder (MDD)*. Washington (DC): Department of Veteran Affairs, Department of Defense; May 2009. http://www.healthquality.va.gov/mdd/mdd_full09_c.pdf.
140. Williams JW, Gierisch JM, McDuffie J, et al. *An Overview of Complementary and Alternative Medicine Therapies for Anxiety and Depressive Disorders: Supplement to Efficacy of Complementary and Alternative Medicine Therapies for Posttraumatic Stress Disorder*. Washington DC; 2011.
141. Papakostas GI, Cassiello CF, Iovieno N. Folates and S-adenosylmethionine for major

- depressive disorder. *Can J Psychiatry*. 2012 Jul;57(7):406-13. Epub: 2012/07/06. PMID: 22762295.
142. Rahimi R, Nikfar S, Abdollahi M. Efficacy and tolerability of *Hypericum perforatum* in major depressive disorder in comparison with selective serotonin reuptake inhibitors: a meta-analysis. *Prog Neuropsychopharmacol Biol Psychiatry*. 2009 Feb 1;33(1):118-27. Epub: 2008/11/26. PMID: 19028540.
143. Linde K, Berner Michael M, Kriston L. St John's wort for major depression. *Cochrane Database Syst Rev*. 2008(4)PMID: CD000448.
144. Appleton KM, Perry R, Sallis H, et al. Omega-3 fatty acids for depression in adults (Protocol). *Cochrane Database Syst Rev*. 2014(5).
145. Hardy ML, Coulter I, Morton SC, et al. S-adenosyl-L-methionine for treatment of depression, osteoarthritis, and liver disease. Evidence Report/Technology Assessment No. 64 AHRQ Publication No. 02-E034. Rockville, MD: Agency for Healthcare Quality and Research; August 2002.
146. Robertson R, Robertson A, Jepson R, et al. Walking for depression or depressive symptoms: A systematic review and meta-analysis. *Mental Health and Physical Activity*. 2012;5(1):66-75.
147. Krogh J, Nordentoft M, Sterne JAC, et al. The effect of exercise in clinically depressed adults: Systematic review and meta-analysis of randomized controlled trials. *J Clin Psychiatry*. 2011;72(4):529-38. PMID: 21034688.
148. Schuch FB, Vasconcelos-Moreno MP, Fleck MP. The impact of exercise on quality of life within exercise and depression trials: A systematic review. *Mental Health and Physical Activity*. 2011;4(2):43-8.
149. Daley A. Exercise and depression: A review of reviews. *J Clin Psychol Med Settings*. 2008;15(2):140-7. PMID: 19104978.
150. Mura G, Moro MF, Patten SB, et al. Exercise as an add-on strategy for the treatment of major depressive disorder: a systematic review. *CNS Spectr*. 2014 Mar 3:1-13. Epub: 2014/03/05. PMID: 24589012.
151. Brenes GA, Williamson JD, Messier SP, et al. Treatment of minor depression in older adults: a pilot study comparing sertraline and exercise. *Aging Ment Health*. 2007 Jan;11(1):61-8. Epub: 2006/12/14. PMID: 17164159.
152. Linde K, Witt CM, Streng A, et al. The impact of patient expectations on outcomes in four randomized controlled trials of acupuncture in patients with chronic pain. *Pain*. 2007 Apr;128(3):264-71. Epub: 2007/01/30. PMID: 17257756.
153. Bradley PR, Association BHM. *British Herbal Compendium: A Handbook of Scientific Information on Widely Used Plant Drugs*. British Herbal Medicine Association; 2006.
154. Russo E, Scicchitano F, Whalley BJ, et al. *Hypericum perforatum*: pharmacokinetic, mechanism of action, tolerability, and clinical drug-drug interactions. *Phytother Res*. 2014 May;28(5):643-55. Epub: 2013/07/31. PMID: 23897801.
155. Goodwin RD. Association between physical activity and mental disorders among adults in the United States. *Prev Med*. 2003;36(6):698-703. PMID: 12744913
156. Xing S, Dipaula BA, Lee HY, et al. Failure to fill electronically prescribed antidepressant medications: a retrospective study. *Prim Care Companion CNS Disord*. 2011;13(1)Epub: 2011/07/07. PMID: 21731832.
157. Trivedi MH, Greer TL, Church TS, et al. Exercise as an augmentation treatment for nonremitted major depressive disorder: A randomized, parallel dose comparison. *J Clin Psychiatry*. 2011;72(5):677-84.
158. Mancini M, Wade AG, Perugi G, et al. Impact of patient selection and study characteristics on signal detection in placebo-controlled trials with antidepressants. *J Psychiatr Res*. 2014;51:21-9. PMID: 24462042
159. Melartin TK, Rytsala HJ, Leskela US, et al. Continuity is the main challenge in treating major depressive disorder in psychiatric care. *J Clin Psychiatry*. 2005 Feb;66(2):220-7. Epub: 2005/02/12. PMID: 15705008.