*Comparative Effectiveness Review*
**Number XX**

# Nonpharmacologic Interventions for Treatment-Resistant Depression in Adults

**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

This information is distributed solely for the purposes of pre-dissemination peer review under applicable information quality guidelines. It has not been formally disseminated by the Agency for Healthcare Research and Quality. 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 determination or policy.

**Contract No.**

**Prepared by**:

*Investigators:*

**AHRQ Publication No. 0X-EXXX-02**
**Publication Month and Year XXX**

**None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report.**

This report is based on research conducted by the XXX (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. XXX). The findings and conclusions in this document are those of the author(s), 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.

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

# Preface

The Agency for Healthcare Research and Quality (AHRQ) conducts the Effective Health Care Program as part of its mission to organize knowledge and make it available to inform decisions about health care. As part of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Congress directed AHRQ to conduct and support research on the comparative outcomes, clinical effectiveness, and appropriateness of pharmaceuticals, devices, and health care services to meet the needs of Medicare, Medicaid, and the State Children's Health Insurance Program (SCHIP).

AHRQ has an established network of Evidence-based Practice Centers (EPCs) that produce Evidence Reports/Technology Assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care. The EPCs now lend their expertise to the Effective Health Care Program by conducting Comparative Effectiveness Reviews (CERs) of medications, devices, and other relevant interventions, including strategies for how these items and services can best be organized, managed, and delivered.

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 are useful because they define the strengths and limits of the evidence, clarifying whether assertions about the value of the intervention are based on strong evidence from clinical studies. For more information about systematic reviews, see http://effectivehealthcare.ahrq.gov/reference/purpose.cfm

AHRQ expects that CERs will be helpful to health plans, providers, purchasers, government programs, and the health care system as a whole. In addition, AHRQ is committed to presenting information in different formats so that consumers who make decisions about their own and their family's health can benefit from the evidence.

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. Comparative Effectiveness Reviews will be updated regularly.

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Sonia Tyutyulkova, M.D., Ph.D.
Task Order Officer
Evidence-based Practice Center Program
Agency for Healthcare Research and Quality

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

# Acknowledgements

## Technical Expert Panel

# Contents

Executive Summary ..... 1

**Evidence Report**

Chapter 1. Introduction ..... 19
Burden and Costs of Disease ..... 19
Purpose of this Report ..... 20
Scope and Key Questions ..... 25
Organization of the Report ..... 26
Chapter 2. Methods ..... 27
Topic Development ..... 27
Technical Expert Panel ..... 27
Literature Search ..... 28
Study Selection ..... 29
Data Extraction and Analytic Strategy ..... 33
Quality Assessment ..... 37
Applicability Assessment ..... 37
Grading Strength of a Body of Evidence ..... 38
Data Synthesis ..... 38
Peer Review ..... 39
Chapter 3. Results ..... 41
Introduction ..... 41
Key Question 1: Organization of Efficacy and Effectiveness Results ..... 42
Key Question 1a: Nonpharmacologic Interventions–Overview of Head-to-Head Comparisons ..... 44
Strength of Evidence: Considering all Tiers ..... 46
Key Question 1a: Nonpharmacologic Interventions–Key Points of Head-to-Head Comparisons ..... 47
Key Question 1a: Nonpharmacologic Interventions–Detailed Analysis of Head-to-Head Comparisons ..... 48
Strength of Evidence Tier One Only ..... 53
Strength of Evidence: Considering all Tiers ..... 55
Key Question 1a: Efficacy or Effectiveness of Nonpharmacologic Interventions for Acute Phase Treatment—Key Points of Active vs. Control Comparisons ..... 59
rTMS vs. Sham ..... 59
VNS vs. Sham ..... 60
Psychotherapy vs. Control ..... 60
Key Question 1a: Efficacy or Effectiveness of Nonpharmacologic Interventions for Acute Phase Treatment—Detailed Analysis of Active vs. Control Comparisons ..... 61
rTMS vs. Sham ..... 61
Vagus Nerve Stimulation Plus Treatment as Usual vs. Treatment as Usual ..... 80
Psychotherapy Plus Treatment as Usual vs. Treatment as Usual ..... 81

Key Question 1b: Comparisons Involving Pharmacologic Interventions for Acute Phase Treatment—Overview of Comparisons ..... 83
Key Question 1b: Comparisons Involving Pharmacologic Interventions for Acute Phase Treatment—Overview of Nonpharmacologic vs. Pharmacologic Treatments ..... 84
Strength of Evidence: Tier 1 Only ..... 84
Strength of Evidence: Considering all Tiers ..... 85
Key Question 1b: Comparisons Involving Pharmacologic Interventions for Acute Phase Treatment—Key Points of Nonpharmacologic vs. Pharmacologic Treatments ..... 85
ECT vs. Pharmacotherapy ..... 87
rTMS Plus Pharmacotherapy vs. Pharmacotherapy ..... 87
Key Question 1b: Comparisons Involving Pharmacologic Interventions for Acute Phase Treatment—Detailed Analysis of Nonpharmacologic vs. Pharmacologic Treatments ..... 87
Repetitive Transcranial Magnetic Stimulation Plus Pharmacotherapy vs. Sham Control Plus Pharmacotherapy ..... 88
CBT vs. Pharmacotherapy ..... 89
Key Question 1b: Pharmacologic Interventions for Acute Phase Treatment—Overview of Pharmacologic vs. Pharmacologic Treatments ..... 91
Key Question 1b: Pharmacologic Interventions for Acute Phase Treatment—Key Points of Direct Comparisons ..... 92
Key Question 1b: Pharmacologic Interventions for Acute Phase Treatment—Detailed Analysis of Direct Comparisons ..... 92
Key Question 2: Efficacy or Effectiveness for Maintaining Remission or Treating Patients With Unresponsive or Recurrent Disease: Overview ..... 99
Strength of Evidence Tier One Only ..... 100
Strength of Evidence: Considering all Tiers ..... 100
Key Question 2: Efficacy or Effectiveness for Maintaining Remission or Treating Patients with Unresponsive or Recurrent Disease: Key Points ..... 102
Electroconvulsive Therapy vs. Repetitive Transcranial Magnetic Stimulation ..... 102
Repetitive Transcranial Magnetic Stimulation vs. Sham ..... 102
Cognitive Behavioral Therapy vs. Usual Care ..... 102
Key Question 2: Efficacy or Effectiveness for Maintaining Remission or Treating Patients With Unresponsive or Recurrent Disease: Detailed Analysis ..... 103
Electroconvulsive Therapy vs. Repetitive Transcranial Magnetic Stimulation ..... 103
Repetitive Transcranial Magnetic Stimulation vs. Sham ..... 105
Cognitive Behavioral Therapy vs. Usual Care ..... 110
Key Question 3: Efficacy or Effectiveness for Treating Treatment-Resistant Depression for Particular Symptom Subtypes ..... 111
Overview ..... 111
Strength of Evidence: Tier One Only ..... 112
Strength of Evidence: Considering all Tiers ..... 112
Key Points ..... 112
Detailed Analysis ..... 113
Key Question 4: Organization of Safety, Adverse Events and Adherence ..... 114
Key Question 4a: Cognitive Functioning—Overview ..... 114
Strength of Evidence: Tier 1 Only ..... 115
Strength of Evidence: Considering all Tiers ..... 115

Key Question 4a: Cognitive Functioning—Key Points ..... 116
Key Question 4a: Cognitive functioning—Detailed Analysis ..... 117
ECT vs. rTMS ..... 117
rTMS vs. sham ..... 120
Key Question 4b: Specific Adverse Events—Overview ..... 122
Strength of Evidence: Tier 1 Only ..... 122
Strength of Evidence: Considering all Tiers ..... 123
Key Question 4b:Specific Adverse Events—Key Points ..... 123
Key Question 4b: Specific Adverse Events—Detailed Analysis ..... 124
ECT vs. rTMS ..... 124
rTMS vs. sham ..... 125
VNS vs. sham ..... 127
Key Question 4c: Tolerability as Measured by Withdrawals due to Adverse Events—Overview ..... 127
Strength of Evidence: Tier 1 Only ..... 128
Strength of Evidence: Considering all Tiers ..... 128
Key Question 4c:Tolerability as Measured by Withdrawals due to Adverse Events—Key Points ..... 130
Key Question 4c:Tolerability as Measured by Withdrawals due to Adverse Events—Detailed Analysis ..... 130
ECT vs. rTMS ..... 130
rTMS vs. sham ..... 131
VNS vs. sham ..... 134
CBT vs. Usual Care ..... 134
Key Question 4d:Adherence as Measured by Overall Withdrawals—Overview ..... 134
Strength of Evidence: Tier 1 Only ..... 136
Strength of Evidence: Considering all Tiers ..... 136
Key Question 4d:Adherence as Measured by Overall Withdrawals—Key Points ..... 137
Key Question 4d:Adherence as Measured by Overall Withdrawals—Detailed Analysis ..... 137
Overall Withdrawals: ECT vs. rTMS ..... 137
Overall Withdrawals: rTMS vs. sham ..... 138
Overall Withdrawals: CBT vs. Usual Care ..... 142
Key Question 5: Efficacy and Harms for Selected Populations ..... 143
Overview ..... 143
Key Points ..... 143
Detailed Analysis ..... 144
Key Question 6: Health-related Outcomes—Overview ..... 146
Strength of Evidence: Tier 1 Only ..... 147
Strength of Evidence: Considering all Tiers ..... 148
Key Question 6: Health-Related Outcomes-Key Points ..... 149
Key Question 6: Health-Related Outcomes—Detailed Analysis ..... 149
Repetitive Transcranial Magnetic Stimulation vs. Sham Control ..... 149
Vagus Nerve Stimulation vs. Sham Control ..... 150
Cognitive Behavioral Therapy vs. Control ..... 152
Chapter 4. Discussion ..... 155
Background ..... 155

Overview of Main Findings ..... 156
KQ 5. Efficacy or Harms of Nonpharmacologic Treatments for Selected Patient Subgroups ..... 166
KQ 6. Health-Related Outcomes of Nonpharmacologic Treatments ..... 167
Applicability ..... 167
Limitations of the Evidence Base ..... 167
Limitations of this Review ..... 169
Future Research ..... 169
Conclusion ..... 171
References ..... 173

## Figures

Figure 1. Phases of treatment for major depression with response to initial treatment ..... 19
Figure 2. Phases of treatment for resistant depression (treatment refractory) ..... 20
Figure 3. Analytic framework for nonpharmacologic interventions for treatment-resistant depression ..... 29
Figure 4. PRISMA tree/disposition of articles ..... 42
Figure 5. Mean difference meta-analysis comparing rTMS vs. control on the HAM-D in a MDD only TRD population (Tier 1) ..... 65
Figure 6. Relative risk meta-analysis of response rates comparing rTMS vs. sham in an MDD-only TRD population (Tier 1) ..... 66
Figure 7. Relative risk meta-analysis of response rates comparing rTMS vs. sham in MDD/Bipolar mix TRD population (Tier 1) ..... 69
Figure 8. Effect size meta-analysis comparing rTMS vs. control on the HAM-D in an MDD and MDD/Bipolar mix TRD population (Tier 1) ..... 70
Figure 9. Relative risk meta-analysis of response rates comparing rTMS vs. sham in an MDD and MDD/Bipolar mix TRD population (Tier 1) ..... 70
Figure 10. Relative risk meta-analysis of remission rates comparing rTMS vs. sham in an MDD and MDD/Bipolar mix TRD population (Tier 1) ..... 71
Figure 11. Effect size meta-analysis comparing rTMS vs. control on the HAM-D for TRD-related studies (all Tiers) in MDD-only population ..... 76
Figure 12. Relative risk meta-analysis of response rates comparing rTMS vs. sham for TRD-related studies (all Tiers) in MDD-only population ..... 76
Figure 13. Relative risk meta-analysis of remission rates comparing rTMS vs. sham for TRD-related studies (all Tiers) in MDD-only population ..... 77
Figure 14. Relative risk meta-analysis of response rates comparing rTMS vs. sham for TRD-related studies (all Tiers) in an MDD/Bipolar mix population ..... 77
Figure 15. Relative risk meta-analysis of response rates comparing rTMS vs. sham for TRD-related studies (all Tiers) in MDD and MDD/Bipolar mix populations ..... 78
Figure 16. Relative risk meta-analysis of remission rates comparing rTMS vs. sham for TRD-related studies (all Tiers) in MDD and MDD/Bipolar mix populations ..... 79

# Tables

Table ES-1. Strength of evidence about nonpharmacologic treatment of adult major depressive disorder, by tier ....4

Table 1. Summary of nonpharmacologic interventions ....22
Table 2. Key questions, outcomes, and study eligibility by key question ....30
Table 3. Tiers of evidence pertaining to populations involving varying proportions of treatment-resistant depression ....34
Table 4. Categories of depressive severity ....36
Table 5. Strength of evidence grades and their definitions ....38
Table 6. Key questions about treatment-resistant depression (TRD) ....41
Table 7. Abbreviations and full names of diagnostic scales and other instruments ....43
Table 8. Number of good and fair quality studies by comparison, tier, and diagnostic mix ....44
Table 9. Comparative clinical outcomes in those with two or more failures ECT vs. rTMS for Tier 1 studies; number of studies and subjects, and strength of evidence ....45
Table 10. Comparative clinical outcomes in those with two or more failures ECT vs. ECT plus rTMS for Tier 1 studies; number of studies and subjects, and strength of evidence ....45
Table 11. Comparative clinical outcomes in those with two or more failures ECT vs. rTMS for Tiers 1-3 combined; number of studies and subjects, and strength of evidence ....46
Table 12. Comparative clinical outcomes in those with two or more failures ECT vs. ECT plus rTMS for Tiers 1-3; number of studies and subjects, and strength of evidence ....47
Table 13. ECT vs. rTMS studies—Tiers 1through 3 ....48
Table 14. ECT vs. ECT plus rTMS studies—Tiers 1-3 ....51
Table 15. Number of studies by comparison and tier ....52
Table 16. rTMS vs. sham: strength of evidence—Tier 1 only ....53
Table 17. VNS vs. sham: strength of evidence—Tier 1 only ....55
Table 18. rTMS vs. sham: strength of evidence Tiers 1-3 ....55
Table 19. Psychotherapy vs. control: strength of evidence—Tiers 1-3 ....58
Table 20. rTMS vs. sham, Tier 1 MDD only studies using an augmentation strategy ....61
Table 21. rTMS vs. sham, Tier 1 MDD only using switch and mixed strategies ....64
Table 22. rTMS vs. sham, tier 1 MDD/BP studies ....66
Table 23. rTMS vs. sham, Tier 2 MDD only ....72
Table 24. rTMS vs. sham, Tier 2 MDD/BP ....73
Table 25. rTMS vs. sham, Tier 3 MDD/BP ....74
Table 26. VNS vs. sham ....80
Table 27. Psychotherapy vs. control, Tier 2 studies ....82
Table 28. Number of good and fair quality studies by comparison, tier, and diagnostic mix ....84
Table 29. ECT vs. pharmacotherapy: strength of evidence—Tier 1 only ....85
Table 30. rTMS + pharmacotherapy vs. pharmacotherapy: strength of evidence—Tiers 1-3 ....86
Table 31. Psychotherapy vs. pharmacotherapy: strength of evidence—Tiers 1-3 ....86
Table 32. ECT vs. pharmacotherapy, Tier 1 ....88
Table 33. Tier 2 studies comparing rTMS + pharmacotherapy vs. sham + pharmacotherapy on populations including bipolar disorder ....88
Table 34. Tier 2 studies comparing psychotherapy to pharmacotherapy in patients with major depressive disorder ....90

Table 35. Number of good and fair quality studies by comparison and definition of treatment resistance (tier) ..... 92
Table 36. Tier 1 studies testing switch strategies in patients with TRD ..... 93
Table 37. Tier 1 studies testing augmentation strategies in patients with TRD ..... 97
Table 38. Mean clinical outcomes for TRD (Tier 1) patients in pharmacologic studies ..... 98
Table 39. Number of studies by comparison and definition of treatment resistance (tier) ..... 99
Table 40. Maintenance of remission in TRD (those with two or more failures) for rTMS vs. sham; number of studies and subjects, and strength of the evidence domains ..... 100
Table 41. Maintenance of remission in TRD-related studies (all tiers) by intervention type; number of studies and subjects, and strength of the evidence domains ..... 101
Table 42. Maintenance of remission in patients with major depressive disorder and probable treatment resistance by intervention; ECT vs. rTMS ..... 104
Table 43. Maintenance of remission in patients with major depression or bipolar disorder and probable treatment resistance by intervention; ECT vs. rTMS ..... 105
Table 44. Maintenance of remission in patients with major depressive disorder failing two or more previous treatments by intervention type; rTMS vs. sham ..... 106
Table 45. Maintenance of remission in patients with major depressive disorder failing one or more previous treatments by intervention; rTMS vs. sham ..... 107
Table 46. Maintenance of remission in patients with major depression or bipolar disorder who failed one or more previous treatments by intervention; rTMS vs. sham ..... 109
Table 47. Maintenance of remission in patients with major depression or bipolar disorder who failed one or more previous treatments by intervention; rTMS vs. sham ..... 110
Table 48. Maintenance of remission in patients with major depressive disorder and probable treatment resistance by intervention; CBT vs. usual care ..... 111
Table 49. Impact of on patients with TRD and an accompanying symptom sub-type; number of studies and subjects, and strength of the evidence domains—considering all tiers ..... 112
Table 50. Study assessing response to ECT or rTMS in patients with treatment-resistant depression (major depressive disorder) and psychosis ..... 113
Table 51. Number of good and fair quality studies by TRD tier and diagnostic mix ..... 115
Table 52. Impact of ECT and rTMS on cognitive functioning in patients; number of studies and subjects; strength of evidence limited to Tier 1 ..... 116
Table 53. Impact of ECT and rTMS on cognitive functioning in patients; number of studies and subjects; strength of evidence ..... 116
Table 54. Cognitive functioning results from rTMS studies with ECT in tier one patients with major depressive disorder ..... 118
Table 55. Cognitive functioning results in Tier 3 patients with major depressive disorder and bipolar disorder ..... 119
Table 56. Cognitive functioning results in Tier 1 patients with major depressive disorder ..... 120
Table 57. Cognitive functioning results from rTMS studies with sham comparison in Tier 1 patients with major depressive disorder and bipolar disorder ..... 121
Table 58. Cognitive functioning results from rTMS study with sham comparison in Tier 2 patients with major depressive disorder ..... 122
Table 59. Number of good and fair quality studies by TRD tier and diagnostic mix that measure adverse events systematically ..... 122
Table 60. Specific adverse events studies; strength of evidence limited to Tier 1 ..... 123
Table 61. Specific adverse events studies; number of studies and subjects; strength of evidence 124

Table 62. Adverse events assessed systematically in tier three patients diagnosed with major depressive disorder and bipolar disorder .....125
Table 63. Adverse events assessed systematically comparing rTMS to sham in Tier 1 patients diagnosed with major depressive disorder .....125
Table 64. Adverse events assessed systematically comparing rTMS to sham in Tier 2 patients diagnosed with major depressive disorder and bipolar disorder .....126
Table 65. Adverse events assessed systematically comparing VNS to sham in Tier 1 patients diagnosed with major depressive disorder and bipolar disorder .....127
Table 66. Number of good and fair quality studies by TRD tier and diagnostic mix that assess withdrawals due to adverse events .....128
Table 67. Strength of evidence for withdrawals due to adverse events limited to Tier 1 .....129
Table 68. Strength of evidence for withdrawals due to adverse events considering all tiers .....129
Table 69. Withdrawals due to adverse events comparing ECT to rTMS in Tier 1 patients diagnosed with major depressive disorder .....130
Table 70. Withdrawals due to adverse events comparing ECT to rTMS in Tier 3 patients diagnosed with major depressive disorder and bipolar disorder .....131
Table 71. Withdrawals due to adverse events comparing rTMS to sham in Tier 1 patients diagnosed with major depressive disorder .....132
Table 72. Withdrawals due to adverse events comparing rTMS to sham in Tier 1 patients diagnosed with major depressive disorder and bipolar disorder .....132
Table 73. Withdrawals due to adverse events comparing rTMS to sham in Tier 2 patients diagnosed with major depressive disorder .....133
Table 74. Withdrawals due to adverse events comparing rTMS to sham in Tier 3 patients diagnosed with major depressive disorder and bipolar disorder .....133
Table 75. Withdrawals due to adverse events comparing VNS to sham in Tier 1 patients diagnosed with major depressive disorder and bipolar disorder .....134
Table 76. Withdrawals due to adverse events comparing rTMS to sham in Tier 2 patients diagnosed with major depressive disorder and bipolar disorder .....135
Table 77. Number of good and fair quality studies by TRD tier and diagnostic mix that assess overall withdrawals .....135
Table 78. Strength of evidence for overall withdrawals during treatment limited to Tier 1 .....136
Table 79. Strength of evidence for overall withdrawals during treatment considering all tiers ..136
Table 80. Adherence/compliance .....137
Table 81. Overall withdrawals comparing ECT to rTMS in Tier 1 patients diagnosed with major depressive disorder .....138
Table 82. Overall withdrawals comparing ECT to rTMS in Tier 3 patients diagnosed with major depressive disorder and bipolar disorder .....139
Table 83. Overall withdrawals comparing rTMS to sham in Tier 1 patients diagnosed with major depressive disorder .....139
Table 84. Overall withdrawals comparing rTMS to sham in Tier 1 patients diagnosed with major depressive disorder and bipolar disorder .....140
Table 85. Overall withdrawals comparing rTMS to sham in Tier 2 patients diagnosed with major depressive disorder .....141
Table 86. Overall withdrawals comparing rTMS to sham in Tier 2 patients diagnosed with major depressive disorder and bipolar disorder .....141

Table 87. Overall withdrawals comparing rTMS to sham in Tier 3 patients diagnosed with major depressive disorder and bipolar disorder .....142
Table 88. Overall withdrawals comparing CBT to sham in Tier 2 patients diagnosed with major depressive disorder.....142
Table 89. Overall withdrawals comparing CBT to usual care in Tier 2 patients diagnosed with major depressive disorder and bipolar disorder .....143
Table 90. Comparative clinical outcomes in those with two or more failures rTMS vs. sham for Tiers 1-3 combined; number of studies and subjects, and strength of evidence.....144
Table 91. Studies of efficacy and harms for patient subgroups .....145
Table 92. Number of good and fair quality studies by TRD tier and diagnostic mix.....147
Table 93. Health related outcome measures by comparison for TRD (Tier 1); number of studies and subjects, and strength of the evidence domains .....148
Table 94. Health related outcome measures by comparison for Tiers 1-3 combined; number of studies and subjects, and strength of the evidence domains .....149
Table 95. Tier 1 studies including depressed populations with ≤20 percent of patients with bipolar disorder .....150
Table 96. Tier 2 studies on patients with major depressive disorder.....151
Table 98. Number of studies of head-to-head comparisons of nonpharmacologic treatments, by comparison and level of treatment resistance .....162
Table 99. Number of studies of nonpharmacologic interventions against controls or usual care, by comparison and level of treatment resistance (tier) .....163
Table 100. Number of studies involving pharmacotherapy comparisons, by level of treatment resistance (tier).....164

## Appendixes

Appendix A. List of Peer Reviewers
Appendix B. Exact Search Strings
Appendix C. Data Abstraction Forms and Quality Rating Criteria
Appendix D. Excluded Studies
Appendix E. Evidence Tables
Appendix F. Abbreviations and Full Names of Diagnostic Scales and Other Instruments
Appendix G. Characteristics of Studies with Poor Internal Validity
Appendix H. Articles by Databases Searched

Appendixes are cited in this report and provided electronically at
http://www.ahrq.gov/downloads/pub/evidence/pdf/___/___.pdf.

# Executive Summary

## Background

Major depressive disorder (MDD) is common and costly. Over the course of a year, between 13.1 and 14.2 million people will experience MDD. Approximately half of these people seek help for this condition, and only 20 percent of those receive adequate treatment. For those who do initiate treatment for their depression, approximately 50 percent will not adequately respond following acute-phase treatment, and they are generally considered to have treatment-resistant depression (TRD, sometimes rendered treatment-refractory disease). Patients failing only one prior treatment are sometimes included in this group, but patients failing two or more prior treatment attempts are a particularly important and poorly understood group.

Patients with TRD incur the highest direct and indirect medical costs among those with MDD. These costs increase with the severity of TRD. Treatment-resistant patients are twice as likely to be hospitalized, and their cost of hospitalization is more than six times the mean total costs of depressed patients who are not treatment resistant. Considering both medical and disability claims from an employer's perspective, one study found that TRD employees cost $14,490 per employee per year, whereas the cost for non-TRD employees was $6,665 per employee per year (1996-1998).

Given the burden of TRD generally, the uncertain prognosis of the disorder, and the high costs of therapy, clinicians and patients alike need clear evidence to guide their treatment decisions. The choices are wide-ranging, include both pharmacologic and nonpharmacologic interventions, and are fraught with incomplete, potentially even conflicting, evidence. Somatic treatments, which may involve use of a pharmacologic intervention or a device, are commonly considered for patients with TRD. Antidepressant medications, which are the most commonly used intervention, have decreasing efficacy for producing remission after patients have experienced two failures. Such drugs also often have side effects, sometimes minor but sometimes quite serious. For these reasons, clinicians often look for alternative strategies for their TRD patients.

This review from the RTI-UNC Evidence-based Practice Center (EPC) provides a comprehensive summary of the available data addressing the comparative effectiveness of four nonpharmacologic treatments—electroconvulsive therapy (ECT), repetitive transcranial magnetic stimulation (rTMS), vagus nerve stimulation (VNS), and cognitive behavioral therapy or interpersonal psychotherapy (CBT or IPT)—as therapies for patients with TRD.

The core patient population of interest was patients with MDD who met our definition of TRD: failure to respond following two or more adequate antidepressant trials. We also included TRD studies in which the patient population could include a "mix" of up to 20 percent of patients with bipolar disorder (i.e., 80 percent or more of patients had only MDD), assuming that this small mix would not substantially alter outcomes seen with MDD-alone populations.

The Effective Health Care Program was initiated in 2005 to provide valid evidence about the comparative effectiveness of different medical interventions for treating health problems. The object is to help consumers, health care providers, and others in making informed choices among treatment alternatives. Through its comparative effectiveness reviews, the program supports systematic appraisals of existing scientific evidence regarding treatments for high-priority health conditions. It also promotes and generates new scientific evidence by identifying gaps in existing scientific evidence and supporting new research. The program puts special emphasis on translating findings into a variety of useful formats for different stakeholders, including consumers.

*The full report and this summary are available at* www.effectivehealthcare.ahrq.gov/reports/final.cfm.

We structured our review to maintain our focus on study populations meeting our TRD definition (≥2 antidepressant failures) while not excluding potentially relevant evidence. We identified different tiers of TRD-related studies to use in our analytic strategy:

- **Tier 1** evidence (TRD as defined in this report): studies in which patients specifically had two or more failures of prior treatment with medications;
- **Tier 2** evidence: studies in which patients had one or more failures of prior treatment;
- **Tier 3** evidence: studies in which the number of prior failed treatments was not specified but the clinical situation suggested a high probability of patients having failed two or more prior antidepressant trials; these data have probable relevance to TRD. Studies that did not specify the number of failed treatments but noted that all subjects were referred for ECT were included in this tier.

This comparative effectiveness review is intended to help policymakers and clinicians make informed choices about the use of *nonpharmacologic* interventions for TRD in adults. Our principal goal is to summarize comparative data on the efficacy, effectiveness, and harms of ECT, rTMS, VNS, and CBT/IPT in patients with TRD. Comparisons of these nonpharmacologic therapies are our main interest. However, because treatment decisions made by patients with TRD and their clinicians are not limited to nonpharmacologic options, we also compare nonpharmacologic options with pharmacologic ones. We address the following six key questions (KQ s) as specified by the Agency for Healthcare Research and Quality (AHRQ). "Trials" in these KQ s refers to treatment attempts, not experimental studies.

- KQ 1a**.** For adults with TRD, defined as two or more failed adequate trials of a biologic [i.e., pharmacologic] intervention), do nonpharmacologic interventions such as electroconvulsive therapy (ECT), repetitive transcranial magnetic stimulation (rTMS), vagus nerve stimulation (VNS), or demonstrated effective psychotherapy (e.g., cognitive therapy [CBT or IPT]) differ in efficacy or effectiveness in treating acute-phase depressive symptoms (e.g., response and remission), whether as a single treatment or part of a combination treatment?
- KQ 1b. How do these nonpharmacologic treatments compare with pharmacological treatments in efficacy or effectiveness in treating acute-phase depressive symptoms after two or more failed adequate trials?
- KQ 2. For adults with TRD, do nonpharmacologic interventions differ in their efficacy or effectiveness for maintaining response or remission (e.g., preventing relapse or recurrence) whether as a single treatment or part of a combination treatment?
- KQ 3. Do nonpharmacologic interventions (single or combination) differ in their efficacy or effectiveness for treating TRD as a function of particular symptom subtypes (e.g., catatonic [frozen or hyper] or psychotic symptoms)?
- KQ 4. For adults with TRD, do nonpharmacologic interventions differ in safety, adverse events, or adherence? Adverse effects of interest include but are not limited to amnesia, memory loss, headaches, and postoperative complications.
- KQ 5. How do the efficacy, effectiveness, or harms of treatment with nonpharmacologic treatments for TRD differ for the following subpopulations:
  - Elderly or very elderly patients; other demographic groups (defined by age, ethnic or racial groups, and sex)?

  - Patients with medical comorbidities (e.g., seizure history, stroke, diabetes, dementia, perinatal depression, ischemic heart disease, cancer)?
- KQ 6. For adults with TRD, do nonpharmacologic interventions differ in regard to other health-related outcomes (e.g., quality of life)?

We searched MEDLINE®, Embase, The Cochrane Library, PsycINFO, and the International Pharmaceutical Abstracts. We searched for systematic reviews, clinical controlled trials, meta-analyses, and nonexperimental studies in which the investigator did not assign group allocation. Sources were searched from 1980 through June 2009. AHRQ Scientific Resource Center (SRC) staff contacted device manufacturers and invited them to submit dossiers, including citations. The SRC also provided our EPC with other relevant data that may not have been captured in the literature search.

For efficacy and effectiveness (KQs 1 and 2), we first focused on head-to-head randomized controlled trials (RCTs) comparing one intervention with another. When sufficient head-to-head evidence was unavailable, we evaluated indirect evidence: nonpharmacologic interventions vs. placebo- or sham-controlled evidence or "treatment as usual" controls. For KQs 3, 4, 5, and 6, we examined data from both experimental and observational studies (generally prospective cohort studies). We did not formally distinguish efficacy from effectiveness trials.

We rated the quality of individual studies as good, fair, or poor; only good or fair studies are included in these analyses. We evaluated the strength of the various bodies of evidence using principles stated in the AHRQ *Methods Guide for Comparative Effectiveness Review*, which grades strength as high, moderate, low, or insufficient. We evaluated the applicability of the body of evidence using a qualitative assessment of the population, intervention/treatment, comparator, outcomes measured, timing of followup, and setting.

Throughout this report we synthesized the literature qualitatively. If data were sufficient, we conducted meta-analyses of data for comparisons involving trials that were fairly homogenous in study populations, treatment intervention, and outcome assessments. Given our focus on clear TRD patient populations (Tier 1), we first present Tier 1 data only; we then give information that combines information from studies involving all three tiers of patients and try to point out when findings or strength of evidence change.

## Results: Overview

From a total of 1,982 citations identified, we ultimately included 58 articles in this review; they represent 49 studies. Of these, five were head-to-head RCTs; 43 were sham- or placebo-controlled RCTs; and one was an observational study. We present evidence that allows comparison of the four nonpharmacologic treatments of interest (ECT, rTMS, VNS, and psychotherapy) stratified by tiers of evidence.

Comparative clinical research on nonpharmacologic interventions in a TRD population is early in its infancy. Many clinical questions about efficacy and effectiveness remain unanswered. The text below presents our principal results; a summary table (Table ES-1) documents major comparisons and outcomes for each key question, gives the overall strength of evidence for that comparison, and outlines key findings. We report first on direct evidence (head to head comparisons) and then on indirect evidence (e.g., trials using controls). We do not include topics on which we had no eligible studies across Tiers 1-3; we retain topics on which we had no eligible studies in Tier 1 but at least one study in Tier 2 or 3.

Table ES-1. Strength of evidence about nonpharmacologic treatment of adult treatment resistant major depressive disorder, by tier

| Key Question and Comparison | Outcome | Tier 1 only | | Tiers 1- 3 | |
|---|---|---|---|---|---|
| | | Strength of Evidence[1] | Findings[2] | Strength of Evidence[1] | Findings[2] |
| **Key Question 1a. Comparative efficacy of nonpharmacologic treatments** | | | | | |
| ECT vs. rTMS | Change in depressive severity | Low | 1 fair trial: both ECT and rTMS improved symptom severity but did not differ significantly. | Insufficient | 1 good trial: ECT produced a significantly greater decrease in symptom severity than rTMS<br>2 fair trials: both ECT and rTMS improved symptom severity but did not differ significantly. |
| | Response rate | Low | 1 fair trial: ECT and rTMS did not differ significantly | Insufficient | 1 good trial: ECT produced significantly greater response rates than rTMS<br>2 fair trials: ECT and rTMS did not differ significantly. |
| | Remission rate | Low | 1 fair trial: ECT and rTMS did not differ significantly | Insufficient | 1 good trial: ECT produced significantly greater remission rates than rTMS.<br>2 fair trials: ECT and rTMS did not differ significantly. |
| ECT plus rTMS vs. ECT | Change in depressive severity | Low | 1 fair trial: Both ECT and ECT plus rTMS improved symptom severity but did not differ significantly. | Low | 2 fair trials: Both ECT and ECT plus rTMS improved symptom severity but did not differ significantly. |
| | Response rate | NA | No eligible studies identified | Low | 1 fair trial: ECT and ECT plus rTMS did not differ significantly. |
| | Remission rate | Low | 1 fair trial: ECT and ECT plus rTMS did not differ significantly. | Low | 1 fair trial: ECT and ECT plus rTMS did not differ significantly. |
| rTMS vs. sham | Change in depressive severity | High | 7 trials (3 good, 4 fair) rTMS had a significantly greater decrease in depressive severity than sham.<br>4 fair trials: rTMS had nonsignificantly greater decrease in depressive severity than sham.<br>1 fair trial: rTMS had greater decrease than sham but significance NR. | High | 13 trials (4 good, 9 fair): rTMS had a significantly greater decrease in depressive severity than sham.<br>5 fair trials: rTMS had a nonsignificantly greater decrease in depressive severity than sham<br>1 fair trial: rTMS had greater decrease than sham but significant NR. |

AE, adverse event; CBT, cognitive behavioral therapy; DBT, dialectical behavioral therapy; ECT ,electroconvulsive therapy; KQ , key question; NA, not applicable; NR, not reported; rTMS, repetitive transcranial magnetic stimulation; VNS, vagus nerve stimulation.

Table ES-1. Strength of evidence about nonpharmacologic treatment of adult treatment resistant major depressive disorder, by tier (continued)

| Key Question and Comparison | Outcome | Tier 1 only | | Tiers 1- 3 | |
|---|---|---|---|---|---|
| | | Strength of Evidence[1] | Findings[2] | Strength of Evidence[1] | Findings[2] |
| **Key Question 1a. Comparative efficacy of nonpharmacologic treatments** | | | | | |
| | Response rate | High | 4 trials (3 good, 1 fair): rTMS had a significantly higher response rate than sham.<br>1 fair trial: rTMS had a nonsignificantly higher response rate than sham.<br>4 fair trials: rTMS had a higher response rate than sham, but significance NR. | High | 5 trials (4 good, 1 fair): rTMS had a significantly higher response rate than sham.<br>2 trials (1 good, 1 fair): rTMS had a nonsignificantly higher response rate than sham.<br>5 fair trials: rTMS had a higher response rate than sham, but significance NR.<br>1 fair trial: rTMS and sham had equal response rates. |
| | Remission rate | Moderate | 2 good trials: rTMS had significantly greater remission rate than sham.<br>1 fair trial: rTMS had a greater remission rate than sham but significance NR. | High | 3 good trials: rTMS had a significantly greater remission rate than sham.<br>2 fair trials: rTMS had a greater remission rate than sham but significance NR.<br>1 fair trial: rTMS and sham did not differ significantly.. |
| VNS vs. sham | Change in depressive severity | Low | 1 good trial: VNS and sham did not differ significantly. | Low | No additional studies identified. |
| | Response rate | Low | 1 good trial: VNS and sham did not differ significantly. | Low | No additional studies identified. |
| Psychotherapy vs. control | Change in depressive severity | NA | No eligible studies identified. | Insufficient | 2 fair trials: psychotherapy group had a significantly greater decrease in depressive severity than control group.<br>1 good trial: psychotherapy and control groups did not differ significantly. |
| | Response rate | NA | No eligible studies identified. | Low | 1 fair trial: psychotherapy group had greater response rate than control group (significance NR) |
| | Remission rate | NA | No eligible studies identified. | Low | 1 good trial showed a significantly higher remission rate for psychotherapy than control<br>1 fair trial favored psychotherapy but significance NR. |

Table ES-1. Strength of evidence about nonpharmacologic treatment of adult treatment resistant major depressive disorder, by tier (continued)

| Key Question and Comparison | Outcome | Tier 1 only | | Tiers 1- 3 | |
|---|---|---|---|---|---|
| | | Strength of Evidence[1] | Findings[2] | Strength of Evidence[1] | Findings[2] |
| **KQ 1b. Comparative efficacy of nonpharmacologic and pharmacologic treatments** | | | | | |
| ECT vs. pharmaco-therapy | Change in depressive severity | Low | 1 fair trial: ECT had significantly greater improvement in symptom severity than pharmacotherapy | Low | No additional studies identified. |
| | Response rate | Low | 1 fair trial: ECT had significantly greater response rates than pharmacotherapy | Low | No additional studies identified. |
| Psychotherap y vs. pharmaco-therapy | Change in depressive severity | NA | No eligible studies identified. | Low | 1 fair trial using both switching and augmenting strategies found no significant differences in change in depressive severity between CBT and pharmacotherapy. 1 small fair trial using a mixed strategy and conducting a treatment completer analysis did not report statistical significance |
| | Response rate | NA | No eligible studies identified. | Low | 1 good trial: CBT and pharmacotherapy did not differ significantly in response rate. |
| | Remission rate | NA | No eligible studies identified. | Low | 1 good trial: CBT and pharmacotherapy did not differ significantly in remission rate. |
| rTMS + pharmaco-therapy vs. pharmaco-therapy | Change in depressive severity | NA | No eligible studies identified. | Low | 1 fair trial: rTMS patients starting a new antidepressant had a significantly greater decrease in depressive severity than sham procedure patients and pharmacotherapy (as measured by effect size). |
| **KQ 2. Comparative efficacy for maintaining remission** | | | | | |
| ECT vs. rTMS | Maintenance of remission | NA | No eligible studies identified | Low | 2 fair trials: ECT and rTMS did not differ significantly |
| rTMS vs. sham | Maintenance of remission | Insufficient | 2 fair trials: relapse rates did not significantly differ between rTMS and sham, but too few patients were followed during the continuation phase to allow for meaningful conclusions. | Low | 6 fair trials: rTMS may maintain remission better than sham (depending on length of follow-up); significance NR for 4 trials) |
| CBT vs. usual care | Maintenance of remission | NA | No eligible studies identified. | Low | 1 fair trial: CBT had lower risk of relapse than usual care at all followup points but significance of results varied by how relapse was defined and was not reported for all follow-up measures. |

Table ES-1. Strength of evidence about nonpharmacologic treatment of adult treatment resistant major depressive disorder, by tier (continued)

| Key Question and Comparison | Outcome | Tier 1 only | | Tiers 1- 3 | |
|---|---|---|---|---|---|
| | | Strength of Evidence[1] | Findings[2] | Strength of Evidence[1] | Findings[2] |
| **KQ 3. Comparative efficacy for particular symptom subtypes** | | | | | |
| ECT vs. rTMS | Change in depressive severity | NA | No eligible studies identified. | Insufficient | 1 fair trial: Mixed results. ECT was significantly more effective than rTMS in the overall population and in psychotic patients, but differential use of antipsychotic with ECT (and not rTMS) substantially biased the results; ECT and rTMS were equally effective in nonpsychotic patients. |
| **KQ 4a. Impact of nonpharmacologic interventions on cognitive functioning** | | | | | |
| ECT vs. rTMS | Cognitive functioning | Low | 1 fair trial and 1 fair cohort study: ECT may have deleterious impact on cognitive functioning compared with rTMS (1 study: significant effect on 1-week recall; both studies: nonsignificant effect on all other measures) | Low | 2 trials (1 good, 1 fair) and 2 fair cohort studies: ECT may have deleterious impact on cognitive functioning compared with rTMS (1 study: significant differences in attention and orientation; 1 study: significant difference in acquisition; all other findings: differences nonsignificant or significance NR) |
| rTMS vs. sham | Cognitive functioning | Low | 4 trials (1 good, 3 fair): rTMS may have positive impact on cognitive functioning (2 trials: significant differences in memory, verbal fluency; all other findings nonsignificant or significance not reported). | Low | 5 trials (1 good, 4 fair): No change in findings for Tier 1 alone. |
| **KQ 4b. Specific adverse events** | | | | | |
| ECT vs. rTMS | Adverse events | NA | No eligible studies identified. | Low | 1 good trial: ECT and rTMS did not differ significantly in specific adverse events. |
| rTMS vs. sham | Adverse events | Low | 1 good trial: rTMS resulted in significantly more scalp pain at the stimulation site than sham. | Low | 3 trials (1 good, 2 fair): rTMS had significantly more scalp pain and difficulty starting urination than sham. |
| *VNS vs. sham* | Adverse events | Low | 1 fair trial: VNS and sham did not differ significantly in specific adverse events | Low | No additional studies identified. |

Table ES-1. Strength of evidence about nonpharmacologic treatment of adult major depressive disorder, by tier (continued)

| Key Question and Comparison | Outcome | Tier 1 only | | Tiers 1- 3 | |
|---|---|---|---|---|---|
| | | Strength of Evidence[1] | Findings[2] | Strength of Evidence[1] | Findings[2] |
| **KQ 4c. Withdrawals due to adverse events** | | | | | |
| ECT vs. rTMS | Withdrawals | Low | 1 fair cohort study: no difference in withdrawals between ECT and rTMS groups. | Insufficient | 2 trials (1 good, 1 fair): ECT and rTMS did not differ significantly in withdrawals attributed to adverse events.<br>1 fair cohort study: ECT group had more withdrawals than rTMS group (significance NR) |
| rTMS vs. sham | Withdrawals | Insufficient | 7 trials (1 good, 6 fair): Trials showed mixed results about withdrawals attributed to adverse events. | Insufficient | 11 RCTs (1 good, 10 fair): Trials showed mixed results about withdrawals attributed to adverse events. |
| VNS vs. sham | Withdrawals | Low | 1 good trial: VNS had greater withdrawals attributed to adverse events than sham (significance NR). | Low | No additional studies identified. |
| CBT vs. usual care | Withdrawals | NA | No eligible studies identified | High | 2 RCTs (1 good, 1 fair): CBT had no withdrawals attributed to adverse events. |
| **KQ 4d. Adherence as measured by overall withdrawals** | | | | | |
| ECT vs. rTMS | Overall withdrawals | Insufficient | 1 fair trial and 1 fair cohort study: Studies showed mixed results about withdrawals. | Insufficient | 3 trials (1 good, 2 fair) and 1 cohort study: Studies showed mixed results about withdrawals. |
| rTMS vs. sham | Overall withdrawals | Insufficient | 7 fair trials: Trials showed mixed results about withdrawals. | Insufficient | 13 trials (1 good, 1 fair): Trials showed mixed results about withdrawals. |
| CBT vs. usual care | Overall withdrawals | NA | No eligible studies identified | Insufficient | 3 trials (1 good, 2 fair): Trials showed mixed results about withdrawals. |
| **KQ 5. Efficacy and harms for selected populations** | | | | | |
| rTMS vs. sham | Changes in depressive severity | NA | No eligible studies identified | Low | 3 fair trials: In post-stroke depression, rTMS groups had significantly better depressive severity than sham groups |
| | Response | NA | No eligible studies identified | Low | 3 fair trials: in post-stroke depression, rTMS groups had better response rates than sham groups, but significance varied. |
| | Remission | NA | No eligible studies identified | Low | 3 fair trials: in post-stroke depression, rTMS groups had better remission rates than sham groups, but significance varied. |

Table ES-1. Strength of evidence about nonpharmacologic treatment of adult major depressive disorder, by tier (continued)

| Key Question and Comparison | Outcome | Tier 1 only | | Tiers 1- 3 | |
|---|---|---|---|---|---|
| | | Strength of Evidence[1] | Findings[2] | Strength of Evidence[1] | Findings[2] |
| **KQ 6. Health-related outcomes** | | | | | |
| rTMS vs. sham | Health-related outcomes | Insufficient | 1 fair trial: Mixed results. rTMS (low-right) group had significantly greater improvement in health status and daily functioning than sham group; rTMS (high-left) group had nonsignificantly greater improvement in health status and daily functioning than sham group. | Low | 2 trials (1 good, 1 fair): rTMS groups had significantly greater improvement in health status and daily functioning than sham groups (except for high left rTMS group in one trial, which had nonsignificantly greater improvement in health status and daily functioning than the sham group). |
| VNS vs. sham | Health-related outcomes | Low | 1 fair trial: VNS and sham groups did not differ significantly in daily functioning | Low | No additional studies identified. |
| CBT/DBT vs. control | Health-related outcomes | NA | No eligible studies identified. | Low | 1 fair trial: CBT group had nonsignificantly greater improvement in daily functioning than waitlist control group. |

[1]Strength of evidence is based on the EPC program's modified version of the GRADE system; see text.

[2]Good and fair designations relate to quality ratings for each study.

The greatest volume of evidence is for ECT and rTMS; however, the direct comparative evidence about even these treatments is quite limited. Available indirect evidence primarily involves rTMS; a little information is available on VNS and psychotherapy (chiefly for efficacy and adverse events) and no available indirect evidence involving ECT. Given the limited number of Tier 1 studies and reporting on the number of failed trials, we were unable to stratify our outcomes by the number of failed trials within Tier 1.

## Efficacy of Nonpharmacologic Interventions Against Other Nonpharmacologic Interventions (KQ 1a)

**Direct evidence.** *Tier 1 trials.* The available head-to-head literature concerning the efficacy of the nonpharmacologic interventions for Tier 1 TRD is limited two fair trials (both in MDD-only populations). One compared ECT and rTMS, and the other compared ECT and ECT plus rTMS. They showed, with low strength of evidence, no differences between treatment options for depressive severity, response rates, and remission rates. No trial involved a direct comparison of psychotherapy with another nonpharmacologic intervention.

*Tiers 1-3 trials.* Considering trials for Tiers 1-3 modified our strength of evidence grades (to insufficient) for comparisons of ECT with rTMS; two additional trials produced conflicting results about these outcomes. Considering trials for Tiers 1-3 did not change findings for ECT

vs. ECT plus rTMS; that is, the strength of evidence remained low that two interventions did not differ in these outcomes.

**Indirect evidence.** We identified trials that compared a nonpharmacologic intervention, generally rTMS, VNS, or psychotherapy, with a control or sham procedure. The numbers of these trials with the same or similar control group were very small, so we could not pool them quantitatively. We could, however, assess the potential benefits of nonpharmacologic interventions vs. controls by calculating mean changes in depressive severity, relative risks of response, and relative risks of remission.

*Tier 1 trials.* rTMS was beneficial relative to controls receiving a sham procedure for all three outcomes (severity of depressive symptoms, response rate, remission rate). rTMS produced a greater decrease in depressive severity (high strength of evidence). Specifically, rTMS averaged a decrease in depressive severity measured by the Hamilton Rating Scale for Depression (HAM-D) of more than 5 points relative to sham control, and this change meets the minimum threshold of the 3-point HAM-D difference that is considered clinically meaningful. Response rates were greater with rTMS than sham (also high strength of evidence); those receiving rTMS were more than three times as likely to achieve a depressive response as patients receiving sham procedure. Finally, rTMS was also more likely to produce remission than the control procedure (moderate strength of evidence); patients receiving rTMS were more than 6 times as likely to achieve remission as those receiving the sham.

In the only other Tier 1 comparison, one good-quality VNS vs. sham control trial (a mixed MDD/bipolar population) reported no significant differences between the groups as measured by a change in depressive severity or response rates (low strength of evidence).

*Data from Tiers 1-3*. Considering evidence from trials in all tiers provided data consistent with just Tier 1 findings favoring rTMS over controls. These additional studies raised the strength of evidence grade for better remission rates for rTMS vs. control from moderate to high.

Compared with Tier 1 quantitative syntheses, meta-analyses involving data from studies in all three tiers provided more conservative point estimates and narrower confidence intervals for each outcome, but these analyses reflect what can be expected in Tier 1 populations. For example, rTMS produced a decrease in severity of approximately 5 HAM-D points relative to sham control. Also, patients receiving rTMS were more than two times as likely to achieve response and more than two times as likely to achieve remission than patients receiving sham control.

Finally, considering evidence from all tiers added studies of psychotherapy against controls. Three studies from MDD-only populations provided, overall, low strength of evidence that CBT groups had better outcomes than control groups in response rates and remission rates. Evidence was insufficient concerning any difference between CBT and controls in changing depressive severity.

## Efficacy of Nonpharmacologic Interventions Compared with Antidepressant Pharmacotherapies (KQ 1b)

**Direct evidence.** *Tier 1 trials.* The available head-to-head literature concerning the efficacy of the nonpharmacologic interventions compared with pharmacologic treatment (in this case, paroxetine) for Tier 1 trials is limited to one fair trial (a mixed MDD/bipolar population). ECT produced a significantly greater decrease in depressive severity (9 points by HAM-D) and

significantly better response rates (71 percent vs. 28 percent) than medications (low strength of evidence).

*Tiers 1-3 trials.* No additional trials were identified for ECT.

For psychotherapy versus pharmacotherapy, data from two fair trials (MDD-only patients) suggested that outcomes for the CBT groups did not differ significantly from those for pharmacotherapy control groups for depressive severity or response or remission rates (low strength of evidence).

Finally, a fair trial reported that rTMS plus a new pharmacotherapy produced a nonsignificantly greater decrease in depressive severity than a sham control plus that new pharmacotherapy (low strength of evidence).

**Indirect evidence**. *Tier 1 trials.* Indirect evidence about procedures or psychotherapy (vs. sham or nonpharmaceutical controls) were presented above as part of KQ 1.

We attempted to determine mean changes in depressive severity, relative risks of response, and relative risks of remission for pharmacologic vs. control studies to allow a comparison with similar outcomes in the nonpharmacologic vs. control trials (KQ 1a, indirect). However, there were no comparable, common control groups not receiving a mood-related medication to allow such comparisons.

## Maintenance of Remission or Prevention of Relapse (KQ 2)

**Direct evidence.** With respect to maintaining remission (or preventing relapse), very little comparative evidence was available about the procedure-based interventions or psychotherapy.

*Tier 1 trials.* We found no direct comparisons for ECT, rTMS, VNS, or CBT.

*Tiers 1-3 trials.* For just Tiers 2-3 (in this case), two fair trials compared ECT with rTMS (one with MDD-only patients and one with a mixed MDD/bipolar patient population). The two interventions did not differ significantly in the likelihood of maintaining remission (low strength of evidence).

**Indirect evidence.** Information about interventions vs. control comparisons testing efficacy for remission or relapse was also very limited.

*Tier 1 trials.* Two fair trials compared rTMS with a sham procedure and found no significant differences except at short-term followup. Because very few patients were followed during the relapse prevention phases of these studies, we concluded that this evidence was insufficient to draw meaningful conclusions about rTMS. We had no eligible studies for ECT, VNS, or psychotherapy.

*Tier 1-3 trials*. For rTMS, six fair sham-controlled rTMS trials suggested that rTMS may maintain remission better than a sham procedure (low strength of evidence with little information about statistical significance of differences).

For CBT, one fair trial reported that CBT maintained remission better than usual care, with one measure showing significant benefit and a second measure nonsignificant benefit for CBT vs. sham (low strength of evidence).

## Efficacy of Nonpharmacologic Interventions for Patients with Different Symptomatology (KQ 3)

**Direct evidence**. Very little comparative evidence addressed whether procedure-based treatments differed as a function of symptom subtypes, and we identified no Tier 1 trials.

For Tier 3 (in this case), a fair trial of a mixed MDD/bipolar population reported that patients with psychotic symptoms who received ECT had a significantly greater decrease in depressive severity than those receiving rTMS; however, the finding had a substantial cointervention bias concerning differences in continuation of psychotropic medications (ECT patients were allowed to continue antipsychotics, while rTMS patients were not). Accordingly, we judged strength of evidence to be insufficient.

No comparative evidence was available about psychotherapy in subgroups defined by symptom clusters.

**Indirect evidence.** We identified no studies testing either procedure-based or psychotherapeutic interventions against sham procedures or other controls.

## Safety, Adverse Events and Adherence (KQ 4)

**Direct evidence.** In examining safety, adverse events, and adherence, we found some differences across the interventions in the harms and negative side effects to patients. For just this set of analyses, we examined both trials and cohort studies, and we focus on cognitive functioning, occurrence of specific adverse events, and withdrawals.

*Cognitive functioning*. For Tier 1 studies on cognitive functioning, ECT may have a greater negative (but transient) effect than rTMS within the six month period following treatment. Across Tier 1-3 studies, findings do not change materially. Strength of evidence in both cases is low.

*Specific adverse events.* We identified no Tier 1 studies comparing two interventions on this outcome. Across Tier 1-3 studies, one good trial reported that ECT and rTMS did not differ significantly (low strength of evidence).

*Withdrawals.* We looked at both withdrawals that investigators attributed to adverse events and overall numbers or rates of withdrawals. Evidence for ECT compared with rTMS (Tier 1 but especially Tiers 1-3) generally produced mixed or inconclusive information as to whether ECT had higher rates of withdrawals than rTMS groups (low or insufficient strength of evidence).

**Indirect evidence.** We attempted to include data from the same types of studies and for the same outcomes as for direct evidence. We identified no studies comparing ECT to control.

*Cognitive functioning.* rTMS may have a positive impact on some aspects of cognitive functioning relative to a sham procedure (for four trials in Tier 1); most comparisons are not statistically significant and the strength of evidence is low in all cases. Findings and strength of evidence do not change materially when considering Tier 1-3 studies.

*Specific adverse events.* rTMS groups had significantly more scalp pain at the stimulation site and, when adding in Tier 2-3 trials, more difficulty starting urination than groups receiving sham procedures (low strength of evidence).

*Withdrawals.* Findings were mixed in both Tier 1 and Tier 1-3 studies as to whether rTMS groups had greater rates of withdrawals than groups receiving sham procedures (insufficient evidence).

There was low strength of evidence that there were greater withdrawals due to adverse events in the vagus nerve stimulation group compared to sham.

No Tier 1 studies reported on withdrawals for CBT groups versus those receiving some form of usual care. Some Tier 1-3 studies suggested that CBT patients did not withdraw (because of adverse events) (high strength of evidence); others reported mixed results about whether CBT and usual care patients differed in adherence, as measured by overall withdrawals (insufficient evidence).

## Efficacy or Harms of Nonpharmacologic Treatments for Selected Patient Subgroups (KQ 5)

**Direct evidence.** We found no studies (in any tier) directly comparing nonpharmacologic interventions in selected populations, such as the elderly, those with stroke, or those with other medical comorbidities.

**Indirect evidence.** No Tier 1 trials provided indirect evidence about patient subgroups. However, three fair trials (all Tier 2), assessed whether the effects of nonpharmacologic treatments differed for a subpopulation comprising patients with post-stroke depression. Specifically, compared with groups receiving sham procedure, patients receiving rTMS had significantly better levels of depression severity (low strength of evidence) and were more likely to have better response and remission (but significance varied) (low strength of evidence).

## Health-Related Outcomes of Nonpharmacologic Treatments (KQ 6)

**Direct evidence.** With respect to patient-reported health-related outcomes, we focused on quality of life (various measures) and ability to function in daily life. No trials (either Tier 1 or Tier 1-3) directly compared the effects of nonpharmacologic treatments for these outcomes.

**Indirect evidence.** *Tier 1 trials.* Two trials (both in mixed MDD/bipolar populations) assessed general health status and mental and physical functioning (all health domains related to quality of life). One fair trial on rTMS vs. sham yielded mixed results (insufficient evidence) as to whether the procedures produced greater improvement in health status and daily functioning (as measured by the Global Assessment of Functioning scale). In the other fair trial, VNS and sham groups did not differ significantly in daily functioning (as measured by the 36-item Medical Outcomes Study Short Form [MOS SF-36]). No studies of psychotherapy were identified.

*Tier 1-3 trials.* Two rTMS vs. sham trials (one good, one fair) provided evidence that rTMS groups had significantly greater improvement in the MOS SF-36 and the Quality of Life, Enjoyment and Satisfaction Questionnaire (QLESQ) than controls (raising the strength of evidence grade to low). No additional studies were available for VNS.

The new CBT vs. control trial reported that the psychotherapy group had a nonsignificantly better level of daily functioning as measured by the Lifework-Range of Impaired Functioning Tool than a wait list control group (low strength of evidence).

## Applicability

For the limited amount and low strength of evidence available, the data for Tier 1 (TRD) is generally applicable to TRD populations. Populations enrolled in these trials appeared representative of our target population. With the exception of rTMS, which had varying doses and length of treatment, the studied interventions appeared comparable with those in routine use. Measured outcomes on the whole reflected the most important clinical outcomes for depression measures, although reporting was inconsistent; outcomes for the other key questions were much more restricted. Follow-up periods were generally shorter than desirable, but most were sufficient to measure an initial acute phase treatment response. Study settings were a mixture of inpatient and outpatient, as ECT is generally an inpatient procedure while the others are generally outpatient. The use of inconsistent definitions of TRD in the trials and the absence of

analyses considering the effect of the number of current treatment failures on outcomes hindered interpretation of data, leading to the use of a tiered system. The evidence base combining data for Tiers 1-3 on the whole produced findings that were consistent with Tier 1 TRD data and also appear applicable to TRD populations.

## Remaining Issues

This area of comparative clinical research is early in its infancy. Key areas for future research need primarily to lay more robust foundations for an evidence base which can better inform decisions for clinicians and patients.

**The field needs a standard definition of TRD that investigators should use in their clinical trials research.** Comparison of any of the potential interventions in the field, nonpharmacologic or otherwise, is hampered by the variability in TRD definitions. Although these definitions appear to be converging on a single meaning—two or more treatment failures in the current episode—very few studies of TRD have applied it. Progress in this area of research requires better standardization of this concept, so that future reviews of the evidence do not need to resort to differentiating, as we did, between "tier 1" studies (i.e., TRD by this definition based on two or more treatment failures) and "tier 2 or 3" types of studies. The latter do provide information that helps illuminate likely impacts of these interventions on patients with TRD, but that is not the same thing as having robust studies focused clearly on the patient population of greatest interest.

**More clinical trials comparing nonpharmacologic interventions with other nonpharmacologic options and with pharmacologic treatments are necessary to inform decisionmaking in TRD.** Clinicians, patients, and policymakers need additional relevant data to guide difficult treatment decisions about what to do next: try another medication trial (and should it be an augmentation, switch, or combination strategy?); add (or switch to) rTMS, ECT, VNS, or psychotherapy?

Also, given that treatment options for many TRD patients include medications, trials should directly compare nonpharmacologic interventions with each other *and* with pharmacologic treatments.

**The numbers of failures in current episode should be delineated carefully.** This information, likely more to be accurate than life-time histories of failures, can help investigators determine whether the particular number of failures, or reaching a particular number of failures in a current episode, can help differentiate between nonpharmacologic treatment choices. For example, for patients with two failures in a current episode, the outcomes may not differ between cognitive therapy and rTMS; however, for patients with a different (higher or lower) number of failures in the current episode, one nonpharmacologic treatment may indeed be better than the other. Currently, we do not know what the proper threshold is for selection of treatment. Clarification of the scientific basis for such a decision would substantially improve decisionmaking.

**Determining whether responses differ for patients with MDD and those with bipolar disorder will help to guide future clinical trial design.** Our decision to include trials with patient populations including up to 20 percent with bipolar disorder (i.e, the "mixed" populations noted earlier) was guided by clinical experience and common sense but not by data. Testing to see whether outcomes differ between the two groups can yield information about inclusion criteria (should the mix be 0%, 10%, 20%, etc.?) that may be useful to investigators in designing

TRD trials and may be important to consider as a potential covariate in analyses involving such mixes.

**Greater consideration should be given to the role that the spectrum of depressive severity plays.** Using a finer gradation of depressive severity than investigators now typically employ might identify whether particularly severe degrees of depression, most commonly understood currently as a HAM-D ≥ 20, may respond differently to the available nonpharmacologic interventions than do less severe levels of depression. These gradations may lead clinicians to a better understanding of severe depression and its role in guiding treatment selection in TRD.

**Direct comparisons of treatment strategies, holding consistent any coexisting or concomitant therapies, are imperative.** Decisionmakers need to know whether outcomes with nonpharmacologic treatments are better when such a treatment augments the current treatment, replaces the current treatment, or replaces the current treatment in combination with another treatment. When ongoing treatment is uncontrolled and reflects a variety of treatments—e.g., some patients continue with atypical antipsychotics, some with mood stabilizers, some with no psychotropic medications—results of study studies are difficult, if not impossible, to interpret.

**Consistent reporting of changes in depressive severity, response rates, and remission rates is crucial.** To allow for better comparisons of clinical outcomes in this difficult-to-treat population, all three measures offer useful information for clinicians. Thus, for either trials or observational studies, investigators should attempt to collect data on all three routinely.

**Application of consistent, accepted protocols in trials.** Making sure that patients receive equivalent doses of different nonpharmacologic interventions is more difficult than making sure of this for pharmacologic interventions. Nevertheless, investigators designing trials of nonpharmacologic therapies can attempt to do so by implementing standard accepted protocols for their trials. Such "dosing" is, admittedly, difficult to control when that protocol is in the process of being developed, as with rTMS, but this standardization is a goal well worth trying to reach.

**More careful and consistent assessment of adverse events is required.** Adverse event reporting is quite limited and over a short time span, and what exists is variable and inconsistent. Systematic collection and more consistent reporting of data on harms—i.e., adverse events and negative side effects—and information about attrition and withdrawal would provide useful information to help balance information now focused on clinical benefits. Use of the CONSORT statement (http://www.consort-statement.org/home/), which guides proper reporting of study information (including the presentation of adverse events), would strengthen reporting both harms and other clinical trial findings; it would also aid in the critical appraisal and interpretation of all study results. Further, a more informative assessment of adverse events would require studies to be able to assess long term and cumulative outcomes.

**Including key relevant measures and subgroups in subsequent research is desirable.** As indicated by the review, nearly no evidence exists on how the effectiveness of nonpharmacologic treatments differs (or not) as a function of symptom subtypes or for subgroups defined by sociodemographic characteristic (such as age) or coexisting medical conditions (e.g., post-stroke or post-myocardial infarction depression; perinatal depression). Also essentially missing is information about health-related outcomes, especially those reported by patients, that concern their quality of life or levels of functional impairment. Subsequent studies should focus on employing known, reliable, and valid measures of patient-reported outcomes, such as the MOS SF-36, the QLESQ, and the EQ-5D.

**Including comparisons of newer nonpharmacologic interventions will be important in future research.** As new nonpharmacologic treatments are developed and tested, investigators should try to include them as potential comparators. At the time we started this comparative effectiveness review, clinical trial data on some of the developing nonpharmacologic interventions, such as magnetic seizure therapy or deep brain stimulation, were insufficient (from the published literature) for us to try to include them. As the evidence bases grow to support the efficacy of such additional nonpharmacologic interventions, the newer strategies should be included in comparative effectiveness study designs.

# Conclusion

Our review suggests that comparative clinical research on nonpharmacologic interventions in a TRD population is early in its infancy, and many clinical questions about efficacy and effectiveness remain unanswered. Interpretation of the data is substantially hindered by varying definitions of TRD and the paucity of relevant studies. The greatest volume of evidence is for ECT and rTMS; however, even for the few comparisons of treatments that are supported by some evidence, the strength of evidence is low for benefits. There was low strength of evidence that ECT and rTMS produced a similar likelihood of maintaining remission in TRD-related populations. The limited data available provided a low strength of evidence that ECT produced a greater negative impact on cognitive functioning compared to rTMS. The few trials addressing subpopulations, subtypes, and health-related outcomes provided low or insufficient evidence of differences between nonpharmacologic interventions. The most urgent next steps for research are to apply a consistent definition of TRD, to conduct more head-to-head clinical trials comparing nonpharmacologic interventions to themselves and to pharmacologic treatments, and careful delineation of the number of adequate treatment failures in the current episode.

# Evidence Report

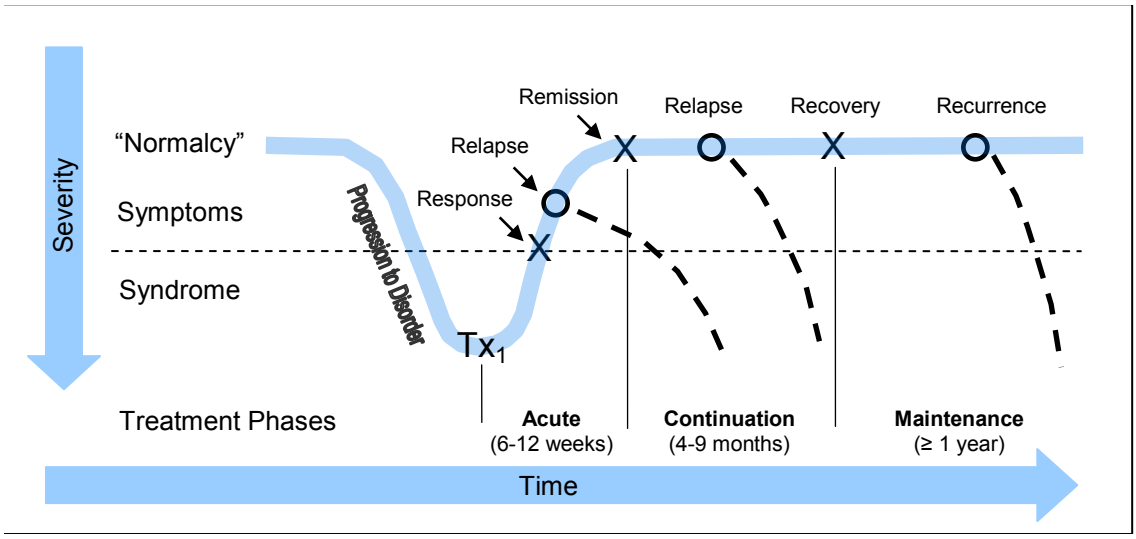
# Chapter 1. Introduction

## Burden and Costs of Disease

Major depressive disorder (MDD) is common and costly. Over the course of a year, between 13.1 and 14.2 million people will experience MDD.[1] Approximately half of these people seek help for this condition, and only 20 percent of those receive adequate treatment.[2]

Among people who do receive adequate treatment, the normal course of treatment consists of an acute phase lasting 6 to 12 weeks with the goal of remission, a complete resolution of the depressive episode (Figure 1). This is followed by a continuation phase of treatment during which the treatment goal is continued absence of depressive symptoms (i.e., relapse prevention) for an additional 4 to 9 months such that the patient's episode can be considered completely resolved. A maintenance phase lasting an additional 1 or more years is recommended in patients who have had two or more previous episodes of depression to prevent the recurrence of a new depressive episode.[3,4]

Figure 1. Phases of treatment for major depression with response to initial treatment



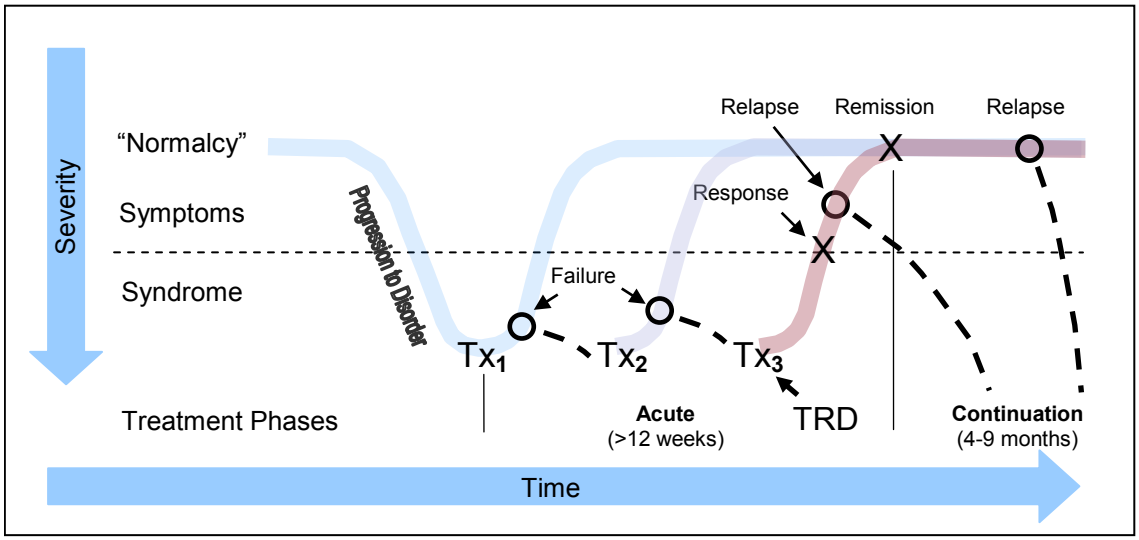

Source: Recreated based on Kupfer, 1991.[5] $Tx_1$ = treatment attempt 1; dashed lines indicate hypothetical worsening of depressive severity, which could indicate failure of treatment, relapse, or recurrence.

Unfortunately, the course of treating patients with depression (especially MDD) often does not follow the idealized treatment phases of reaching, continuing, and maintaining remission as depicted in Figure 1. In the acute phase of treatment, only 30 percent reach the treatment goal of remission. The remaining 70 percent will either obtain response (usually defined as ≥ 50 percent reduction in depressive severity) without remitting (about 20 percent) or not respond at all (50 percent).[6]

This 50 percent of people whose depressive disorder does not adequately respond following acute-phase treatment are generally considered to have treatment-resistant depression (TRD, sometimes rendered treatment-refractory disease).[7] Patients failing only one prior treatment are sometimes included in this group, but patients failing two or more prior treatment attempts are a particularly important and poorly understood group (Figure 2).[8] Indeed, for patients whose depression does not remit after two adequate treatment attempts in the current episode, the likelihood of recovery with subsequent medication treatment decreases by half to approximately

15%.[8] In contrast with Figure 1, which depicts the course of treatment for a patient responding to first-line treatment (i.e., $Tx_1$), treatment resistant patients depicted in Figure 2 require additional treatments (i.e., $Tx_2$, $Tx_3$, or more) and thus have prolonged depressive symptoms during unsuccessful acute phase treatment. Patients failing two or more treatments during the same depressive episode, i.e., those marked TRD as $Tx_3$ in the figure, are also believed to have more resistant disease than patients failing two or more prior treatments during their entire lifetime. The former group of patients seemingly has a more uncertain prognosis for their condition over time than do patients not seen as treatment-resistant (as defined here); by extension, they face longstanding and greater burden of disease.

Figure 2. Phases of treatment for resistant depression (treatment refractory)




Source: Adopted from Kupfer, 1991[5] $Tx_{1-3}$ = Treatment attempt 1, 2, and 3, respectively; TRD = treatment-resistant depression; dashed lines indicate hypothetical worsening of depressive severity, which could indicate failure of treatment, relapse, or recurrence.

Patients with TRD incur the highest direct and indirect medical costs among those with major depressive disorder. These costs increase with the severity of TRD.[9] Treatment-resistant patients are twice as likely to be hospitalized and their cost of hospitalization is more than six times the mean total costs of depressed patients who are not treatment resistant.[10] Considering both medical and disability claims from an employer's perspective, one study found that TRD employees cost $14,490 per employee per year, whereas the cost for non-treatment-resistant depressed employees was $6,665 per employee per year (1996-1998).[11]

## Purpose of this Report

Given the burden of TRD generally, the uncertain prognosis of the disorder, and the high costs of therapy, clinicians and patients alike need clear evidence to guide their treatment decisions. The choices are wide-ranging, include both pharmacologic and nonpharmacologic interventions, and are fraught with incomplete, potentially even conflicting, evidence. Somatic treatments, which may involve use of a pharmacologic intervention or a device, are commonly considered for patients with TRD. Antidepressant medications, which are the most commonly used intervention, have decreasing efficacy for producing remission after patients have experienced two failures. Such drugs also often have side effects,[8] sometimes minor but

sometimes quite serious.[12] For these reasons, clinicians often look for alternative strategies for their TRD patients.

This comparative effectiveness review (CER) is intended to help policymakers and clinicians make informed choices about the use of nonpharmacologic interventions for TRD in adults. Our principal goal is to summarize comparative data on the efficacy, effectiveness, and harms of ECT, rTMS, VNS, and CBT/IPT in patients with TRD. Comparisons between two or more nonpharmacologic interventions are our main interest; however, because patients with TRD and their clinicians often decide between another medication treatment and a nonpharmacologic option, we also compare nonpharmacologic options with pharmacologic ones, both directly and indirectly. The goal is to produce a rough estimate of how these strategies compare for this patient population.

## Included interventions

Nonpharmacologic somatic treatments and nonsomatic psychotherapy treatments offer alternatives to antidepressant medications, although the evidence base for many of these treatments is limited. At the time the protocol for this review was developed, only four types of interventions had a sufficient evidence base establishing their efficacy and were therefore considered appropriate for a Comparative Effectiveness Review. Interventions that offer promising options for patients with TRD include: electroconvulsive therapy (ECT), repetitive transcranial magnetic stimulation (rTMS), vagus nerve stimulation (VNS), and evidence-based psychotherapy (e.g., cognitive therapy, such as cognitive behavioral therapy [CBT] or interpersonal psychotherapy [IPT]). In some cases, these therapies or procedures can be used in combination (e.g., ECT and rTMS). Table 1 provides a summary of these principal nonpharmacologic interventions, including their uses, technical parameters, common side effects, and contraindications. They are described in more detail below. Generally, although these interventions may be safe and effective options for TRD, little evidence exists to guide decisions about their comparative efficacy. Further, how the nonpharmacologic options compare with pharmacologic treatments remains unclear.

## Electroconvulsive Therapy (ECT)

ECT has been available for use in the United States since the 1930s. Current evidence indicates that ECT has a role in the treatment of people with depression and in certain subgroups of people with schizophrenia, catatonia, and mania.[13,14] Its primary role in depression is for treatment resistance or intolerance.[15]

ECT involves passing an electric current through the brain to produce a convulsion. Electrodes are usually placed at the bifrontal, bilateral, or right unilateral position. It is not commonly used as a first-line therapy or in primary care practice. The exceptions are uses in an emergency in which the person's life is at risk because of refusing to eat or drink or being in a catatonic state or in cases of attempted suicide. The effectiveness of ECT may be related to the stimulus parameters used, including position of electrodes, dosage, and waveform of electricity.

ECT is covered by major insurance plans, Medicaid, and Medicare. Reimbursement for each treatment is approximately $275 per treatment.[16] ECT usually consists of 2 to three treatments per week for 3 to 4 weeks.

Table 1. Summary of nonpharmacologic interventions covered in this report

| Major Factors about Nonpharma-cologic Interventions | Electroconvulsive therapy (ECT) | Repetitive Transcranial Magnetic Stimulation (rTMS) | Vagus Nerve Stimulation (VNS) | Cognitive Behavioral Therapy (CBT) or Interpersonal Therapy (IPT) |
|---|---|---|---|---|
| Description | Passing an electric current through the brain after administering anesthetic and muscle relaxants, to produce a convulsion | Focal magnetic stimulation through the scalp without the use of anesthesia[17] | Surgically placed electrodes around the left vagus nerve to modulate mood and seizure control | Psychotherapy to identify negative depressogenic cognitions[18] or interpersonal behaviors[19] |
| Uses | Depression, schizophrenia, catatonia, mania | Depression, mania, anxiety, schizophrenia, epilepsy, Parkinson's disease[20] | Depression, epilepsy | Depression, bipolar disorder, psychosis, anxiety, personality disorders, eating disorders |
| Common Placement Sites | Bifrontal/bilateral or unilateral electrode placement | Dorsolateral prefrontal cortex | Left vagus nerve | Not applicable |
| Average duration | Administered 2 or 3 times a week for 3-4 weeks[21] | 40 minutes daily (usually weekdays) for 2-6 weeks[22] | 30 seconds every 5 minutes, generally for 10 weeks[23] | Weekly sessions for 3-4 months |
| Usual dosage | Millicoulombs of charge[15] | <1-20 Hertz | Current >1 milliamperes (mA), Frequency 1-145 Hertz | Not applicable |
| Contra-indications | Increased risk of complications in patients with unstable cardiac disease, ischemia, arrhythmias, hemorrhage, or increased intracranial pressure[15] | Presence of metal objects anywhere in the body except the mouth such as cardiac pacemakers, medication pumps, cochlear implants. Patients with high risk of seizure. | Bilateral or left cervical vagatomy. Patients with implants should not receive short wave diathermy, microwave diathermy, or ultrasound diathermy | Patients with cognitive disorders, cognitive impairment, or limited cognitive functioning. |

ECT shows greater improvement in patients with suicidal intent than other antidepressant treatments; thus, it may be used as an early therapeutic option in suicidal patients.[24] Research also indicates that despite physical illness, coexisting diseases, and cognitive impairment, older patients tolerate ECT as well as younger patients and may demonstrate better response.[25,26] Because ECT is a procedure that involves anesthesia, it also poses slight risks to patients from the procedure itself. Other potential risks include seizure and adverse cognitive effects.[15]

## Repetitive Transcranial Magnetic Stimulation

rTMS involves magnetic focal stimulation through the scalp. The current elicited by the electromagnetic coil stimulates nerve cells in the region of the brain involved in mood regulation and depression. It can be administered in an office setting without the use of anesthesia. Patients with metal objects anywhere in the body should not undergo this procedure. Sessions are usually 40 minutes in length, administered daily (usually only weekdays) for 2 to 6 weeks. rTMS is

reportedly less expensive than ECT, costing between $100 to $300 per session.[27] It also is perceived by patients as less threatening than ECT.[28] rTMS is usually considered a reasonable option for acute treatment of TRD as opposed to VNS and pharmacotherapy, which are predominantly used as long-term treatments for TRD.[29] The FDA states that rTMS is "indicated for the treatment of Major Depressive Disorder in adult patients who have failed to achieve satisfactory improvement from one prior antidepressant medication at or above the minimal effective dose and duration in the current episode,"[30] an indication that does not specifically address TRD. Possible side effects with rTMS include mild headaches, syncope, and transient hearing changes.[22] rTMS reportedly does not have the seizure or cognitive risks of ECT.[22]

## Vagus Nerve Stimulation

VNS involves surgically placed electrodes around the left vagus nerve. The VNS device consists of a round battery-powered generator that is implanted into the chest wall and attached to wires threaded along the vagus nerve. The therapy includes minor surgery, lasting approximately 30 to 60 minutes. Once implanted, the generator pulses the nerve for 30 seconds once every 5 minutes.[31] The total duration of this intervention is generally 10 weeks, although the stimulation can be extended for longer intervals.[23]

VNS was first used in patients with epilepsy; it was also found simultaneously to improve mood.[32] While the US FDA approved VNS for TRD in July 2005, the Centers for Medicare & Medicaid Services decided not to cover VNS in February 2007, citing lack of evidence.[31] VNS devices cost approximately $10,000 to $20,000, not including the cost of surgery and hospital fees. Although the initial cost of VNS is very high, it may save money for TRD patients in the long run. One study reported long-term savings with VNS compared with usual TRD care, estimating savings of $2,974 and $23,539 per patient per year at 5 and 8 years of device life, respectively.[33]

The place in therapy for VNS may be for patients who have failed four or more adequate antidepressant treatments.[34] Considerations also include a more length onset of antidepressant action than other treatments, as VNS benefits for TRD may not be fully realized for 6 to 12 months.[35] Further, VNS poses surgical risks and is associated with several side effects such as voice alteration, cough, neck pain, paresthesia, and dyspnea.[36]

## Cognitive Behavioral Therapy or Interpersonal Therapy

Use of CBT began in the 1960s. It is a type of psychotherapy that aims to modify distorted, maladaptive, and depressogenic cognitions and related behavioral dysfunction.[18] The therapist first introduces the patient to the cognitive model. Agendas, feedback, and psychoeducational procedures are used to structure sessions. For treating depressed patients with CBT, therapists emphasize negatively distorted thinking and deficits in learning and memory functioning.

IPT developed in the 1970s. IPT helps the patient explore social and interpersonal issues that relate to depressive symptoms. Depressive symptoms identified are related to one of the four key problem areas: grief, disputes, transitions, and deficits.[19] After selecting a focus area, later sessions help the patient develop strategies to deal with the problem.[37]

Both CBT and IPT have been studied extensively for depression, eating disorders, anxiety, and personality disorders, but understanding their role in the treatment of TRD is more limited. Both therapies involve weekly sessions with the therapist, which last for 30 to 60 minutes. CBT may be carried out in a group setting if deemed beneficial for the patient. The therapy generally

lasts between 3 to 4 months for acute phase treatment, although treatment duration may be occur for longer periods. Cost of CBT and IPT depend on the facility and the therapist; on average, these interventions cost around $150 per session.

CBT and IPT do not have any risks or side effects associated with them. Patients need to have normal cognitive functioning to comprehend the therapist's questions. CBT and IPT are comparable psychotherapies for major depression and appear to be as effective as antidepressant medication treatment,[38-40] although CBT may be more effective in patients with severe depression.[37]

## Pharmacologic Interventions

For many patients with TRD, the consideration of another pharmacologic intervention (whether a single agent or combination) remains the next decision step. To place the comparative effectiveness of nonpharmacologic treatments within the context of pharmacologic considerations, we also consider clinical outcomes for a next step pharmacologic treatment based on augmentation and combination mediations commonly used in clinical practice.[41] Given the limited evidence base addressing this topic for TRD, we only consider pharmacologic information for clinical outcomes during acute phase treatment for our main population of interest (see Key Question 1b below).

## Patient Populations Included

**Treatment resistance defined by prior failures.** The primary focus of this review is on patients with MDD who failed two or more prior treatment attempts within the current episode. Definitions of TRD vary considerably and controversially, must often by the number of treatment failures (e.g., one failure, or one or more failures, or two or more failures), whether or not the treatment failures occur during the current episode, and whether or not treatment failures required different classes of antidepressants; no universally accepted definition of TRD currently exists.[7,42-45] This variability is reflected in the differing operational definitions and selection criteria used for TRD trials. Nevertheless, a consensus appears to be forming around a definition of 2 or more treatment failures in the current episode.[42] We view the most applicable evidence to be derived from *patients failing two or more treatment attempts that are of adequate dose and duration during the current depressive episode*. This population represents a homogenous group with known treatment resistance, and we believe these studies are most relevant to our key questions concerning efficacy, effectiveness, safety, and tolerability. However, given the evolving nature of the TRD definition, studies have often not clarified the number of failures within the current episode. Consequently, for the purposes of this report, we will define TRD as **an episode of MDD that has not recovered following two or more adequate antidepressant medication trials, regardless of the class of antidepressant used or whether the treatment failures were required to be in the current episode**.

The variance of the TRD classification makes the interpretation of the available data involving our interventions of interest challenging. Studies addressing TRD and these nonpharmacologic interventions are not always designed with the above specifications in mind. Rather, some studies focus more broadly on the efficacy and/or safety of the interventions in populations of patients with poorly specified characteristics with respect to treatment failures. In particular, they may require patients to have only one previous treatment failure rather than two, or they may be conducted in samples of patients for whom the investigators have not been

completely clear about failures but still give enough information to regard the subjects as "probable" failures (e.g., patients referred for ECT). In such studies, baseline characteristics may provide data indicating that a subset of these patients have failed two or more treatments; however, it is often unclear what proportion of the sample would fit the TRD definition of two or more selected for this report. Although these study populations do not involve homogenous TRD populations, their samples likely include a substantial proportion of TRD patients, and hence can provide data relevant to TRD. Consequently, while we will focus on studies strictly meeting our TRD definition, we will secondarily consider how data from two other groups of studies—those requiring 1 or more treatment failures (which involve patients with only 1 failure as well as those with TRD) and those with probable TRD—may enhance our results.

**Treatment-resistant depression defined for two classes of mood disorder.** Studies of treatment resistance often consider patients with bipolar disorder in addition to patients with major depressive disorder. Our primary focus is evidence about TRD in study patients who clearly have MDD and not any another mood disorder. However, clinical trials of TRD patients frequently allow a mixture of MDD and Bipolar disorder in their samples. Given that depressive episodes in MDD may have a different prognosis than those in bipolar disorder,[46] such a mixture may distort the true effect seen in MDD only patients. At the same time, studies in which a small fraction of the patient population has bipolar disorder rather than purely MDD, are still likely to produce some information on the main topic (i.e., MDD alone). We attempted to select a threshold that would allow inclusion of studies with a proportion of bipolar disease that would not change the likelihood of response. No evidence exists that indicates a proper threshold for such a mixture. After conferring with a Technical Expert Panel, we chose to include trials in our synthesis when the patient population as a whole consists of no more than 20 percent bipolar patients, assuming that such a mix would not substantially alter outcomes from what one would see with MDD alone. The type of bipolar diagnosis could include Type 1 (with manic episodes) or Type 2 (with hypomanic episodes).

## Scope and Key Questions

This review compares the efficacy, effectiveness, and harms of nonpharmacologic interventions for TRD. To that end, we address the following six key questions (KQ s). "Trials" in these KQ s refers to treatment attempts, not experimental studies.

- KQ 1a**.** For adults with treatment-resistant depression (TRD, defined as two or more failed adequate trials of a biologic[*] intervention), do nonpharmacologic interventions such as electroconvulsive therapy (ECT), repetitive transcranial magnetic stimulation (rTMS), vagus nerve stimulation (VNS), or demonstrated effective psychotherapy (e.g., cognitive therapy [CBT or IPT]) differ in efficacy or effectiveness in treating acute-phase depressive symptoms (e.g., response and remission), whether as a single treatment or part of a combination treatment?
- KQ 1b. How do these nonpharmacologic treatments compare with pharmacological treatments in efficacy or effectiveness in treating acute-phase depressive symptoms after two or more failed adequate trials?

[*] biologic: pharmacologic

- KQ 2. For adults with TRD, do nonpharmacologic interventions differ in their efficacy or effectiveness for maintaining response or remission (e.g., preventing relapse or recurrence) whether as a single treatment or part of a combination treatment?
- KQ 3. Do nonpharmacologic interventions (single or combination) differ in their efficacy or effectiveness for treating TRD as a function of particular symptom subtypes (e.g., catatonic [frozen or hyper] or psychotic symptoms)?
- KQ 4. For adults with TRD, do nonpharmacologic interventions differ in safety, adverse events, or adherence? Adverse effects of interest include but are not limited to amnesia, memory loss, headaches, and postoperative complications.
- KQ 5. How do the efficacy, effectiveness, or harms of treatment with nonpharmacologic treatments for TRD differ for the following subpopulations:
  - Elderly or very elderly patients; other demographic groups (defined by age, ethnic or racial groups, and sex)?
  - Patients with medical comorbidities (e.g., seizure history, stroke, diabetes, dementia, perinatal depression, ischemic heart disease, cancer)?
- KQ 6. For adults with TRD, do nonpharmacologic interventions differ in regard to other health-related outcomes (e.g., quality of life)?

## Organization of the Report

The remainder of this report describes our methods, presents the results of our synthesis of the literature, discusses our conclusions and provides other information relevant to the interpretation of this work. Chapter 2 describes our methods in detail. Chapter 3 presents our results for all the key questions and subquestions; it includes summary tables as well. In the discussion (Chapter 4), we summarize the findings, present the strength of evidence for critical comparisons or outcomes, and discuss the implications for practice and further research. A complete list of references is located immediately following the discussion chapter, along with a glossary of terms and a list of abbreviations used throughout this report.

This report also contains the following appendices. Appendix A[†] lists the peer reviewers of the draft of this report. [To be completed following peer review.] Appendix B contains the exact search strings we used in our literature searches. Appendix C documents all the data abstraction forms and our quality rating criteria. Our excluded studies with reasons for exclusion are presented in Appendix D. Evidence tables appear in Appendix E. Appendix F is our table of scales used for measuring neurocognitive and other adverse effects. Appendix G lists poor quality studies and reasons for exclusion from relevant Key Question analyses. Finally, Appendix H lists all of the sources from which we identified all of the studies for this review.

[†] Appendixes are cited in this report and provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/___/___.pdf.

# Chapter 2. Methods

In this chapter, we document the procedures that the RTI International–University of North Carolina Evidence-based Practice Center (RTI–UNC EPC) used to develop this comparative effectiveness review (CER) on nonpharmacologic treatments for adults with treatment-resistant depression (TRD). We briefly describe the topic development process below. We then document our literature search and retrieval process and describe methods of abstracting relevant information from the eligible articles to generate evidence tables. We also document our criteria for rating the quality of individual studies and for grading the strength of the evidence as a whole.

## Topic Development

The topic of this CER and preliminary questions arose through an open process involving the public, the Scientific Resource Center (SRC) for the Effective Health Care program of the Agency for Healthcare Research and Quality (AHRQ) at Oregon Health & Science University, and various stakeholder groups (http://effectivehealthcare.ahrq.gov/index.cfm/who-is-involved-in-the-effective-health-care-program1/about-the-stakeholder-group/). Our EPC was asked to develop provisional "key questions" based on the issues submitted by the nominator of the topic. We conducted a preliminary literature review and worked with key informants to develop a set of provisional "key questions". These key questions were posted by AHRQ for public comment before they were assigned to the RTI-UNC EPC for this full comparative effectiveness review.

## Technical Expert Panel

In designing the study questions and methodology at the topic development stage, we consulted several technical and content experts, seeking broad expertise and perspectives. We worked with seven key informants and all were invited to participate in the Technical Expert Panel (TEP) for the full comparative effectiveness review. Five accepted and in one case a replacement from the consumer organization was made due to the original person no longer being with the organization. In addition, we also invited an expert in psychotherapy, and another psychiatrist conducting a similar evidence review on pharmacotherapy options after one failed treatment, creating a total of eight members (listed in Acknowledgements on page iv). We note that two TEP members had undisclosed conflicts of interest (COI) related to the TMS device. However, upon further inquiry and clarification of the specifics of the form, both individuals filed more complete COI forms.

The TEP provided assistance throughout the project and contributed to AHRQ's broader goals of (1) creating and maintaining science partnerships as well as public-private partnerships and (2) meeting the needs of an array of potential customers and users of its products. Thus, the TEP was both an additional resource and a sounding board during the project.

To ensure robust, scientifically relevant work, we called on the TEP to provide reactions to work in progress and advice on substantive issues or possibly overlooked areas of research. Specifically, TEP members participated in conference calls and discussions through e-mail to:

- review the key questions and the analytic framework, at the beginning of the project;
- discuss the preliminary assessment of the literature, including inclusion/exclusion criteria, including the review of the protocol; and

- provide input on the information and categories included in evidence tables.

Our key questions were posted on AHRQ's Effective Health Care website (http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction = displayproduct&productid = 369) on December 9, 2009. After discussions with the TEP, we added an additional question, KQ 1b, as described in Chapter 1.

# Literature Search

## Databases and Search Terms

To identify articles relevant to each of the six key questions defined in Chapter 1, we searched MEDLINE®, Embase, The Cochrane Library, PsycINFO, and the International Pharmaceutical Abstracts. The full search strategy is presented in Appendix B. We used Medical Subject Headings (MeSH or MH) as search terms when available as well as key words when appropriate. The first step was to locate all articles on depression limiting these to humans, adults, and those published in the English language. We combined terms for treatment-resistant depression including the terms refractory, resistant and drug resistance. The search was further narrowed to specific pharmacological and nonpharmacological treatments. Nonpharmacological interventions included the terms, socioenvironmental therapy, interpersonal psychotherapy, ipt, psychotherapy, cognitive therapy, cognitive behavioral therapy, cbt, electroconvulsive therapy, ect, transcranial magnetic stimulation, (r)tms and vagus nerve stimulation. We searched for systematic reviews, clinical controlled trials, and non-experimental studies in which the investigator did not assign group allocation. Sources were searched from 1980 to June 2009.

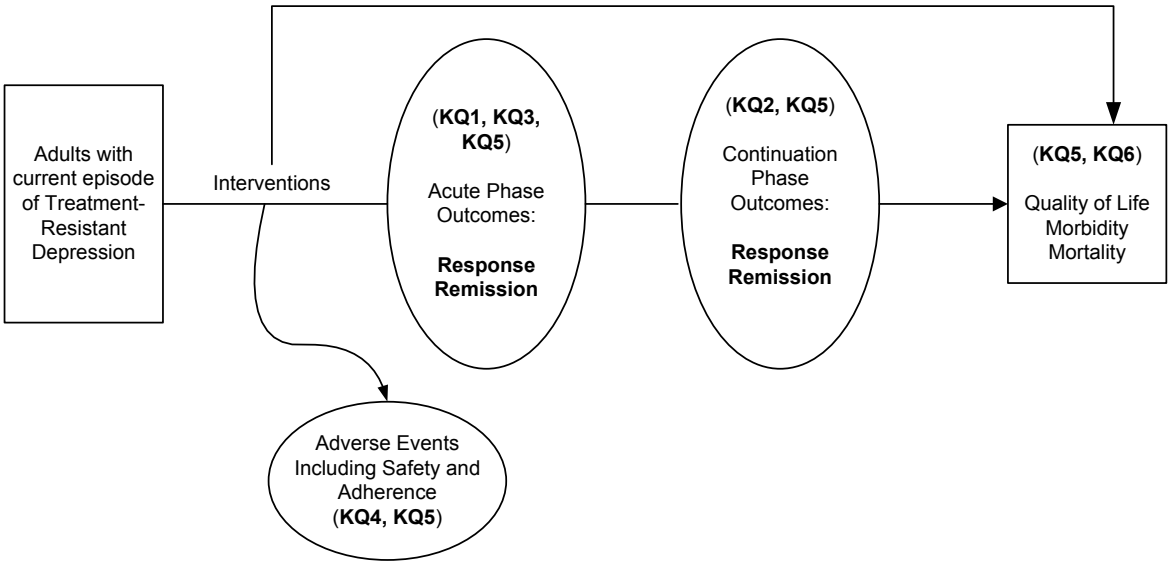
We used the National Library of Medicine (NLM) publication type tags to identify reviews, RCTs, and meta-analyses. We also manually searched reference lists of pertinent review articles and letters to the editor. We imported all citations into an electronic database (EndNote X3). Additionally, we hand-searched the Center for Drug Evaluation and Research (CDER) database to identify unpublished research submitted to the US Food and Drug Administration (FDA).

AHRQ Scientific Resource Center (SRC) staff contacted device manufacturers and invited them to submit dossiers, including citations. We reviewed dossiers received from Cyberonics and Neuronetics. The SRC also provided our EPC with the results of their grey literature search: relevant articles, conference proceedings and meeting abstracts to assist our center to identify other eligible studies that may not have been captured in the literature search.

## Analytic Framework

Based on the six key questions, we developed an analytic framework to guide the systematic review (Figure 3). Specifically, the first two key questions pertain to the efficacy and effectiveness of obtaining (KQ 1) and maintaining (KQ 2) response and remission using these nonpharmacologic treatments; KQ 1 addresses the acute phase of treatment and KQ 2 the continuation or maintenance phases of treatment (as had been depicted in Figure 3). KQ 3 addresses response and remission for psychiatric subtypes of TRD (e.g., coexisting anxiety), and KQ 5 focuses on certain population subgroups (e.g., the elderly). KQ 4 focuses on safety and tolerability issues—i.e., harms—with each of the interventions. Finally, KQ 6 looks at how these interventions affect other health outcomes, such as quality of life.

Figure 3. Analytic framework for nonpharmacologic interventions for treatment-resistant depression




# Study Selection

To summarize, interventions included for one or more of the key questions (KQ s) are:

- Nonpharmacologic therapies, for KQ s 1-6:
  - Electroconvulsive therapy (ECT)
  - Repetitive transcranial magnetic stimulation (rTMS)
  - Vagus nerve stimulation (VNS)
  - Evidence-based psychotherapy -- specifically cognitive therapy (CBT or IPT)
- Pharmacologic,[41] for KQ 1b only, at least one of the antidepressants listed below:
  - Selective serotonin reuptake inhibitors (SSRIs): citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine, sertraline.
  - Serotonin-noepinephrine reuptake inhibitors: desvenlafaxine, duloxetine, mirtazapine, venlafaxine
  - Serotonin modulators: nefazodone and trazodone;
  - Tetracyclic: mirtazapine;
  - Other antidepressants: bupropion;
  - Tricyclic antidepressants (TCAs): e.g., amitriptyline, clomipramine, desipramine, doxepin, imipramine, maprotiline, mianserin, nortriptyline;
  - Monoamine oxidase inhibitors (MAOIs): e.g., phenelzine, tranylcypromine
  - Augmentation strategies with methylphenidate; T4/cytomel; liothyronine; buspirone; lithium or amilsupride; apripazole; olanzapine; quetiapine; risperidone; ziprasidone.

For each key question, we specified inclusion and exclusion criteria for studies and specified the outcome measures of interest; these are listed in Table 2. For efficacy and effectiveness (all KQ s except KQ 4), we first focused on head-to-head randomized controlled trials (RCTs) comparing one intervention with another. This body of work provides direct evidence about the comparisons. When sufficient head-to-head evidence was unavailable, we evaluated placebo- or sham-controlled evidence; in some cases, studies might have used “treatment as usual” as the

control arm. In any of these cases, the evidence provides only indirect evidence. Meta-analyses based on a systematic literature search were eligible for inclusion for each key question. For reviewing adverse events (KQ 4), per our standard approach, we include observational studies. Finally, given the dearth of randomized controlled data our preliminary review suggested was available for KQ 3 on psychiatric subtypes, KQ 5 on subgroups, and KQ 6 on quality of life, for these key questions we included observational studies (limited to prospective and retrospective cohort studies, case control studies). We do not formally distinguish efficacy from effectiveness trials.

Table 2. Key questions, outcomes, and study eligibility by key question

| Key Question and Outcomes | Study Eligibility Criteria (Inclusion and Exclusion Criteria |
|---|---|
| Key Question 1a and 1b<br>**Efficacy and effectiveness**<br>Outcomes<br>• Response<br>• Remission<br>Measurement Scales<br>• Hamilton Rating Scale for Depression Scale (HAM-D)<br>• Montgomery-Åsberg Depression Rating Scale (MADRS)<br>• Beck Depression Inventory (BDI)<br>• Inventory of Depressive Symptomatology<br>• Clinical Global Impression (CGI)<br>• Other relevant scales if none of the above is reported (e.g., Patient Health Questionnaire [HQ-9]) | Study design<br>KQ 1a:<br>• RCTs of nonpharmacologic vs. nonpharmacologic<br>• RCTs of nonpharmacologic vs. pharmacologic (an antidepressant, with or without additional pharmacologic agent[s])<br>• Good or fair quality meta-analyses<br>KQ 1b:<br>• RCTs of nonpharmacologic vs. placebo or sham<br>• RCTs of pharmacologic (an antidepressant, with or without additional pharmacologic agent[s]) vs. placebo or sham<br>• Good or fair quality meta-analyses<br>Minimum study duration<br>• Any duration<br>Sample size<br>• No minimum |
| Key Question 2<br>**Maintenance of response or remission (or prevention of relapse or recurrence)**<br>Outcomes<br>• Relapse (continuation phase)<br>• Recurrence (maintenance phase)<br>Measurement Scales<br>• All efficacy/effectiveness scales listed for KQ 1 above | Study design<br>• RCTs of nonpharmacologic vs. nonpharmacologic<br>• RCTs of nonpharmacologic vs. placebo or sham<br>• RCT designs include continued treatment for prevention or assessment of duration of effect after treatment stopped<br>• Good or fair quality meta-analyses<br>Minimum study duration<br>• ≥ 1 month for relapse prevention<br>• ≥ 3 months for recurrence prevention<br>Sample size<br>• No minimum |
| Key Questions 3<br>**Efficacy and effectiveness by subtype**<br>Outcomes<br>• Response<br>• Remission<br>Measurement Scales<br>• All efficacy/effectiveness scales (see KQ 1 above) | Study design<br>• RCTs of nonpharmacologic vs. nonpharmacologic<br>• RCTs of nonpharmacologic vs. placebo or sham<br>• Good or fair quality meta-analyses<br>• Observational studies (limited to prospective and retrospective cohort studies, case control studies)<br>Minimum study duration<br>• Any duration |

Table 2. Key questions, outcomes, and study eligibility by key question (continued)

| Key Question and Outcomes | Study Eligibility Criteria (Inclusion and Exclusion Criteria |
|---|---|
| Symptom Subtypes<br>• Psychotic-paranoia/hallucinations<br>• Chronic depression –2 years or more<br>• Melancholic<br>• Atypical<br>• Post-partum | Sample size<br>• No minimum |
| Key Questions 4<br>**Safety, adverse events, and adherence**<br>Outcomes<br>• Neurocognitive<br>  ○ Amnesia<br>  ○ Memory loss<br>• Headaches<br>• Postoperative complications<br>• Other reported events<br>• Discontinuations<br>• Adherence/compliance<br>Measurement Scales<br>• All reported adverse events (AE) measurement scales<br>• Discontinuations (overall and attributed to AEs)<br>• Adherence or compliance measures | Study design<br>• RCTs of nonpharmacologic vs. nonpharmacologic<br>• RCTs of nonpharmacologic vs. placebo or sham<br>• Good or fair quality meta-analyses<br>• Observational studies (limited to prospective and retrospective cohort studies, case control studies)<br>Minimum study duration<br>• Any duration<br>Sample size<br>• No minimum, case reports excluded |
| Key Questions 5<br>**Population Subgroups**<br>Outcomes<br>• Response/remission<br>• Relapse/recurrence<br>• Adverse events<br>• Discontinuations<br>Measurement Scales<br>• All reported efficacy/effectiveness scales (see KQ 1 above)<br>• All reported AE measurement scales (see KQ 4above)<br>• Discontinuations and adherence rates<br>Population Subgroups<br>• Age<br>• Medical comorbidity<br>• Race or ethnicity | Study design<br>• RCTs of nonpharmacologic vs. nonpharmacologic<br>• RCTs of nonpharmacologic vs. placebo or sham<br>• Good or fair quality meta-analyses<br>• Observational studies (limited to prospective and retrospective cohort studies, case control studies)<br>Minimum study duration<br>• Any duration<br>Sample size<br>• No minimum, case reports excluded |
| Key Questions 6<br>**Health-Related Outcomes**<br>Outcomes<br>• Quality of life<br>• Satisfaction/enjoyment<br>• Physical or mental functioning<br>• Work productivity or employment<br>Measurement Scales<br>• Global Assessment of Functioning Ability (GAF)<br>• Quality Of Life Enjoyment And Satisfaction Questionnaire (Q-LES-Q) | Study design<br>• RCTs of nonpharmacologic vs. nonpharmacologic<br>• RCTs of nonpharmacologic vs. placebo or sham<br>• Good or fair quality meta-analyses<br>• Observational studies (limited to prospective and retrospective cohort studies, case control studies)<br>Minimum study duration<br>• Any duration<br>Sample size<br>• No minimum |

Table 2. Key questions, outcomes, and study eligibility by key question (continued)

| Key Question and Outcomes | Study Eligibility Criteria (Inclusion and Exclusion Criteria |
|---|---|
| • Medical Outcomes Study Short Form: SF-36, SF-12, or others<br>• Employment/productivity scales<br>• Activities of Daily Living<br>• Other relevant measures | |

Two persons independently reviewed article abstracts using the criteria presented in Appendix C[‡] for Level One. If both reviewers agreed that the study did *not* meet eligibility criteria, we excluded it; otherwise it moved forward to the next step for full-text review, Level Two. We retrieved the full articles for all studies retained at this stage.

Two reviewers then independently reviewed the full-text articles and applied a more detailed set of inclusion criteria; these involved explicit reasons for exclusion, such as wrong design, and specifics on levels of treatment resistance. Appendix C includes copies of all reviewer forms. We resolved conflicts about inclusion at this stage through a consensus with conflicts adjudicated by a third party. Studies excluded at this stage, along with reasons for exclusion, are listed in Appendix D.

For this review, results from well-conducted, valid head-to-head trials—i.e., direct comparisons—provide the strongest evidence to compare treatments with respect to efficacy and harms. The many possible comparisons, set out in Chapter 1, are complex; in some cases, studies compared a treatment with a combination of that treatment and a second intervention. We defined head-to-head trials as those comparing one treatment with another treatment either by itself or in combination with other interventions.

We did not examine placebo-controlled or sham-controlled trials in detail if a sufficient number of head-to-head trials were available. If limited head-to-head evidence was published, we reviewed placebo-controlled trials to provide an overview of efficacy. For harms (i.e., evidence pertaining to tolerability and adverse events), we examined data from both experimental and observational studies.

We did not set any minimum criteria for study duration or sample size, though case reports were excluded when observational study designs were allowed. The exception to this involved relapse and recurrence prevention studies, for which we required at least 1 and 3 months of followup, respectively.

We reviewed studies with health outcomes as primary outcome measures. Outcomes for efficacy or effectiveness, for example, were a decrease in depressive severity, treatment response and remission, quality of life, relapse, functional capacity, and hospitalization. We reviewed response and remission when based on changes in scores on depression scales as proxies for health outcomes (e.g., 50 percent improvement of depression scores for response). For harms, we looked for both overall and specific outcomes related to neurocognitive functioning, specific adverse events including but not limited to amnesia, memory loss, headache, procedure related complications, recorded systematically and spontaneously and tolerability as reflected by withdrawals and withdrawals attributable to adverse events.

[‡] Appendixes are cited in this report and provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/___/___.pdf.

# Data Extraction and Analytic Strategy

We designed and used a structured data abstraction form to ensure consistency of data abstraction and quality appraisal for each study (reproduced in Appendix C). All data abstraction originally employed SRS 4.0 Mobius Analytics™ (www.mobiusanalytics.com/e/index.cfm). Trained reviewers abstracted data from each included study into predesigned evidence tables for each key question; they also assigned an initial quality rating (described below). A senior reviewer read each abstracted article, evaluated the accuracy and completeness of the data abstraction, and independently did a second quality rating. Final evidence tables can be found in Appendix E. [§]

We abstracted data on study design, baseline population characteristics, specifications of the intervention, and relevant outcome assessments for both efficacy and harms. We abstracted data for the efficacy and quality-of-life outcome assessments when the studies used validated measures. We also abstracted data on compliance, attrition, and harms. Finally, we recorded whether analyses were done according to intention-to-treat methods (ITT) if such information was available in the articles. A detailed list of the data elements abstracted is presented in Appendix C.

## Treatment Resistant Depression Definition and Tier Classification

As already noted, the definitions of TRD vary along several dimensions: How many previous treatment failures are considered? What treatments were failed? Were dose and duration of previous treatments adequate? Were the failures during the current episode or over a lifetime? Moreover, the populations included in clinical studies differ by numerous factors. In regard to the variability of the definitions used in studies of TRD, as laid out in Chapter 1, we extracted specific information to create the three-tiered classification system used in presenting results in Chapter 3. We specifically collected data on the study's definition of a failed "trial" (i.e., a treatment in this context). These variables included a specific drug or drug class failed, the specified duration and or dose of an "adequate" trial, the required number of failed trials (whether in the current episode or in a previous episode (termed "lifetime") for inclusion, and baseline characteristics (i.e., mean number of failed trials and other pertinent descriptors) of the sample.

While our working definition of TRD is two or more treatment failures, we realize that many studies involving TRD populations often do not use this definition when formulating their inclusion criteria and that these criteria may not accurately reflect the average number of failed antidepressant trials for a study population. For example, while some studies may require only a single antidepressant failure for a participant to be included in a study, the inclusion criteria may not accurately indicate the average number of antidepressant failures for the study population which could be higher than the cut point set by study inclusion criteria.

When devising the analytic strategy for this report, variation in study inclusion criteria and the overlap in the actual number of antidepressant failures was considered. We considered options, and discussed possible approaches with our TEP, who supported the use of tiered study classification system. We have attempted to maintain our focus on study populations meeting our TRD definition (≥2 antidepressant failures) while not excluding potentially relevant evidence.

---

[§] Appendixes are cited in this report and provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/___/___.pdf.

Our approach to stratifying the literature—into three "tiers"—is highlighted in Table 3. We primarily differentiate studies based on how investigators for included studies defined TRD:

Table 3. Tiers of evidence pertaining to populations involving varying proportions of treatment-resistant depression

| | Relevance to Treatment-Resistant Depression (TRD) per CER protocol | | |
|---|---|---|---|
| **Relevant Study Population** | **Tier 1. TRD per CER protocol (All patients required to have 2 or more treatment failures)** | **Tier 2. All patients required to have 1 or more prior treatment failures** | **Tier 3. Involves those with probable TRD (but number of failures not specified)** |
| MDD alone | All MDD patients who failed ≥ 2 previous treatments | All MDD patients who failed ≥ 1 previous treatment | All MDD patients with TRD not defined |
| Mixed MDD and bipolar disease, with bipolar patients constituting > 0% but ≤ 20% of the study population | MDD/Bipolar mix who failed ≥ 2 previous treatments | MDD/Bipolar mix who failed ≥ 1 previous treatment | MDD/Bipolar mix with TRD not defined |

CER, comparative effectiveness review; MDD, major depressive disorder; TRD, treatment-resistant depression.

- **Tier 1** evidence: involves studies requiring failure to recover following 2 or more adequate antidepressant treatment trials (Tier 1, our working definition of TRD).
- **Tier 2** evidence: involves studies requiring patients to have one or more failed adequate antidepressant treatment trials, so may include both those with only 1 prior treatment failure in addition to those with two or more failed trials. By virtue of including those with only one failure, on average this group has an overall lesser degree of treatment resistance than TRD patients (Tier 1).
- **Tier 3** evidence: involves studies where the number of prior failed treatments was not specified but the clinical situation suggested a high probability of patients having failed two or more prior antidepressant trials; this data has probable relevance to TRD. For example, an included study may refer to TRD without characterizing it, or the clinical presentation strongly suggests two or more prior treatment failures. Studies which did not specify the number of failed treatments but noted that all subjects were referred for ECT were included in this tier.

## Psychiatric Diagnosis

Also, as described in Chapter 1, we included study populations of patients with Major Depressive Disorder (MDD) and study population that include a small number of patients with bipolar disorder. We explicitly extracted data regarding the psychiatric diagnosis—i.e., major depressive disorder (MDD) or bipolar disorder—to allow us to limit the percent of patients with a bipolar TRD to ≤ 20%, a proportion that we determined would be unlikely to influence the outcomes form what was expected for an MDD TRD population. If the study clarified whether the included Bipolar patients were Type 1 (with manic episodes) or Type 2 (with hypomanic episodes), we collected this information.

## Nonpharmacologic Intervention Treatment Characteristics

During data abstraction, characteristics of each mode of nonpharmalogical intervention which impacted treatment dose or intensity were collected and used in our analytic approach. Parameter variables were unique for the each mode of intervention. For ECT, data were collected on the location of the stimuli (e.g. unilateral/bilateral), treatment intensity (e.g. as a function of seizure threshold), number of treatments per week, the mean number of treatment sessions. In the results chapter, ECT implementation for an intervention group is described using the proportion receiving bilateral stimulation and the mean number of treatment sessions received with additional treatment description parameters listed in the evidence tables (Appendix E).

For rTMS, data were abstracted on the location of stimuli (e.g. left or right dorsolateral prefrontal cortex), frequency (e.g. Hertz) and intensity (e.g. as a function of motor threshold) of the stimuli, stimuli or pulses per session (abbreviated " pps"), total number of sessions, and duration of treatment (in weeks). These variables were not always presented in this fashion within our included studies. The following formula was used to calculate pps when the number of treatments per week was not explicitly provided: frequency (Hz) * duration of each train (seconds) * number of trains = pps.[20]

A range of treatment parameters for both active and sham stimulations are utilized in rTMS efficacy studies. In the treatment of depression, stimuli are most often applied at either a high frequency (>1 Hz) to the left or low frequency (≤1 Hz) to the right dorsolateral prefrontal cortex.[20] To simplify reporting in the results chapter, the location of stimulation and frequency is specified only studies deviating from these conventions. All other interventions are described as either high rTMS or low rTMS with complete descriptions of all rTMS stimulation parameters as provided in individual studies are reported in the evidence tables (Appendix E).[**]

Some methods of sham rTMS stimulation have been shown to have a smaller, but noteworthy amount of active stimulation.[47,48] If an included study used one of these methods of sham stimulation, investigators assessed the possibility this impacted the results of the study with potential issues acknowledged in the description of the results. Full descriptions of all sham stimulation parameters are found in the evidence tables (Appendix E).

For VNS, data were collected on the frequency (Hz), pulse width (in seconds), on/off cycle schedule, and duration of treatment. Only treatment parameters outside of the standard range are described in the results with full intervention methodology, including sham stimulation procedures, presented in the evidence tables (Appendix E).

Lastly, for psychotherapeutic interventions, data were collected on the method of therapy implementation (i.e. individual or group therapy), content of the curriculum (e.g. cognitive-based therapy), intensity of the treatment (in sessions per week), total number of sessions, and treatment duration (in weeks). Psychotherapeutic interventions are defined by curriculum content in the results; other parameters are reported in the evidence tables (Appendix E).

## Antidepressant Medication Treatment Strategy

In addition to the nonpharmacologic interventions utilized in studies, investigators used different strategies for managing patients' antidepressant pharmacotherapy which included antidepressants and augmenting agents such as antipsychotics and mood stabilizers. All included

---

[**] Appendixes are cited in this report and provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/___/___.pdf.

studies were categorized into one of five groups according to how the antidepressant pharmacotherapy is addressed as part of a study design. Anti-anxiety medications were allowed by some studies; however these medications were not assessed as part of the antidepressant strategy categorization as there is no evidence basis supporting their benefit as an augmentation agent.

*Switch* studies are those which all patients discontinued their prior antidepressant treatment prior to initiating their next step treatment. Other studies allowed patients to continue their prior antidepressant pharmacotherapy and initiated next step treatment as an add-on or augmentation to their current treatment; these treatment strategies were termed *augmentation* strategies. In some augmentation studies, a small proportion of patients were not taking any psychotropic medications before or during the trial. The inclusion of such patients is acknowledged in the study description.

A third set of studies used both switch and augmentation strategies and were categorized as *mixed*. Two types of mixed studies exist in the included literature. One group of studies encourages but does not require patients to discontinue their antidepressant medications resulting in a study population that contains both switchers and augmenters in all study groups. These studies allow different antidepressant medication strategies within research groups and are called mixed-*within.* Other studies compare patients who *switch* to patients who *augment;* these studies use a mixed antidepressant medication strategy with between group differences and are called *mixed-between*.

In another subset of studies, all patients initiated a new psychotropic medication at the same time in which active groups began the nonpharmacologic intervention. This strategy was termed *combination treatment*. Lastly, in a small group of studies, medications were not limited or initiated by the study (e.g. patients sought treatment as usual which allowed them to change medications or continue the same regimen at the discretion of their treating doctor). This group of studies was described as having an *unlimited* psychotropic medication strategy. A small number of studies allowed (or disallowed) antidepressant medications and potential augmenting agents differently (e.g. antidepressants were discontinued while patients were allowed to continue antipsychotics); pharmacologic strategies of these studies are described in the text and summary tables. Details of each studies antidepressant medication strategy are provided in the evidence tables (Appendix E).[††]

## Disease Severity

Lastly, to enable us to examine differences based on disease severity, we grouped baseline scores into three categories: none to mild, moderate, and severe to very severe (Table 4).[49]

Table 4. Categories of depressive severity

| | 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 |

[††] Appendixes are cited in this report and provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/___/___.pdf.

## Quality Assessment

To assess the quality (internal validity or risk of bias) of all included studies, we used predefined criteria based on those described in the AHRQ *Methods Guide for Comparative Effectiveness Reviews* (ratings: good, fair, poor).[50] Two independent reviewers assigned quality ratings. They resolved any disagreements by discussion and consensus or by consulting with a third reviewer.

Elements of quality assessment for trials included, among others, the methods used for randomization, allocation concealment, and blinding; the similarity of compared groups at baseline; maintenance of comparable groups; overall and differential loss to followup; and the use of ITT analysis. We assessed observational studies based on the potential for selection bias (methods of selection of subjects and loss to followup), potential for measurement bias (equality, validity, and reliability of ascertainment of outcomes), adjustment for potential confounders, and statistical analysis.

In general terms, a "good" study has the least bias and results are considered to be valid. We rated studies that met all criteria as good quality. Fair studies presumably fulfilled all quality criteria but did not report their methods to an extent that answered all our questions. A "fair" study is susceptible to some bias but probably not sufficient to invalidate its results. The fair-quality category is likely to be broad, so studies with this rating will vary in their strengths and weaknesses. A "poor" rating indicates significant bias (stemming from, e.g., serious errors in design, analysis reporting large amounts of missing information, or discrepancies in reporting) that may invalidate the study's results. Studies that had a fatal flaw (defined as a methodological shortcoming that leads to a high probability of bias) in one or more categories were rated poor quality.

Poor-quality studies and reasons for that rating are presented in Appendix F.[‡‡] In this CER, we excluded poor quality studies from our analyses if there were enough good or fair studies with significant findings. In some cases, a poor study may offer the only pertinent information about an important outcome or comparison, and we may comment on it in the relevant section of the results (Chapter 3), but it will not be included in summary tables there.

## Applicability Assessment

Using the parameters for evaluation on guidance provided by AHRQ's *Methods Guide for Comparative Effectiveness Reviews,*[50] we evaluated the applicability of the studies included and evaluated in this CER. Applicability is essentially the generalizability or external validity of the studies included in the evidence base. We evaluated applicability using a qualitative assessment of the population, intervention/treatment, comparator, outcomes measured, timing of followup, and setting. We specifically considered whether populations enrolled in these trials or studies differed from target populations as laid out above, whether studied interventions are comparable with those in routine use, whether comparators reflect best alternatives, whether measured outcomes reflect the most important clinical outcomes, whether followup was sufficient, and whether study settings were representative of most settings.

---

[‡‡] Appendixes are cited in this report and provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/___/___.pdf.

## Grading Strength of a Body of Evidence

We evaluated the strength of evidence based on the AHRQ *Methods Guide for Comparative Effectiveness Reviews*.[50] Strength of evidence is graded only for major comparisons and major outcomes for the topic at hand. The strength of evidence for each outcome or comparison that we graded incorporates scores on four mandatory domains: risk of bias, consistency, directness, and precision; it can also reflect ratings for other domains that can be factored in when relevant (e.g., dose-response relationships). As described in Owens et al., the evaluation of risk of bias includes assessment of study design and aggregate quality of studies.[50] We judged good quality studies with strong designs to result in evidence with low risk of bias. We graded evidence as consistent when effect sizes across studies were in the same direction and had a narrow range. When the evidence linked the interventions directly to health outcomes, we graded the evidence as being direct. For active versus sham control comparisons, we graded the evidence as direct for general efficacy, which should not be interpreted as direct comparative effectiveness for the head-to head comparisons considered in this report (e.g. rTMS vs. VNS, rTMS vs. ECT). For the main head to head comparisons for this report (ECT, rTMS, VNS, and psychotherapy), we graded evidence as being precise when results had a low degree of uncertainty. We had two separate reviewers evaluate the overall strength of evidence for each major outcome based on a qualitative assessment of strength of evidence for each domain and reconciled all disagreements. The levels of strength of evidence are shown in Table 5.

Table 5. Strength of evidence grades and their definitions

| Grade | Definition |
|---|---|
| High | **High confidence that the evidence reflects the true effect.** Further research is very unlikely to change our confidence in the estimate of effect. |
| Moderate | **Moderate confidence that the evidence reflects the true effect.** Further research may change our confidence in the estimate of effect and may change the estimate. |
| Low | **Low confidence that the evidence reflects the true effect.** Further research is likely to change our confidence in the estimate of effect and is likely to change the estimate. |
| Insufficient | Evidence either is unavailable or does not permit estimation of an effect. |

We present our strength of evidence findings in our Overview sections. We will perform two strength of evidence assessments. First, we will grade strength of evidence for Tier 1 studies only (those with TRD). Subsequently, we will assess strength of evidence for the body of data that includes Tier 2 and 3 studies, highlighting how taking into account this additional information affects results relative to those found in Tier 1 alone.

# Data Synthesis

Although we use the tiers as a guide to describe all the included evidence, our primary focus is on the populations with a Tier 1 TRD definition (2 or more previous treatment failures). Some studies do not clarify whether failures occurred in a "current" episode or during one or more previous episode(s) (which can be characterized as over a "lifetime"). For that reason, our tiers may include a mix of studies that assess failing treatments in the current episode or failing treatments over a more extended period that may involve more than one episode. We highlight this distinction as appropriate. We also highlight other aspects of how treatment resistance, diagnosis, or severity of illness might vary.

As described above, a complex and broad array of factors have the potential to shape the answers to the key questions. Throughout this report we synthesized the literature qualitatively.

If data were sufficient, we also augmented findings with quantitative analyses. We conducted meta-analyses of data for comparisons involving trials that were fairly homogenous in study populations, treatment intervention, and outcome assessments. For efficacy, we used three outcome measures:

1. The weighted mean difference of changes on the HAM-D. We chose this outcome measure to have an estimate of the actual difference in effect sizes between treatments.
2. The relative risk (RR) of being a responder (more than 50 percent improvement from baseline) on the Hamilton Rating Scale for Depression (HAM-D) or the Montgomery-Asberg Depression Rating Scale (MADRS) at study endpoint.
3. The relative risk (RR) of achieving remission on the Hamilton Rating Scale for Depression (HAM-D) or on the Montgomery-Asberg Depression Rating Scale (MADRS) at study endpoint. The HAM-D definition for the 17-item version was ≤ 8, and for the 21-item version was ≤ 10. For the MADRS, the remission definition was a score of ≤ 8. If a study used a slightly different definition for remission, this difference was noted in the study's summary table and was included if in the authors' judgment did not substantially differ from the above.

For each meta-analysis, we conducted a test of heterogeneity ($I^2$ index) and applied both a random and a fixed effects model. We report the results from random effects models because, in all meta-analyses, the results from random and fixed effects models were very similar. If the RR was statistically significant, we calculated the number needed to treat (NNT) from the pooled relative risk or the pooled risk differences if variations in baseline risks were small.

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.

If meta-analyses were not possible but we deemed that an estimation of a treatment effect was of particular interest, we conducted descriptive statistics of the above mentioned outcome measures. We calculated weighted means and 95 percent confidence intervals of changes on HAM-D or MADRS, and the percentages of responders and remitters for specific interventions or treatment strategies. Findings provide an estimate of the average, to be expected treatment effect for a specific intervention. Nevertheless, they have to be interpreted cautiously. Because of the lack of control groups no general efficacy can be inferred from such results. Furthermore, the magnitude of treatment effects should not be compared across interventions.

## Peer Review

This draft CER received external peer review from the TEP members and individuals who were experts in fields relevant to treatment-resistant depression and from various stakeholder and user communities (listed in Appendix A[§§] [to be completed for final report]). The SRC managed the peer review process. If reviewers provided additional references to consider for inclusion in the final report, we reviewed all suggested references and included those that were appropriate

[§§] Appendixes are cited in this report and provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/___/___.pdf.

and within the scope of this CER. We also addressed all comments and revised the report accordingly.

# Chapter 3. Results

## Introduction

This chapter presents the results of our synthesis of the evidence on all six key questions (KQ , summarized in Table 6) about nonpharmacologic interventions for treating patients with treatment-resistant depression (TRD). To summarize, for all KQ s except KQ 1, we are concerned with four major nonpharmacologic interventions: electroconvulsive therapy (ECT), repetitive transcranial magnetic stimulation (rTMS), vagus nerve stimulation (VNS), and cognitive behavioral therapy or interpersonal psychotherapy (CBT or IPT). As noted in Table 6, KQ 1b asks about *pharmacologic* interventions in patients who have failed two or more previous treatments.

Table 6. Key questions about treatment-resistant depression (TRD)

| Key Questions |
|---|
| KQ 1a. Efficacy of nonpharmacologic interventions for acute-phase TRD (response or remission)<br>KQ 1b. Efficacy of pharmacologic interventions for acute-phase TRD (response or remission), for patients failing two or more prior treatments |
| KQ 2. Efficacy for maintaining response or remission (e.g., preventing relapse or recurrence) |
| KQ 3. Efficacy for acute-phase TRD as a function of particular symptom subtypes (e.g., catatonia or psychosis) |
| KQ 4. Harms of nonpharmacologic interventions (i.e., safety, adverse events, or adherence issues) |
| KQ 5. Efficacy or harms of nonpharmacologic treatments for selected subgroups defined by sociodemographic characteristics or coexisting conditions |
| KQ 6. Health-related outcomes of nonpharmacologic treatments (e.g., quality of life) |

This chapter is organized as follow: first by KQ ; second by intervention comparison, third by type of treatment failure (i.e., tier), and then by MDD or MDD and bipolar study populations. In addition, according to the specifications from the Agency for Healthcare Research and Quality (AHRQ) for comparative effectiveness reviews (CER), within each KQ section, we present an overview, then key points, and finally detailed analyses. Finally, as explained in Chapter 2, we graded the strength of evidence for all major comparisons and outcomes. In order to provide our readers with all the available evidence we have stratified the strength of evidence assessment by tier . We first present the strength of evidence assessments for Tier 1 studies alone (TRD). Then, we provide a second series of tables describing the strength of evidence when including studies from Tiers 1 through 3 and describe how the Tier 2 and 3 evidence supplements the evidence provided in Tier 1 studies alone.

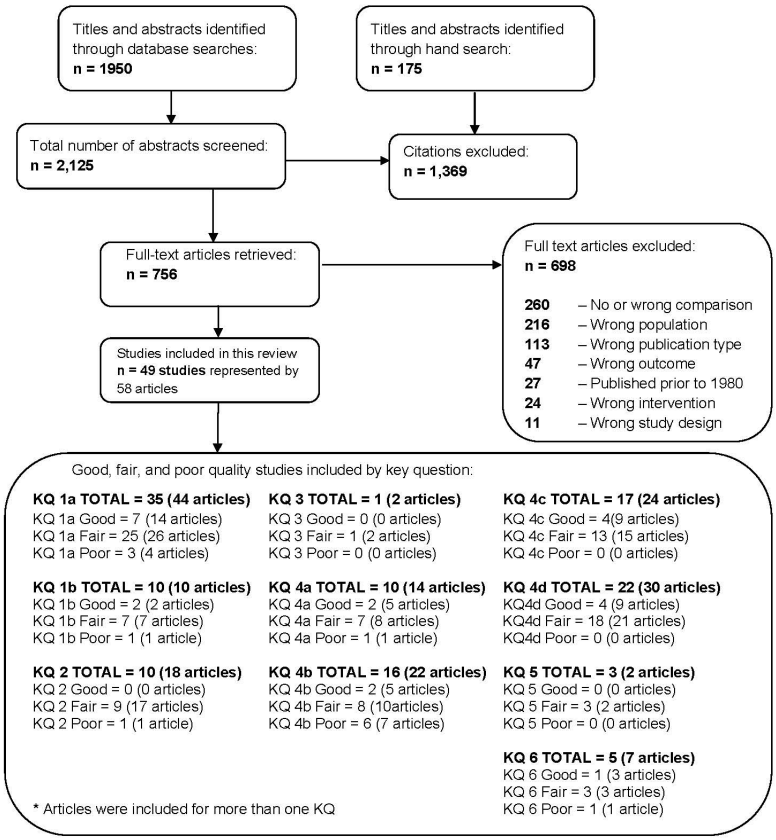
We focus in this chapter chiefly on trials, which can be head-to-head investigations or trials with control arms involving sham procedures or, for behavioral interventions, various forms of "usual care" that can include physician (psychiatrist) visits, medications, or both. For KQ 4 on harms, we also include observational studies. Evidence tables for all studies are presented in Appendix E.

We include information only on studies for which our quality ratings were good or fair; most studies were rated fair, so we specifically call out quality ratings only for good trials or studies. Poor-quality studies are listed in Appendix G;[***] in a very few cases in which a poor-quality study may have had the only relevant information on a major comparison or outcome, we will cite information about statistically significant findings in the detailed analysis text. Summary tables in the detailed analyses subsections have only good or fair quality studies.

[***] Appendixes are cited in this report and provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/___/___.pdf.

We identified 1,982 citations from searches across databases. Additionally, we detected 187 articles from manually reviewing the reference lists of pertinent review articles. Figure 4 documents the disposition of the 58 articles in this review. Working from 767 articles retrieved for full review, 711 were excluded at this stage (Appendix D). We included 58 published articles reporting on 49 studies: 48 randomized controlled trials (RCTs) (56 articles) and 1 observational study (1 article). Evidence tables for included studies, by key question, can be found in Appendix E.

Figure 4. PRISMA tree/disposition of articles




Of the 58 included articles, 12 (21 percent) were supported by pharmaceutical or device manufacturers; 37(64 percent) were funded by governmental or independent funds. We could not determine the source of support for 9 (15 percent) studies.

Reasons for exclusion were based on eligibility criteria or methodological criteria. Studies that originally met eligibility criteria but were later rated as poor quality for internal validity are located in Appendix E. For KQ 1a, 3 studies were rated as poor. KQ 1b, KQ 2, KQ 4a, and KQ 6 each rated one study as poor. KQ 4b excluded 6 studies for poor internal validity. Within the subsections KQ 4c and KQ 4d, no poor studies were identified. The main reason for rating as poor of studies was due to poor reporting of methodology.

# Key Question 1: Organization of Efficacy and Effectiveness Results

The presentation of KQ 1, which deals only with efficacy and effectiveness of interventions undertaken in acute phase treatment, is complex. Such clinical outcomes are one of a number of variables guiding the selection of therapy. Other considerations in acute phase treatment, such as effectiveness for subgroups, harms, and other health-related outcomes such as quality of life are addressed by KQ 3 through 6. KQ 2, in contrast, assesses the role of treatment selection in maintaining response or remission during continuation phase treatment.

Our primary focus is on comparisons of nonpharmacologic interventions—ECT, rTMS, VNS, and psychotherapy—presented as KQ 1a. We present evidence that stratifies first by which interventions are being compared, then by tier, and then by whether the population was MDD

only or MDD/Bipolar mix. Within each tier, we attempt to assess the effect on outcomes of key PICOTS elements: whether the population is MDD vs. MDD/Bipolar mix; whether treatment failure is required in the current episode; the level of depressive severity; treatment characteristics (e.g., number of treatment sessions, treatment location); and treatment strategy (e.g., whether patients switched to a new treatment or added it to augment their current treatment). We focus on Tier 1 TRD data first, and then we consider potentially relevant data from Tiers 2 and 3. We begin by reviewing this head-to-head literature.

Given the limited number of head-to-head comparisons available, we also review the nonpharmacologic interventions vs. control to assess whether we might be able to extend our analyses through indirect comparison. Such indirect analyses require a suitable number of comparisons with placebo or sham groups across the interventions.

Next, in KQ 1b, we compare nonpharmacologic to pharmacologic interventions. We present the evidence in a similar order. First, we review head-to-head nonpharmacologic vs. pharmacologic comparisons. Second, we review available pharmacologic vs. pharmacologic literature addressing response to antidepressant management to provide a comparison of what might be expected with a next step pharmacologic treatment for TRD. These comparisons involve only MDD-only, Tier 1 study populations. In reviewing the pharmacologic literature, we attempt to identify adequate control groups that would allow us to generate indirect measures of the relative outcomes of pharmacologic vs. control interventions that we can compare to the nonpharmacologic effect sizes. Throughout KQ 1, we provide a qualitative synthesis of the evidence; this is paired with a quantitative analysis of this data when an adequate number of studies are identified.

Our main outcomes of interest are changes in depressive severity, rates of response, and rates of remission. Most studies report these outcomes using a version of the HAM-D, so we focus on this result; however, in the absence of HAM-D scores, we used MADRS, BDI, or QIDS-SR scores. In Table 7, information is provided for these scales. For each outcome, we report the results of appropriate statistical tests comparing results between groups. All statistics are based on an intent-to-treat analysis unless otherwise specified (e.g., as last observation carried forward, or LOCF analyses). In studies in which the mean change in depression severity or proportion of responders or remitters is not reported but in which sufficient information is provided to calculate these variables, we made the calculations and include this information in the tables. We also categorized each population for depression severity using the chart described in Table 7 of Chapter 2 on methods. We consider only studies assessed as good or fair quality.

Table 7. Abbreviations and full names of diagnostic scales and other instruments

| Abbreviated Name | Complete Name of Measure or Instrument | Range of Scores | Improvement Denoted by |
|---|---|---|---|
| BDI | Beck Depression Inventory | 0-63 | Decrease |
| HAM-$D_{17}$ | Hamilton Rating Scale for Depression – 17 item | 0-52 | Decrease |
| HAM-$D_{21}$ | Hamilton Rating Scale for Depression – 21 item | 0-64 | Decrease |
| HAM-$D_{24}$ | Hamilton Rating Scale for Depression – 24 item | 0-75 | Decrease |
| HAM-$D_{25}$ | Hamilton Rating Scale for Depression – 25 item | 0-52 | Decrease |
| MADRS | Montgomery-Asberg Depression Rating Scale | 0-60 | Decrease |
| QID-SR | Quick Inventory of Depressive Symptomology – Self Report | 0-27 | Decrease |

# Key Question 1a: Nonpharmacologic Interventions–Overview of Head-to-Head Comparisons

Five head-to-head comparisons were available, three comparing ECT vs. rTMS and two comparing ECT vs. a combination of ECT plus rTMS (Table 8). Only two of these studies involved TRD (Tier 1) populations, and both of these studies enrolled MDD only patients.[51,52] These two studies showed no difference in outcome between tested treatments. Tier 2 evidence added one study of an MDD only population comparing ECT vs. rTMS[53] showing no difference in outcomes. Tier 3 evidence added two studies. One study was a good quality trial (reported in three articles) involved an MDD/Bipolar mix population and compared ECT vs. rTMS using an augmentation strategy, showing better outcomes for the ECT group.[54-56] The second study added a comparison of ECT vs. ECT plus rTMS in an MDD only population, showing no difference in outcome between the two treatments.[57]

Table 8. Number of good and fair quality studies by comparison, tier, and diagnostic mix

| Comparison, Tier | Diagnosis | |
|---|---|---|
| | MDD only | MDD and Bipolar Disorder |
| **ECT vs. rTMS** | | |
| Tier 1 (≥ 2 failures) | 1 | 0 |
| Tier 2 (≥ 1 failures) | 1 additional | 0 |
| Tier 3 (probable) | 0 | 1 additional |
| **ECT vs. ECT plus rTMS** | | |
| Tier 1 (≥ 2 failures) | 1 | 0 |
| Tier 2 (≥ 1 failures) | 0 | 0 |
| Tier 3 (probable) | 1 additional | 0 |

ECT, electroconvulsive therapy; MDD, major depressive disorder; rTMS, repetitive transcranial magnetic stimulation.

All studies included patients with severe depression and none required a failure in the current episode, preventing an assessment of the role of these variables on outcome. There was no clear difference in outcome as a function of varying the Tier definition, but the few studies limit observation of any true pattern. Due to the small number of studies, we were unable to determine whether outcomes varied by treatment strategy or treatment characteristics.

We could not assess how type of treatment strategy affected outcomes because of the limited number of studies and the multiple types of treatment strategies used. Studies varied by whether the trial tested interventions as a switch strategy (switching from the current failed treatment to a new strategy), an augmentation strategy (adding the new intervention to the current regimen), or some mixture of the two. Most treatment strategies involved switching from the current failed treatment.[51,53,57,58] One study reported in three articles was a pure augmentation trial.[54-56] Other studies included a combination of the two strategies.[52,53] Finally, some studies compared combinations of treatments (such as ECT vs. ECT plus rTMS[52,57]).

Strength of evidence assessments were made for three outcomes: change in depressive severity, response rates, and remission rates. We first will present the strength of evidence for Tier 1 studies alone, and then present strength of evidence for all three tiers considered together. When possible, within each comparison we report results by treatment strategy since this is a fundamental aspect of the antidepressant therapy.

## Strength of Evidence: Tier 1 Only

One study provides a low strength of evidence that there were no difference in either depressive severity, response rates or remission rates between switching to ECT vs. switching to

rTMS, (Table 9) Similarly, a second study provides a low strength of evidence that there were no differences in changes in depressive severity or between groups augmenting with ECT or with ECT plus rTMS (Table 10).

Table 9. Comparative clinical outcomes in those with two or more failures ECT vs. rTMS for Tier 1 studies; number of studies and subjects, and strength of evidence

| Outcome | Number of Studies; Subjects | Risk of bias Design/ Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| Change in depressive severity | 1; 42 | Medium<br>RCT<br>1 fair | Consistent | Direct | Imprecise | 1 fair trial using a switch strategy found both ECT and rTMS improved symptom severity; there were no significant differences between ECT and rTMS in symptom improvement<br>Low |
| Response | 1; 42 | Medium<br>RCT<br>1 fair | Consistent | Direct | Imprecise | 1 fair trial using a switch strategy found no significant differences in response rates between ECT (20%) and rTMS (45%)<br>Low |
| Remission | 1; 42 | Medium<br>RCT<br>1 fair | Consistent | Direct | Imprecise | 1 fair trial using a switch strategy found no significant difference in remission rates between ECT (15%) and rTMS (9%)<br>Low |

ECT, electroconvulsive therapy; RCT, randomized controlled trials; rTMS, repetitive transcranial magnetic stimulation.

Table 10. Comparative clinical outcomes in those with two or more failures ECT vs. ECT plus rTMS for Tier 1 studies; number of studies and subjects, and strength of evidence

| Outcome | Number of Studies; Subjects | Risk of bias Design/ Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| Change in depressive severity | 1; 22 | Medium<br>RCT<br>1 fair | Consistent | Indirect (compares combination to ECT rather than rTMS to ECT) | Imprecise | 1 fair trial using an augmentation strategy found both ECT and ECT plus rTMS improved symptom severity; there were no significant differences between ECT and ECT+ rTMS in symptom improvement<br>Low |
| Response | 0; 0 | — | — | — | — | — |
| Remission | 1; 22 | Medium<br>RCT<br>1 fair | Unknown | Indirect (compares combination to ECT rather than rTMS to ECT) | Imprecise | 1 fair trial using an augmentation strategy found no significant difference in remission rates between ECT and rTMS.<br>Low |

ECT, electroconvulsive therapy; RCT, randomized controlled trials; rTMS, repetitive transcranial magnetic stimulation.

## Strength of Evidence: Considering all Tiers

Considering evidence from all tiers added three studies, each to one of the existing two comparisons, producing a total of five head-to-head studies. Of these five, three compared ECT vs. rTMS[51,53,54] and 2 compared ECT plus rTMS vs. ECT.[52,57]

In comparing ECT and rTMS, considering all three tiers added two studies: a Tier 2 study in an MDD only population indicating no difference in outcome between ECT and rTMS,[53] and a Tier 3 study in an MDD/Bipolar mix population, the only good quality head-to-head study identified, which showed benefit for augmenting with ECT over augmenting with rTMS for all three outcomes.[54] Because of these conflicting results, the addition changed the strength of evidence for ECT vs. rTMS from a finding of low strength of evidence of no difference to one of an insufficient strength of evidence for changes in depressive severity, response rates, and remission rates (Table 11).

Table 11. Comparative clinical outcomes in those with one or more failures ECT vs. rTMS for Tiers 1-3 combined; number of studies and subjects, and strength of evidence

| Outcome | Number of Studies; Subjects | Risk of bias Design Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| Change in depressive severity | 3; 126 | Medium<br>RCT<br>1 good<br>2 fair | Inconsistent | Direct | Imprecise | 2 fair trials using switch strategies found both ECT and rTMS improved symptom severity; there were no significant differences between ECT and rTMS in symptom improvement.<br>1 good trial using an augmentation strategy found ECT to result in a significantly greater decrease in symptom severity compared to rTMS<br>Insufficient |
| Response | 3; 126 | Medium<br>RCT<br>1 good<br>2 fair | Inconsistent | Direct | Imprecise | 2 fair trials using switch strategies found no significant differences in response rates between ECT and rTMS.<br>1 good trial using an augmentation strategy found significantly greater response rates for ECT compared to rTMS.<br>Insufficient |
| Remission | 3; 126 | Medium<br>RCT<br>1 good<br>2 fair | Inconsistent | Direct | Imprecise | 2 fair trials using switch strategies found no significant difference in remission rates between ECT and rTMS.<br>1 good trial using an augmentation strategy found significantly greater remission rates for ECT compared to rTMS.<br>Insufficient |

ECT, electroconvulsive therapy; RCT, randomized controlled trials; rTMS, repetitive transcranial magnetic stimulation.

In comparing ECT and ECT plus rTMS, considering all tiers added one Tier 2 trial in an MDD only population that indicated no difference between the two interventions for all three outcomes (Table 12).[57] These results were consistent with the single Tier 1 finding; accordingly, the strength of evidence remained low that there was no difference between the two interventions.

Table 12. Comparative clinical outcomes in those with one or more failures ECT vs. ECT plus rTMS for Tiers 1-3; number of studies and subjects, and strength of evidence

| Outcome | Number of Studies; Subjects | Risk of bias Design Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| Change in depressive severity | 2; 44 | Medium<br>RCT<br>2 fair | Consistent | Indirect (compares combination to ECT rather than rTMS to ECT) | Imprecise | 2 fair trials (1 using switch, 1 using augmentation) found both ECT and ECT plus rTMS improved symptom severity; there were no significant differences between ECT and ECT+ rTMS in symptom improvement.<br>Low |
| Response | 1; 22 | Medium<br>RCT<br>1 fair | Unknown | Indirect (compares combination to ECT rather than rTMS to ECT) | Imprecise | 1 fair trial using an augmentation strategy found no significant differences in response rates between ECT and rTMS.<br>Low |
| Remission | 1; 22 | Medium<br>RCT<br>1 fair | Unknown | Indirect (compares combination to ECT rather than rTMS to ECT) | Imprecise | 1 fair trial using an augmentation strategy found no significant difference in remission rates between ECT and rTMS.<br>Low |

ECT, electroconvulsive therapy; RCT, randomized controlled trials; rTMS, repetitive transcranial magnetic stimulation.

# Key Question 1a: Nonpharmacologic Interventions–Key Points of Head-to-Head Comparisons

## Electroconvulsive Therapy vs. Repetitive Transcranial Magnetic Stimulation

Two fair MDD only studies, one Tier 1[51] and one Tier 2[53] found no differences in changes in depressive symptomatology, response, or remission. A single good quality Tier 3 MDD/Bipolar mix study found a greater change in depressive symptomatology and higher response and remission rates in the ECT group.[54,56]

## Electroconvulsive Therapy vs. ECT Plus Repetitive Transcranial Magnetic Stimulation

Two fair studies, one Tier 1 MDD only[52] and one Tier 3 MDD only[57] found no difference in changes in depressive symptomatology, response, or remission.

# Key Question 1a: Nonpharmacologic Interventions–Detailed Analysis of Head-to-Head Comparisons

## Electroconvulsive Therapy vs. Repetitive Transcranial Magnetic Stimulation

**Tier 1: Patients with two or more failures.** One trial comparing ECT with rTMS was identified in Tier 1 (Table 13).

Table 13. ECT vs. rTMS studies—Tiers 1through 3

| Author, Year<br>Endpoint<br>Current Episode<br>Failure Requirement<br>Quality | Intervention and Sample Size<br>Study Details | Population<br>Characteristics | Change in Depressive Symptoms | Response<br>Remission |
|---|---|---|---|---|
| | | **Tier 1** | | |
| Rosa et al., 2006[51]<br>2-4 weeks of active treatment (after week 2, rTMS non-responders withdrawn with LOCF)<br>Did not require failure in the current episode<br>Fair | **ECT (n = 20)**<br>% bilateral NR, mean number of sessions 10 (1.5)<br>rTMS (n = 22)<br>High frequency (10Hz), up to 20 sessions, 2500 pps (slightly outside safety guidelines)<br>**Treatment Strategy**<br>Switch<br>Definitions<br>Remission Ham-$D_{17}$ ≤ 7 | **Mean number of failed anti-depressant trials:**<br>ECT: NR<br>rTMS: NR<br>**Baseline Depression:**<br>HAM-$D_{17}$, mean (SD)<br>ECT: 32.1 (5.0)*<br>rTMS: 30.1 (4.7)*<br>*completers analysis<br>ECT: n = 15<br>rTMS: n = 20 | **HAM-$D_{17}$**<br>Change, mean (SD): NR<br>*P* = 0.86 | **HAM-$D_{17}$**<br>Response, n (%)<br>ECT: 6 (20)<br>rTMS: 10 (45)<br>*P* = 0.35<br>Remission, n (%)<br>ECT: 3 (15)<br>rTMS: 2 (9)<br>*P* = 0.65 |
| | | **Tier 2** | | |
| Grunhaus et al., 2003[53]<br>4 weeks for rTMS; ECT was at physician discretion, all reported pts included in analysis<br>Did not require failure in the current episode<br>Fair | **ECT (n = 20)**<br>35% bilateral, mean sessions = 10.25 (3.1)<br>rTMS (n = 20)<br>High frequency, 20 sessions<br>**Treatment Strategy**<br>Switch<br>Definitions<br>Response defined as a decrease ≥ 50% or HAM-D17 score ≤ 10 and a GAF rating ≥ 60<br>Remission defined as HAM-D17≤ 8 | **Number of failed antidepressant trials:**<br>% with ≥ 2 failed<br>ECT: 60<br>rTMS: 65<br>**Baseline Depression:**<br>HAM-$D_{17}$, mean (SD)<br>ECT: 25.5 (5.9)<br>rTMS: 24.4 (3.9) | **HAM-$D_{17}$**<br>Change, mean (SD)<br>ECT: -12.3<br>rTMS: -11.1<br>*P* = NS | **HAM-$D_{17}$**<br>Response, n (%)<br>ECT: 12 (60)<br>rTMS: 11 (55)<br>*P* = NS<br>Remission, n (%)<br>ECT: 6 (30)<br>rTMS: 6 (30)<br>*P* = NS |

AD, antidepressants; ECT, electroconvulsive therapy; G, group; GAF, global assessment of functioning; HAM-$D_{17}$, 17 item Hamilton Depression Scale; Hz, hertz; LOCF, last observation carried forward; mITT, _; n, number; NR, not reported; *P*, p-value; pps, pulses per session; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation.

Table 13. ECT vs. rTMS studies—Tiers 1through 3 (continued)

| Author, Year<br>Endpoint<br>Current Episode<br>Failure Requirement<br>Quality | Intervention and Sample Size<br>Study Details | Population Characteristics | Change in Depressive Symptoms | Response<br>Remission |
|---|---|---|---|---|
| | | **Tier 3** | | |
| McLoughlin et al. 2007,[54] Eranti et al. 2007,[55] and Knapp et al. 2008[56]<br><br>End of treatment (at clinicians discretion for ECT group, 3 weeks in rTMS), m-ITT<br><br>Did not require failure in the current episode<br><br>Tier 3—referred for ECT<br><br>Good | ECT (n = 22)<br>82% bilateral, mean session 6.3 (2.5)<br><br>rTMS (n = 24)<br>High frequency, 15 sessions<br><br>**Treatment strategy**<br>Augmentation<br><br>Definitions<br>• Remission defined as ≤ 8 | **Diagnosis**<br>Bipolar (%)<br>ECT: 9.1<br>rTMS: 8.3<br><br>**Number of failed antidepressant trials:**<br>Mean (SD)<br>ECT: 2.5 (1.4)<br>rTMS: 2.4 (1.0)<br><br>**Baseline Depression:**<br>HAM-$D_{17}$, mean (SD)<br>ECT: 24.8 (5.0)<br>rTMS: 23.9 (7.0) | **HAM-$D_{17}$**<br>Change, mean (SD)*<br>ECT: -14.1<br>rTMS: -5.4<br>*P* = 0.017<br><br>*only pts with post-baseline assessment<br>G1: n = 22<br>G2: n = 23 | **HAM-$D_{17}$**<br>Response, n (%)*<br>ECT: 13 (59.1)<br>rTMS: 4 (17.4)<br>*P* = 0.005<br><br>Remission, n (%)*<br>ECT: 13 (59.1)<br>rTMS: 4 (17.4)<br>*P* = 0.005 |

*MDD only.* One trial directly compared four weeks of ECT (n = 20 patients) to high frequency rTMS (n = 22 patients) (Table 13).[51] The mean baseline HAM-$D_{17}$ for treatment completers was 32.1 (SD (5.0) (ECT; n = 15 patients) and 30.1 (SD 4.7) (rTMS; n = 20 patients), indicating that the groups were severely depressed. ECT was initially unilateral, and it was switched to bilateral if there were no response after two weeks; the mean number of treatments was 10. If rTMS patients had not responded after two weeks, they exited the study with their last observation carried forward (LOCF). The treatment strategy was a switch. ITT analyses indicated no difference between the likelihood of response with ECT vs. rTMS (20% vs. 45%, $P = 0.35$), nor was there any difference between the likelihood of remission (15% vs. 9%, $P = 0.65$).

*MDD/Bipolar.* There were no eligible studies.

**Tier 2: Patients with one or more failures.** One trial comparing ECT with rTMS was identified in Tier 2 (Table 13).

*MDD only.* One additional study was captured considering Tier 2 (Table 12).[53] This trial directly compared up to four weeks ECT (n = 20 patients) with 20 sessions of high frequency rTMS (n = 20 patients) after patients were switched from antidepressant pharmacotherapy (Table 8). Patients were severely depressed (mean HAM-$D_{17}$ for ECT group 25.5 [SD 5.9] and for rTMS group 24.4 [SD 3.9]). For the ECT group, patients began with unilateral treatment but were switched to bilateral treatment if response was limited. Although rTMS treatment totaled 20 sessions, ECT treatment continued until the treating physician assessed that a therapeutic response had been obtained or no further benefit was expected. The authors' analyses accounted for all patients who were randomized. At the end of treatment, ECT and rTMS patients did not differ significantly in either depressive severity (- 12.3 vs. -11.1), the response rate (60 percent vs. 55 percent), or the remission rate (30 percent vs. 30 percent).

*MDD/Bipolar.* There were no eligible studies.

**Tier 3: Patients with probable treatment resistant depression.** One trial comparing ECT with rTMS was identified in Tier 3 (Table 13).

*MDD only.* There were no eligible studies.

*MDD/Bipolar mix.* One additional study reported in three articles was identified for Tier 3 (Table 13).[54-56] This study, the only good trial involving a head-to-head comparison, used an augmentation strategy to compare outcomes following two-to-three weeks of ECT (n = 22 patients) vs. 3 weeks of rTMS (n = 24) in a group of patients referred for ECT. While failure of a prior antidepressant treatment was not a selection criterion for the study, the mean number of previous antidepressant failures was approximately 2.5 in each treatment group. The ECT group had 9.1 percent bipolar disorder (n = 2), and the rTMS groups had 8.3 percent (n = 2) with bipolar disorder. Patients were moderately depressed at baseline (mean HAM-$D_{17}$ 23.9 [SD 7.0] for rTMS and 24.8 [SD 5.0] for ECT). In a modified ITT analysis, ECT patients had better outcomes in all depression domains recorded at the end of treatment. Compared to the rTMS group, those receiving ECT experienced a greater decrease in depressive severity (mean HAM-$D_{17}$ change -14.1 vs. -5.4, $P = 0.017$) and higher rates of both response and remission (59.1% vs. 17.4%, $P = 0.005$ for each, as all who responded also remitted).

**Tiers 1-3: Combined results.** Two fair studies found no differences between groups in change in depressive severity, response, or remission;[51,53] while one good study found ECT resulted in greater efficacy across the three measures.[54-56] With only three studies identified for this comparison, it is difficult to assess what study design, participant, or treatment characteristics may have contributed to different results in both intervention efficacy and between group comparisons.

While the study indicating greater efficacy for ECT was identified in Tier 3, the mean number of failed trials (N = 2.4-2.5) indicates substantial overlap with patients included in Tier 1 and Tier 2 studies.[54-56] Baseline characteristics reported in the Tier 2 study also show overlap with Tier 1 populations with over 60 percent of participants failing two or more antidepressant treatment trials.[53] None of the studies comparing ECT with rTMS required an antidepressant failure in the current episode. Average baseline depression scores indicate severe depression for all study populations. Only four participants with bipolar disorder were included in the single study allowing patients with this diagnosis to participate; these patients were equal distributed between treatment groups.[54-56]

Both studies finding no differences used switch strategies[51,53] while the study showing greater efficacy for ECT used an augmentation strategy.[54-56] All studies used high frequency rTMS and bilateral ECT. Both studies finding no differences had similar treatment durations for both ECT (mean number of sessions = 10) and rTMS (up to 20 sessions),[51,53] while the study finding ECT resulted in greater efficacy had a shorter duration of treatment for both interventions (ECT mean number of sessions = 6.3; rTMS sessions = 15).[54-56]

## Electroconvulsive Therapy vs. ECT Plus Repetitive Transcranial Magnetic Stimulation

**Tier 1: Patients with two or more failures.** One trial comparing ECT with ECT plus rTMS was identified in Tier 1 (Table 14).

*MDD only.* One trial directly compared two weeks of unilateral ECT (n = 11 patients) to a combination of one day of unilateral ECT followed by 4 days of high frequency rTMS (n = 11 patients).[52] Patients were severely depressed at entry (median HAM-$D_{17}$ for ECT group = 30 and

Table 14. ECT vs. ECT plus rTMS studies—Tiers 1-3

| Author, Year<br>Endpoint<br>Episode Failure<br>Quality | Intervention and Sample Size<br>Study Details | Population Characteristics | Change in Depressive Symptoms | Response Remission |
|---|---|---|---|---|
| | | **Tier 1** | | |
| Pridmore, 2000[52]<br>2 weeks of treatment<br>Did not require failure in the current episode<br>Fair | **ECT (n = 11)**<br>100% unilateral, 6 sessions<br>**ECT plus rTMS (n = 11)**<br>ECT: 100% unilateral (day 1), plus high frequency rTMS: (days 2-5)<br>Repeated in week 2<br>**Treatment Strategy**<br>Primarily augmentation (4 patients not on AD at start). ADs and mood stabilizers continued but other psychotropics discontinued<br>**Definitions**<br>Remission HAM-$D_{17}$ < 9 | **Mean number of failed anti-depressant trials:**<br>ECT: NR<br>ECT+rTMS: NR<br>**Baseline Depression :**<br>HAM-$D_{17}$, median<br>ECT: 30<br>ECT+rTMS: 28 | **HAM-$D_{17}$**<br>Change, median<br>ECT: -23<br>ECT+rTMS G2: -20<br>*P* = 0.6 | **HAM-$D_{17}$**<br>Remission, n (%)<br>ECT: 6 (54.5)<br>ECT+rTMS G2: 6 (54.5)<br>*P* = NR |
| | | **Tier 3** | | |
| Chistyakov et al. 2005[57]<br>3 weeks, all reported patients included<br>Did not require failure in the current episode<br>Tier 3 –referred for ECT<br>Fair | ECT plus sham (n = 10)<br>100% Unilateral ECT (2 days a week) plus sham rTMS (4 days a week)<br>**ECT plus rTMS (n = 12)**<br>100% Unilateral ECT (2 days a week) plus low frequency rTMS(4 days a week)<br>**Treatment** strategy<br>Switch | **Mean number of failed anti-depressant trials:**<br>ECT + sham: NR<br>ECT+rTMS: NR<br>**Baseline Depression:**<br>HAM-D mean reported in graph only | **HAM-$D_{NR}$**<br>Change, mean (SD)<br>ECT+sham: NR<br>ECT+rTMS: NR<br>*P* > 0.05 | **HAM-$D_{NR}$**<br>Response, n (%)<br>Overall: 19 (86)<br>ECT+sham: NR<br>ECT+rTMS: NR<br>*P* = NS |

AD, antidepressants; ECT, electroconvulsive therapy; HAM-$D_{17}$, 17 item Hamilton Depression Scale; HAM-$D_{NR}$, _ Hamilton Depression Scale; n, number; NR, not reported; NS, not sufficient; *P*, p-value; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation.

for ECT plus rTMS group = 28). For the majority of patients, this trial tested an augmentation strategy. However four patients (2 in each group) were not taking any antidepressant medication at study entry and patients were allowed to continue any mood stabilizers they were taking (1 in each group). ITT analyses showed no clear difference in outcomes between the two groups. Specifically, there was no difference in change in depressive severity (-23 vs. -20, *P* = 0.6) or remission rates (54.5% vs. 54.5%, *P* = not reported).*MDD/Bipolar.* There were no eligible studies.

**Tier 2: Patients with one or more failures.** There were no eligible studies.

**Tier 3: Patients with probable treatment resistant depression.** One trial comparing ECT with ECT plus rTMS was identified in Tier 1 (Table 14).

*MDD only.* Following discontinuation of antidepressant pharmacotherapy (switch strategy), a three week study compared six sessions of unilateral ECT plus 12 sessions of low frequency rTMS (n = 12) vs. six sessions of unilateral ECT plus 12 sessions of sham rTMS (n = 10).[57] Depressive severity was not reported in text, but figures indicate HAM-D (NR) was above 40 for each group, suggesting very severe depression. The treatment strategy was a switch, and no other

psychotropic medications were allowed. All patients were included in the final analysis. There was no clear difference in response rates between ECT plus rTMS vs. rTMS alone (data not reported, *P* = NS).

*MDD/Bipolar mix.* There were no eligible studies.

**Tier 1-3 combined results.** Two fair studies found no differences between groups in change in depressive severity, response, or remission.[52,57] With only two studies identified for this comparison, it is difficult to assess how study design, participant, or treatment characteristics may have affected treatment efficacy; furthermore, one of the two studies did not report specific data points impeding additional analysis.

Overall, studies appeared similar with the exception of Tier. One study fell into Tier 1[52] and one into Tier 3[57] with no information provided regarding the average number of antidepressants failed prior to study entry for the Tier 3 study.[57] Neither study required a failure in the current episode. All patients were diagnosed with MDD and the average baseline depression scores indicate severe depression for both study populations. Dosing strategies for the combination groups in both two studies were similar with patients receiving one to two sessions of ECT and four sessions of rTMS per week. ECT strategies were also similar with patients receiving 2-3 ECT sessions per week. All ECT treatments were unilateral while one study used high frequency rTMS[52] and the other used low frequency.[57] Lastly, one study duration was two weeks while the other was three weeks.

# Key Question 1a: Nonpharmacologic Interventions–Overview of Active vs. Control Comparisons

A total of 23 comparisons of an active nonpharmacologic intervention vs. a sham or control group were identified (Table 15), providing a total of 3 distinct comparisons. No comparisons of ECT vs. control were identified. The small number of studies within some comparisons (e.g. VNS: 1 study, psychotherapy N = 3) did not allow an indirect comparison of nonpharmacologic interventions. A sufficient number of studies comparing rTMS to sham stimulation allowed for some comparisons across variables. Results for Tier 1 vs. Tiers 1-3 combined were consistent and generally consideration of all tiers provided more conservative point estimates with narrower confidence intervals, suggesting that the tier results might be reasonably combined. Results for MDD only and MDD/Bipolar mix populations were in the same direction and of similar magnitude, suggesting that combining results from these two populations was reasonable. A limited number of studies within comparisons restricted analysis

Table 15. Number of studies by comparison and tier

| Comparison and Population Included | Studies Included | |
|---|---|---|
| | MDD Only | MDD and Bipolar Disorder |
| **rTMS vs. sham** | | |
| Tier 1 (≥ 2 failures) | 8 | 4 |
| Tier 2 (≥ 1 failures) | 3 additional | 1 additional |
| Tier 3 (probable) | 0 | 3 additional |
| **Psychotherapy vs. control** | | |
| Tier 1 (≥ 2 failures) | 0 | 0 |
| Tier 2 (≥ 1 failures) | 3 additional | 0 |
| Tier 3 (probable) | 0 | 0 |
| **VNS + TAU vs. TAU** | | |
| Tier 1 (≥ 2 failures) | 0 | 1 additional |
| Tier 2 (≥ 1 failures) | 0 | 0 |
| Tier 3 (probable) | 0 | 0 |

MDD, major depressive disorder; rTMS, repetitive transcranial magnetic stimulation. TAU, treatment as usual; VNS, vagus nerve stimulation.

and prevented assessment of whether outcomes differed by depressive severity, treatment strategy, treatment characteristics, and whether failure in the current episode was required.

Strength of evidence assessments were made for three outcomes: change in depressive severity, response rates, and remission rates. Below, we present the results first for Tier 1 only, and then for Tiers 1-3 combined, highlighting how consideration of all tiers affected results. Within the text of the strength of evidence sections, we combine MDD and MDD/Bipolar mix studies together, as there were minimal differences between these two groups, but within the strength of evidence table results column, we stratify by the two populations to allow the reader to see this breakdown. When possible, within each comparison we report results by treatment strategy since this is a fundamental aspect of the antidepressant therapy.

## Strength of Evidence Tier One Only

A total of twelve different Tier 1 trials compared rTMS vs. sham control for at least one for the three outcomes (Table 16). For changes in depressive severity, twelve rTMS vs. sham control studies (8 MDD only, 4 MDD/Bipolar mix) involving 400 participants provide a high degree of evidence that rTMS produces a greater decrease in depressive severity. Studies that did not report significant differences had small samples. A random effects meta-analysis of ten Tier 1 studies indicated that rTMS produces a decrease in HAM-D depressive severity of over 5 points relative to sham control.

Table 16. rTMS vs. sham: strength of evidence—Tier 1 only

| Comparison | Number of Studies; Subjects | Risk of bias Design/ Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| Change in depressive severity | 12; 400 | Low<br>RCT<br>3 good<br>9 fair | Consistent | Indirect | Precise | 2 good and 3 fair trials using augmentation strategies found rTMS to have a significantly greater decrease in depressive severity compared to sham. The same relationship was seen in 3 more fair studies but these results were not statistically significant; and another fair study where the results of a significance test were not reported.<br>2 trials using a mixed strategy, 1 good, 1 fair, found a significantly greater decrease in depressive severity for rTMS.<br>1 fair trial using a switch strategy found a nonsignificantly greater decrease for rTMS.<br>High |

NR, not reported; PT, patient; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation.

Table 16. rTMS vs. sham: strength of evidence—Tier 1 only (continued)

| Comparison | Number of Studies; Subjects | Risk of bias Design/ Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| Response | 9; 329 | Low<br>RCT<br>3 good<br>6 fair | Consistent | Indirect | Precise | 1 good and 4 fair trials using augmentation strategies found rTMS to have a significantly higher response rate compared to sham. The same relationship was seen in 2 more trials, 1 good and 1 fair, but results of test of statistical significance were not reported. Two other fair augmentation trials showed the same relationship but results did not reach statistical significance.<br>1 fair trial, using a switch strategy found rTMS to have a higher response rate compared to sham (*P* = NR).<br>1 good trial using a mixed strategy found a significantly higher response rate for rTMS.<br>High |
| Remission | 3; 151 | Low<br>RCT<br>2 good<br>1 fair | Consistent | Direct | Precise | 1 good trial using a mixed strategy found a significantly greater remission rate for rTMS<br>1 good trial using an augmentation strategy found a significant difference between groups while 1 fair study showed the same relationship but did not report on significance.<br>Moderate |

For changes in response rates, nine rTMS vs. sham control studies involving 329 participants (6 MDD only studies and 3 MDD/Bipolar mix studies) provided a high degree of evidence that rTMS is more likely to produce a response than sham control. A random effects meta-analysis of nine Tier 1 studies shows that patients receiving rTMS are over 3 times more likely to achieve a depressive response that patients receiving sham control.

For changes in remission rates, three rTMS vs. sham control studies (1 MDD only, 2 MDD/Bipolar mix) involving 151 patients provided moderate strength of evidence that rTMS produces greater remission rates than rTMS (Table 16). A random effects meta-analysis of four Tier 1 studies shows that patients receiving rTMS are over 6 times more likely to achieve remission that patients receiving sham control.

In the only other Tier 1 comparison, one good quality VNS vs. sham control study in an MDD/Bipolar mix population involving 222 participants provides low evidence that neither a change in depressive severity nor response rates following VNS substantially differ from a sham control (Table 17).

Table 17. VNS vs. sham: strength of evidence—Tier 1 only

| Comparison | Number of Studies; Subjects | Risk of bias Design/ Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| Change in depressive severity | 1; 222 | RCT<br>Low<br>1 good | Unknown | Indirect | Precise | 1 good trial that augmented treatment as usual found no significant difference in decreases in depressive severity between patients receiving VNS and sham.<br>Low |
| Response | 1; 222 | RCT<br>Low<br>1 good | Unknown | Indirect | Precise | 1 good trial that augmented treatment as usual found no significant difference in response rates between patients receiving VNS and sham.<br>Low |
| Remission | 0; 0 | — | — | — | — | — |

MDD, major depressive disorder; NS, not significant; PT, patient; VNS, vagus nerve stimulation.

## Strength of Evidence: Considering all Tiers

Considering all tiers added seven total studies to the comparison of rTMS vs. sham control (i.e. total number of studies = 19), all consistent with Tier 1 only findings. The only effect on strength of evidence was to advance a finding from moderate strength of evidence to high that remission rates for rTMS are greater than sham control (Table 18). We identified no additional VNS vs. sham studies.

Table 18. rTMS vs. sham: strength of evidence Tiers 1-3

| Comparison | Number of Studies; Subjects | Risk of bias Design Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| Change in depressive severity | 19; 821 | Low<br>RCT<br>4 good<br>15 fair | Consistent | Indirect | Precise | 2 good and 5 fair trials using augmentation strategies found rTMS to have a significantly greater decrease in depressive severity compared to sham. The same relationship was seen in 3 more fair studies but these results were not statistically significant; |

MDD, major depressive disorder; NR, not reported; NS, not significant; PT, patient; RCT, randomized controlled trials; rTMS, repetitive transcranial magnetic stimulation; VNS, vagus nerve stimulation; (x), x = number of comparisons.

Table 18. rTMS vs. sham: strength of evidence Tiers 1-3 (continued)

| Comparison | Number of Studies; Subjects | Risk of bias Design Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| | | | | | | and 1 other fair study where the results of a significance test were not reported.<br><br>3 trials using a mixed strategy, 1 good, 2 fair, found a significantly greater decrease in depressive severity for rTMS.<br><br>1 good and 2 fair trial using switch strategies found rTMS to have a significantly greater decrease in depressive severity compared to sham. 2 fair trials using a switch strategy found no significant differences between rTMS and sham.<br>High |
| Response | 13; 739 | Low<br><br>RCT<br><br>4 good<br>9 fair | Consistent | Indirect | Precise | 2 good and 1 fair trial using augmentation strategies found rTMS to have a significantly higher response rate compared to sham. 3 fair augmentation trials found higher response rates for rTMS but did not report statistical significance. 1 fair augmentation trial found a nonsignificantly higher response rate for rTMS.<br><br>1 good trial using a switch strategy found rTMS to have a significantly higher response rate compared to sham. 1 fair switch trial found a nonsignificantly higher response rate for rTMS. 2 fair switch trial found a higher response rate for rTMS but did not report statistical significance. 1 fair switch trial found that rTMS and sham had equal response rates.<br><br>1 good trial using a mixed strategy found a significantly higher response rate for rTMS.<br><br>High |
| Remission | 6; 490 | Low<br><br>RCT<br><br>3 good<br><br>3 fair | Consistent | Direct | Precise | 1 good trial using an augmentation strategy found a significantly greater remission rate for rTMS. 1 fair study showed the same |

Table 18. rTMS vs. sham: strength of evidence Tiers 1-3 (continued)

| Comparison | Number of Studies; Subjects | Risk of bias Design Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| | | | | | | relationship but did not report on significance.<br>1 good trial using a switch study found a significantly greater remission rate for rTMS compared to sham; 1 fair trial found the same relationship but did not report on significance; 1 fair trial using a switch strategy found no differences between groups.<br>1 good trial using a mixed strategy found a significantly greater remission rate for rTMS<br>High |

Further, considering evidence from all tiers added a new comparison, psychotherapy vs. control, which included three studies involving 207 participants from MDD only populations and provided, overall, a low strength of evidence that CBT produced better outcomes than control groups (Table 19). Results among the three outcomes differed. There was insufficient evidence for any difference in changes in depressive severity in the comparison of CBT vs. control, but there was a low degree of evidence that CBT procedure greater response and remission rates than a control group.

For rTMS vs. control, seven studies (three MDD only studies with 390 participants and four MDD/Bipolar mix studies with 83 participants) addressing changes in depressive severity were added, bringing the total number of studies to 19 trials with 915 participants (Table 18). The strength of evidence favoring rTMS over control remained high. A random effects meta-analysis of 16 Tier 1-3 studies indicated that rTMS produce a decrease in HAM-D depressive severity of approximately 5 HAMD points relative to sham control. Compared to a Tier 1 only meta-analysis, the point estimate from the Tier 1-3 analysis was more conservative and provided a narrower confidence interval, suggesting greater precision.

For response rates in rTMS vs. control, consideration of all tiers added four studies (three MDD only studies with 390 participants, and one MDD/Bipolar mix study with 20 participants), for a total of thirteen studies with 740 participants (Table 19). The data were consistent with Tier 1 outcomes, and the strength of evidence favoring rTMS over sham remained high. A random effects meta-analysis of these 13 Tier 1-3 studies shows that patients receiving rTMS are over two times more likely to achieve a depressive response than patients receiving sham control. Compared to the analogous Tier 1 only meta-analysis, the point estimate was more conservative and the confidence interval narrowed, providing more precision.

Table 19. Psychotherapy vs. control: strength of evidence—Tiers 1-3

| Comparison | Number of Studies; Subjects | Risk of bias Design/ Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| Change in depressive severity | 3; 207 | Low<br>RCT<br>1 good<br>2 fair | Inconsistent | Indirect | Imprecise | 2 fair trials using augmentation strategies found a significantly greater decrease in depressive severity for the psychotherapy group compared to control.<br>1 good trial, also augment, found no significant differences between groups.<br>Insufficient |
| Response | 1; 25 | Medium<br>RCT<br>1 fair | Unknown | Indirect | Imprecise | 1 fair study using an augment strategy found a greater response rate in the psychotherapy group compared to control (*P* = NR)<br>Low |
| Remission | 2; 182 | Low<br>RCT<br>1 good<br>2 fair | Consistent | Indirect | Imprecise | 1 good trial showed a significantly higher remission rate for psychotherapy than control.<br>1 fair trial favored psychotherapy but significance NR.<br>Low |

NR, not reported; RCT, randomized controlled trial.

For remission rates in rTMS vs. control, consideration of all tiers added three studies, all Tier 2 MDD only with a sum of 349 participants, for a total of 7 studies with 490 participants (Table 18). The date remained consistent with Tier 1 data, and the strength of evidence remained moderate-to-high that rTMS produces greater remission rates than sham control. A random effects meta-analysis of 7 Tier 1-3 studies shows that patients receiving rTMS are over two times more likely to achieve a remission than patients receiving sham control. Compared to a Tier 1 only meta-analysis, the point estimate was more conservative and the confidence interval narrower, again providing greater precision.

In the additional comparison provided, three MDD only studies compared psychotherapy to control in a total of 207 participants (Table 19). For changes in depressive severity, these threes studies provide insufficient data for whether CBT produces a better outcome than control. One large good quality study that found no difference[59,60] included patients with mild depressive severity, while the other two studies involving participants with greater depressive severity patients showed a benefit for psychotherapy augmentation.

For response rates in CBT vs. control, a single Tier 2 study involving 25 severely depressed patients provides a low degree of evidence that psychotherapy produces better outcomes. For

remission rates in CBT vs. control, two Tier 2 studies involving 182 severely depressed patients provide a low degree of evidence that psychotherapy produces better remission rates.

# Key Question 1a: Efficacy or Effectiveness of Nonpharmacologic Interventions for Acute Phase Treatment—Key Points of Active vs. Control Comparisons

Active vs. control comparisons were also limited and the small number studies within comparisons prevented an indirect meta-analytic synthesis. Available comparisons of nonpharmacologic treatment compared to control were available for three comparisons: rTMS vs. sham (n = 11 studies reported in 12 articles for an MDD only population[61-72] and n = 8 studies for an MDD/Bipolar mix population[17,73-79]); VNS vs. sham (n=1 study involving an MDD/Bipolar mix population);[80] and psychotherapy to a control group (n = 3 studies reported in four articles);[59,60,81,82]

## rTMS vs. Sham

For Tier 1, eight MDD only and four MDD/Bipolar mix studies were identified. Four studies, two in MDD only (three articles)[67,72,83] and two in MDD/Bipolar[17,74] were deemed good quality, while the remaining studies were assessed as fair. Though some studies did not report tests of statistical significance or had very small sample sizes, evidence generally supported the benefit of rTMS over sham for a decrease in depressive symptomatology and a greater likelihood of response and remission. Results from MDD only and from MDD/Bipolar mix studies were in the same direction and of similar magnitude, and results from combining these two populations did not substantially differ from MDD only, suggesting that combining these two populations was reasonable. Meta-analyses in TRD (Tier 1) involving both MDD only and MDD/Bipolar mix populations indicated benefit for rTMS over sham. TRD patients treated with rTMS had significantly greater decreases in depressive symptomatology (decrease in HAMD -5.56, 95% CI, -6.90 to -4.22). rTMS patients were also over three times as likely to respond (pooled relative risk for response 3.26, 95% CI, 1.83 to 5.81, which translates to a Number Needed to Treat = 5 [95% CI, 4-9]), and over 6 times as likely to remit (pooled relative risk for response 6.40, 95% CI, 2.07 to 19.76), with a Number Needed to Treat of 4 (95% CI, 3—6).

Consideration of all Tiers together for the combined MDD and MDD/Bipolar mix populations provided results consistent with those from Tier 1 alone combined but with more conservative point estimates and narrower confidence intervals, suggesting that results from analyses of studies from all three tiers reflect what can be expected in TRD (Tier 1) populations. The weighted mean difference in HAM-D depressive severity was -4.98 (95% CI, -5.86 to -4.11). The pooled relative risk indicated that patients receiving rTMS were over twice as likely to respond than those receiving placebo (pooled relative risk 2.31, 95% CI, 1.62 to 3.30), which translates into a Number Needed to Treat of 5 (95% CI, 4—10). Remission rates also favored rTMS. The pooled relative risk for remission was 2.33 (95% CI, 1.39 to 3.93), which translates to a Number Needed to Treat of 8 (95% CI, 6—15).

This finding of the above clinical outcomes from Tiers 1, 2, and 3 reflecting what was found with Tier 1 alone held whether the population included was MDD only, or MDD/Bipolar mix, respectively.

Findings addressing the remaining key PICOTS elements were limited. Three quarters of the Tier 1 studies utilized an augmentation strategy (n = 5 MDD only,[61-63,65,68] and n = 4 MDD/Bipolar mix[17,73-75] while others (all MDD only) used a switch (n = 1)[64] or a mixed strategy (n = 2).[66,67]. There was no clear difference in outcome as a function of strategy, but the limited number of comparisons prevented a firm conclusion. The consideration of additional tiers of evidence did not affect this finding.

For the few Tier 1 studies, we were unable to detect clear differences by treatment characteristics (i.e. pharmacotherapy strategy, rTMS frequency, or treatment duration) through qualitative analysis due to other potentially confounding variables resulting from study design or participant characteristics. The consideration of additional tiers of evidence did not affect this finding.

For Tier 1, one study did not report baseline depressive severity[73] and one study focused on patients with moderate disease severity[62] while the remaining ten studies were on patients with severe depression. With little variation by depressive severity, we were unable to detect any differences by this variable. The consideration of additional tiers of evidence did not affect this finding.

Only three studies required a failure in the current episode, two in MDD only[62,63] and one in MDD/Bipolar mix[73], with no differences in outcomes apparent, but the small number of studies prevented a more formal analysis. The consideration of additional tiers of evidence did not affect this finding.

Finally, data addressing the effect of varying treatment characteristics was limited. One Tier 1 study suggested that the length of treatment may be positively associated with greater benefit for rTMS,[67] but the small number of studies with acute phase rTMS treatment beyond two weeks restricts any conclusion. The consideration of additional tiers of evidence did not affect this finding.

## VNS vs. Sham

We identified only one study comparing VNS to sham, conducted in a Tier 1 MDD/Bipolar mix population.[80] The majority of measures utilized by this study found no difference between VNS and sham on changes in depressive severity or rates of response and remission. Since only a single study was identified for this comparison, further assessment by key variables was not possible.

## Psychotherapy vs. Control

For the third comparison, one good study reported in two articles[59,60] and two fair studies[81,82] supported greater outcomes for patients in psychotherapy compared to a control group. All three studies were in Tier 2, involved MDD only populations, and utilized similar treatment characteristics in the psychotherapy groups. There was no clear effect of treatment strategy or the requirement for a failure in the current episode.

# Key Question 1a: Efficacy or Effectiveness of Nonpharmacologic Interventions for Acute Phase Treatment—Detailed Analysis of Active vs. Control Comparisons

## rTMS vs. Sham

Nineteen studies provided an rTMS vs. sham comparison for TRD-related depressive illnesses. Twelve were Tier 1 studies, four were Tier 2 studies, and three were Tier 3 studies. Eleven involved MDD only population, while 8 had an MDD/Bipolar mixture (<20 percent with Bipolar disorder).

**Tier 1: Patients with two or more failures.** Twelve Tier 1trials comparing rTMS with sham were identified in Tier 1.

*MDD Only.* Of the eight Tier 1 MDD only studies identified,[61-68,83] only one trial, reported in two articles, was good quality.[67,83]

Five of these studies tested rTMS as an augmentation strategy (Table 20).[61-63,65,68] The largest augmentation study was a two week trial that compared high frequency rTMS (n = 20 patients) to a sham control (n = 20 patients).[63] Participants' depression was severe (mean HAM-$D_{21}$ in rTMS group = 27.1, and 25.6 in control). In an analysis of treatment completers, rTMS patients had a greater decrease in depressive severity (-7.05 vs. -1.77, $P$ = 0.003). Including all participants, rTMS patients had a greater likelihood of response (25 percent vs. 5, $P$ = NR) compared to control patients.

Table 20. rTMS vs. sham, Tier 1 MDD only studies using an augmentation strategy

| Author, Year<br>Endpoint<br>Episode Failure<br>Quality | Intervention and Sample Size<br>Study Details | Population Characteristics | Change in Depressive Symptoms | Response Remission |
|---|---|---|---|---|
| Garcia-Toro et al., 2001[63]<br>2 weeks, completers analysis<br>Required failure in the current episode<br>Fair | rTMS (n = 20)<br>High frequency, 10 sessions<br>Sham (n = 20)<br>Treatment Strategy<br>Augmentation | **Mean number of failed anti-depressant trials:**<br>rTMS: NR<br>Sham: NR<br>**Baseline Depression**<br>HAM-$D_{21}$, mean (SD)<br>rTMS: 27.11 (6.65)<br>Sham: 25.6 (4.92) | **HAM-$D_{21}$***<br>Change, mean (SD)<br>rTMS: -7.05 (5.66)<br>Sham: -1.77 (3.78)<br>*P = 0.003*<br>*all results based on completers (G1: n = 17, G2: n = 18) | **HAM-$D_{21}$***<br>Response, n (%)<br>rTMS: 5 (25)<br>Sham: 1 (5)<br>*P* = NR |

DLPFC, dorsolateral prefrontal cortex; G, group; HAM-$D_{21}$, 21-Item Hamilton Depression Scale; HAM-$D_{25}$, 25-Item Hamilton Depression Scale; Hz, hertz; n, number; NR, not reported; NS, not significant; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation.

Table 20. rTMS vs. sham, Tier 1 MDD only studies using an augmentation strategy (continued)

| **Author, Year**<br>**Endpoint**<br>**Episode Failure**<br>**Quality** | Intervention and Sample Size<br>Study Details | **Population Characteristics** | **Change in Depressive Symptoms** | **Response**<br>**Remission** |
|---|---|---|---|---|
| Garcia-Toro et al., 2006[65]<br><br>2 weeks, all reported paticipants included in analysis<br><br>Did not require failure in the current episode<br><br>Fair | rTMS-1 (n = 10)<br>High frequency plus low frequency, 10 session<br><br>rTMS-2 (n = 10)<br>Same as above, but with individually assessed location<br><br>Sham rTMS (n = 10)<br>Double winged coil angled at 45 degrees<br><br>Treatment strategy<br>Augmentation | **Mean number of failed anti-depressant trials:**<br>rTMS: NR<br>Sham: NR<br><br>**Baseline Depression**<br>HAM-$D_{21}$, mean (SD)<br>rTMS-1: 27.30 (4.97)<br>rTMS-2: 25.00 (4.14)<br>Sham: 25.10 (7.28) | **HAM-$D_{21}$**<br>Change, mean (SD)<br>rTMS-1: -7.2<br>rTMS-2: -6.9<br>Sham G3: -1.5<br><br>rTMS-1 plus rTMS-2 (-7.05) vs. sham, *P* = 0.048 | **HAM-$D_{21}$**<br>Response, n (%)<br>rTMS-1: 2 (20)<br>rTMS-2: 2 (20)<br>Sham G3: 0 (0)<br>*P* = NR |
| Boutros et al., 2002[68]<br><br>2 weeks<br><br>Did not require failure in the current episode<br><br>Fair | rTMS (n = 12)<br>High frequency, 10 sessions<br><br>Sham (n = 9)<br><br>Treatment strategy<br>Augmentation<br><br>Definitions<br>Response1 definition: >30% decrease in HAM-$D_{25}$<br><br>Response2 definition: ≥50% decrease in HAM-$D_{25}$<br>**calculated from table | **Mean number of failed anti-depressant trials:**<br>rTMS: NR<br>Sham: NR<br><br>**Baseline Depression**<br>HAM-$D_{25}$, mean (SD)<br>rTMS: 34.4 (10.1)<br>Sham: 31.7 (4.9) | **HAM-$D_{25}$**<br>Change, mean (SD)<br>rTMS: -11.75<br>Sham: -6.22<br>*P* = NS | **HAM-$D_{25}$**<br>Response1, n (%)<br>rTMS: 7 (58.3)<br>Sham: 2 (22.2)<br>*P* = NR<br><br>Response2, n (%)**<br>rTMS: 3 (25.0)<br>Sham: 2 (22.2)<br>*P* = NR |
| Kauffmann et al., 2004[62]<br><br>2 weeks<br><br>Did not require failure in the current episode<br><br>Fair | rTMS (n = 7)<br>Low frequency, 10 sessions<br><br>Sham (n = 5)<br><br>Treatment Strategy<br>Augmentation, pts encouraged to discontinue mood stablizers<br><br>Definitions<br>Response: HAM-$D_{21}$ < 10 | **Mean number of failed anti-depressant trials:**<br>rTMS: NR<br>Sham: NR<br><br>**Baseline Depression**<br>HAM-$D_{21}$, mean (SD)<br>rTMS: 21.86 (2.31)<br>Sham: 18.20 (2.20) | **HAM-$D_{21}$**<br>Change, mean (SD)<br>rTMS: -10.57<br>Sham: -6.31<br>*P* = NS | **HAM-$D_{21}$**<br>Response, n (%)<br>rTMS: 4 (57)<br>Sham: 2 (40)<br>*P* = NS<br><br>Response2, n (%)<br>rTMS: 4 (57)<br>Sham: 1 (20)<br><br>*P* = NS |
| Padberg et al., 1999[61]<br><br>1 week<br><br>Required failure in the current episode<br><br>Fair | **rTMS (n = 6)**<br>High frequency, 5 sessions<br><br>Low-left rTMS (n = 6)<br>0.3 Hz, Left-DLPFC, 5 sessions<br><br>Sham rTMS (n = 6)<br><br>**Treatment** strategy<br>Augmentation, 16.7% not on medication at study entry | **Mean number of failed anti-depressant trials (current episode):**<br>High rTMS: 4.0 (2.2)<br>Low-left rTMS: 3.2 (0.8)<br>Sham: 3.2 (1.2)<br><br>**Baseline Depression**<br>HAM-$D_{21}$, mean (SD)<br>High rTMS: 30.2 (9.5)<br>Low-left rTMS: 26.7 (9.4)<br><br>Sham: 22.2 (8.8) | **HAM-$D_{21}$**<br>Change, mean (SD)<br>High rTMS: -1.7<br>Low-left rTMS: -5.2<br>Sham: -1.3<br>*P* = NS | **HAM-$D_{21}$**<br>Response: NR<br>Remission: NR<br>*P* = NR |

Another two week study testing augmentation compared two active rTMS treatments (n = 10 patients each) with each other and with 10 sessions of sham stimulation (n = 10 patients).[65] Enrolled patients were severely depressed (mean HAM-$D_{21}$ item scores for each group between 25 and 27.3). The three groups did not appear to differ by decrease in depressive severity. However, the two active groups combined did have a greater 2 week decrease in depressive severity than the sham control group (-7.05 vs. -1.5, *P* = 0.048). Also, two patients in each of the active groups responded at two weeks, compared to no patients in the control group (*P* = NR).

A two week augmentation study compared high frequency rTMS (n = 12 patients) to sham rTMS treatment (n = 9 patients).[68] At entry, patients in the two groups were severely depressed (mean HAM-$D_{25}$ item scores were 34.4 in the rTMS groups and 31.7 in the control group). Analysis was modified ITT. Patients in the rTMS group had a mean change in HAM-$D_{25}$ severity of -11.75 vs. -6.22 in the sham stimulation group (*P* = ns); the small sample size likely limited the power to detect a difference. Using the study's definition of response (> 30 percent in HAM-$D_{25}$ item), 58.3 percent of rTMS patients responded compared to 22.2 percent of the sham stimulation group (*P* = not reported). Using a more standard definition of response as 50 percent or greater decrease (able to be calculated from study information), 22.2 percent of rTMS patients responded compared to 20 percent of the group receiving sham stimulation (*P* = NR).

A small trial comparing outcomes at two weeks after 10 sessions of low frequency rTMS treatment (n = 7 patients) compared to sham rTMS treatment (n = 5 patients).[62] The groups had moderate depressive severity (HAM-$D_{21}$, 21.86 for rTMS, 18.2 for control). While mostly an augmentation study, patients were advised to discontinue benzodiazepines and mood stabilizers. ITT analyses showed that patients receiving rTMS had a 10.57 decrease in HAM-$D_{21}$ compared to a 6.31 decrease for the sham stimulation group (*P* = NS). Response rates did not differ between the two groups (57 percent vs. 40 percent, *P* = NS). Again, small sample sizes may have limited the power to detect differences.

The last augmentation study compared outcomes after 1 week of treatment with high frequency rTMS (n = 6 patients), low frequency rTMS to the left DLPFC (n = 6 patients), or sham rTMS stimulation (n = 6 patients).[61] One treatment failure needed to have occurred in the current episode. Enrolled patients were moderately-to-severely depressed (mean HAM-$D_{21}$ score 30.2, 26.7, and 22.2 for high frequency, low frequency, and control groups, respectively). Patients receiving low frequency rTMS had a significant decrease in depressive severity relative to baseline (mean HAM-$D_{21}$ change -1.7 for high frequency, -5.2 for low frequency to the left DLPFC, and - 1.3 for sham stimulation), but there was no difference in treatment effect between groups in this small study.

Of the remaining three studies identified, one tested a switch strategy and two used a mixed strategy (Table 21). The single switch study tested was a small two week trial that compared high frequency rTMS (n = 7 patients) to sham rTMS stimulation (n = 8 patients).[64] Patients were severely depressed (mean HAM-$D_{17}$ for the two groups was between 20 and 23). At two week followup, ITT analysis indicated that the decrease in depressive severity did not differ between the two groups (-8.1 for rTMS, -5.5 for sham, *P* = NS). Similarly, the rate of response did not appear to differ (28.6 percent vs. 14.3 percent, *P* = NR).

Table 21. rTMS vs. sham, Tier 1 MDD only using switch and mixed strategies

| Author, Year<br>Endpoint<br>Episode Failure<br>Quality | Intervention and Sample Size<br>Study Details | Population Characteristics | Change in Depressive Symptoms | Response<br>Remission |
|---|---|---|---|---|
| | | **Switch Strategy** | | |
| Holtzheimer et al., 2004[64]<br>2 weeks<br>Did not require failure in the current episode<br>Fair | rTMS (n = 7)<br>High frequency rTMS, 10 sessions<br>Sham rTMS (n = 8)<br>Treatment strategy<br>Switch | **Mean number of failed anti-depressant trials:**<br>rTMS: NR<br>Sham: NR<br>**Baseline Depression**<br>HAM-$D_{17}$, mean (SD)<br>rTMS: 22.7 (5.3)<br>Sham: 20.8 (6.3) | **HAM-$D_{17}$**<br>Change, mean (SD)<br>rTMS: -8.1<br>Sham: -5.5<br>*P* = NS | **HAM-$D_{17}$**<br>Response, n (%)<br>rTMS: 2 (28.6)<br>Sham: 1 (14.3)<br>*P = NR* |
| | | **Mixed Strategy** | | |
| Avery et al. 2006[67]<br>Patients treated over 4 weeks and primary endpoint 1 week after final txt<br>Did not require failure in the current episode<br>Good | rTMS (n = 35)<br>High frequency, 15 sessions over 4 weeks<br>Sham (n = 33)<br>Treatment strategy<br>Mixed-within group differences 31% of rTMS group and 27% of control group continued taking medications<br>Definitions<br>Remission definition: HAM-$D_{17}$ < 10 | **Mean number of failed anti-depressant trials:**<br>rTMS: 3.2 (2.44)<br>Sham: 3.3 (1.72)<br>**Mean number of failed anti-depressant trials (current episode):**<br>rTMS: 1.46 (0.78)<br>Sham: 1.48 (0.67)<br>**Baseline Depression**<br>HAM-$D_{17}$, mean (SD)<br>rTMS: 23.5 (3.9)<br>Sham: 23.5 (2.9) | **HAM-$D_{17}$**<br>Change, mean (SD)<br>rTMS: -7.8 (7.8)<br>Sham: -3.7 (6.3)<br>*P = 0.002* | **HAM-$D_{17}$**<br>Response, n (%)<br>rTMS: 11 (31.4)<br>Sham: 2 (6.1)<br>*P* = 0.008<br>Remission, n<br>rTMS: 7 (20.0)<br>Sham: 1 (3.0)<br>*P* = 0.033 |
| Pascual-Leone et al., 1996[66]<br>Crossover trial, 1 week<br>Did not require failure in the current episode<br>Fair | rTMS (n = 17)<br>High frequency, 5 sessions<br>Sham (n = 17)<br>Combined data from 4 control stimulations<br>Treatment strategy<br>Mixed—within group differences and combination (All pts in both groups given 30 mg/d nimodipine) | **Mean number of failed anti-depressant trials:**<br>rTMS: NR<br>Sham: NR<br>**Baseline Depression**<br>HAM-$D_{21}$, mean: NR | **HAM-$D_{21}$**<br>Change, mean:<br>TMS: NR<br>Sham: NR<br>*P* < 0.0005 | **HAM-$D_{21}$**<br>Response: NR<br>Remission: NR |

DLPFC, dorsolateral prefrontal cortex; G, group; HAM-$D_{17}$, 17-item Hamilton Depression Scale; HAM-$D_{21}$, 21-item Hamilton Depression Scale; HAM-$D_{25}$, 25-item Hamilton Depression Scale; Hz, hertz; m-ITT, modified-intent-to-treat, mg/d, milligram per day; MT, motor threshold; n, number; NR, not reported; NS, not significant; *P*, p-value; pts, patients; pps, pulses per session; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation; txt(s), treatment(s); vs., versus.

Two studies tested a mixed strategy (Table 20). One of these trials was a good quality four week study that compared 15 sessions of left-sided high frequency rTMS (n = 35 patients) to control treatment (n = 33 patients), and was the only one to report remission rates in this tier.[67,83] Groups enrolled were in general severely depressed (mean HAM-D17 score 23.5). This mixed

strategy was primarily a switch, although a substantial percentage of patients continued antidepressants (31 percent of rTMS group, 27 percent of control group) and benzodiazepines (26 percent and 24 percent, respectively). Outcomes were measured one week after completing the 4 week treatment, and all ITT analyses favored the rTMS group. Compared to controls, the rTMS group had greater decrease in depressive severity (-7.8 vs. -3.7, $P = 0.002$), a greater response rate (31.4 percent vs. 6.1 percent, $P = 0.008$), and a greater remission rate (20.0 percent vs. 3.0 percent, $P = 0.033$).

One small mixed study used a cross-over design to compare 17 TRD patients with psychotic symptoms randomized to receive different orderings of 1 high frequency rTMS intervention and 4 different sham rTMS interventions over a five week period.[66] Patients had at least three episodes of depression that had been resistant to multiple medications. Baseline depressive severity was not reported. Though patients attempted to discontinued their antidepressant medication, many were unable to do so making this strategy mixed (within group differences). All patients received nimodipine (which appears to have mood stabilizing effects) as a combination treatment with both the active rTMS and control interventions. Results suggested that the active rTMS produced greater improvement in HAM-$D_{21}$ scores than comparison groups ($P < 0.0005$)

*Meta-analytic Synthesis of Tier 1 MDD Only.* Meta-analyses supported the benefit of rTMS over sham control. The weighted mean difference in HAM-D depressive severity was -4.52 (95% CI, -6.25 to -2.80) (Figure 5).

The pooled relative risk indicated that patient receiving rTMS were over 2 ½ times more likely to have a treatment response than those receiving sham treatment (pooled relative risk = 2.62, 95% CI, 1.31 to 5.24) (Figure 6), which translates to a number needed to treat of 5 (95% CI, 4-10).

There was only one study reporting remission for Tier I MDD only, indicating greater remission rates for rTMS over sham (20% vs. 3%).[67]

Figure 5. Mean difference meta-analysis comparing rTMS vs. control on the HAM-D in a MDD only TRD population (Tier 1)

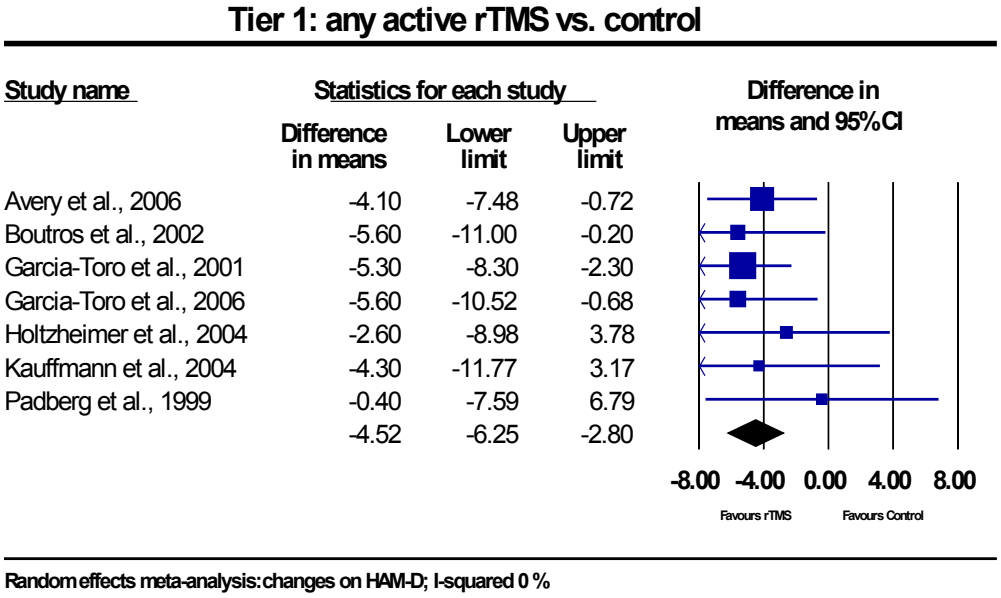
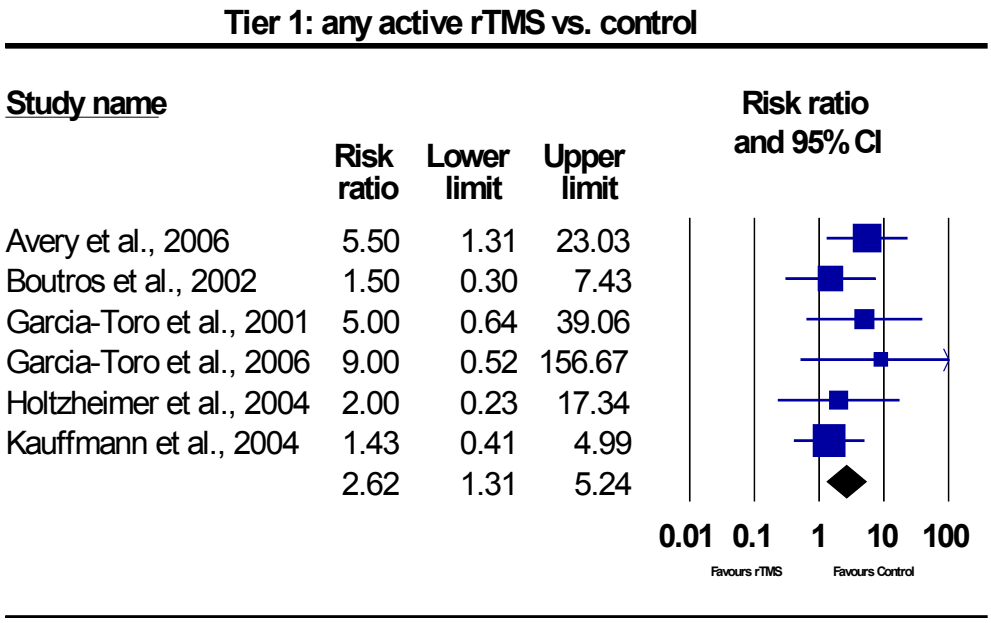


Figure 6. Relative risk meta-analysis of response rates comparing rTMS vs. sham in an MDD-only TRD population (Tier 1)




*MDD/Bipolar*. For rTMS vs. sham, four Tier 1 studies involving MDD/Bipolar mix populations, all using augmentations strategies, were identified.[17,73-75] These studies are summarized in Table 22 with detailed descriptions provided in the evidence tables (Appendix E).[†††]

Table 22. rTMS vs. sham, tier 1 MDD/BP studies

| Author, Year<br>Endpoint<br>Episode Failure<br>Quality | Intervention and Sample Size<br>Study Details | Population Characteristics | Change in Depressive Symptoms | Response<br>Remission |
|---|---|---|---|---|
| Bocchio-Chiavetto et al. 2008[73]<br><br>Crossover, 1 week, all reported patients included in the analysis<br><br>Required failure in the current episode<br><br>Fair | rTMS (n = 36)<br>Low Frequency rTMS (n = 18)<br>5 sessions<br>OR<br>High Frequency rTMS (n = 18)<br>5 sessions<br><br>Sham (n = 15)<br><br>Treatment strategy<br>Augmentation | **Diagnosis**<br>Bipolar (%)<br>Overall: 13.9<br>**Mean number of failed anti-depressant trials:**<br>Overall: 2.89<br><br>**Baseline Depression**<br>HAM-$D_{21}$, mean (SD)<br>rTMS: 23.19 (5.12)<br>Sham: 24.53 (4.79) | **HAM-$D_{21}$**<br>Change, mean (SD)<br>rTMS: -5.69<br>Sham: -3.40<br>*P* = NR | **HAM-$D_{21}$**<br>Response, n (%)<br>NR<br>Remission, n (%)<br>NR |

G, group; HAM-$D_{17}$, 17 item Hamilton Depression Scale; HAM-$D_{21}$, 21 item Hamilton Depression Scale; Hz, hertz; LOCF, last observation carried forward; MADRS, Montgomery-Åsberg Depression Rating Scale; n, number; NR, not reported; *P*, p-value; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation; txt, treatment.

[†††] Appendixes are cited in this report and provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/___/___.pdf.

Table 22. rTMS vs. sham, tier 1 MDD/BP studies (continued)

| **Author, Year**<br>**Endpoint**<br>**Episode Failure**<br>**Quality** | Intervention and Sample Size<br>Study Details | **Population Characteristics** | **Change in Depressive Symptoms** | **Response Remission** |
|---|---|---|---|---|
| Fitzgerald et al., 2006[17]<br><br>6 weeks of txt (after 2 weeks, patients with < 20% decrease in score exited with LOCF)<br><br>Did not require failure in current episode<br><br>Good | High plus Low rTMS (n = 25)<br>High frequency rTMS up to 30 sessions plus low frequency rTMS up to 30 sessions<br><br>Sham (n = 25)<br><br>Treatment strategy<br>Augmentation, 23% not taking any medication at start of study<br><br>Definitions<br>Remission definition: HAM-$D_{17}$ < 8 | **Diagnosis**<br>Bipolar (%)<br>rTMS: 16<br>Sham: 16<br><br>**Mean number of failed anti-depressant trials:**<br>rTMS: 5.6 (3.1)<br>Sham: 6.2 (3.0)<br><br>**Baseline Depression**<br>HAM-$D_{17}$, mean (SD)<br>rTMS: 22.5 (7.4)<br>Sham: 19.8 (4.4) | **HAM-$D_{17}$**<br>rTMS: -10.2<br>Sham: plus1.1<br>*P* < 0.001 | **HAM-$D_{17}$**<br>Response, n (%)<br>rTMS: 13 (52)<br>Sham: 2 (8)<br>*P* = 0.001<br><br>Remission (%)<br>rTMS: 10 (40)<br>Sham: 0 (0)<br>*P* = 0.001 |
| Fitzgerald et al. 2003[74]<br><br>2 weeks<br><br>Did not require failure in the current episode<br><br>Good | High rTMS (n = 20)<br>High frequency, 10 sessions<br><br>Low rTMS (n = 20)<br>Low frequency, 10 sessions<br><br>Sham (n = 20)<br><br>Treatment strategy<br>Augmentation<br><br>Definitions<br>Response1 definition: >20% decrease in MADRS score<br><br>Response2 definition: ≥50% decrease in MADRS | **Diagnosis**<br>Bipolar (%)<br>High rTMS: 5<br>Low rTMS G2: 5<br>Sham: 20<br><br>**Mean number of failed anti-depressant trials:**<br>Overall: 5.68 (3.40)<br><br>**Baseline Depression**<br>MADRS, mean (SD)<br>High rTMS: 36.05 (7.55)<br>Low rTMS G2: 37.70 (8.36)<br><br>Sham: 35.75 (8.14) | **MADRS**<br>Change, mean (SD)<br>High rTMS: -5.25<br>Low rTMS G2: -5.5<br>Sham: -0.35<br><br>High rTMS vs. sham, low rTMS vs. sham, *P* < 0.005 | **MADRS**<br>Response1, n (%)<br>High rTMS: 8 (40)<br>Low rTMS G2: 7 (35)<br>Sham: 2 (10)<br>*P* = 0.07<br><br>Response2, n (%)<br>High rTMS: 0 (0)<br>Low rTMS G2: 1 (5)<br>Sham: 0 (0)<br>*P = NR* |
| Su et al., 2006[75]<br><br>2 weeks, completers analysis<br><br>Did not require failure in the current episode<br><br>Fair | 20Hz rTMS (n = 11)<br>High frequency (20Hz), 10 sessions<br><br>5Hz rTMS (n = 11)<br>High frequency (5Hz),10 sessions<br><br>Sham (n = 11)<br><br>Treatment strategy<br>Augmentation<br><br>Definitions<br><br>Remission defined as HAM-$D_{21}$ < 8 | **Diagnosis**<br>Bipolar (%)<br>20 Hz rTMS: 10<br>5 Hz rTMS G2: 20<br>Sham G3: 20<br><br>**Mean number of failed anti-depressant trials:**<br>20 Hz rTMS: NR<br>5 Hz rTMS G2: NR<br>Sham G3: NR<br><br>**Baseline Depression**<br>HAM-$D_{21}$, mean (SD)<br>20 Hz rTMS: 23.2 (7.5)<br>5 Hz rTMS G2: 26.5 (5.2)<br><br>G3: 22.7 (4.7) | **HAM-$D_{21}$***<br>Change, mean (SD)<br>20 Hz rTMS: -13.4 (4.9)<br>5 Hz rTMS G2: -14.2 (6.0)<br>Sham G3: -3.7 (9.3)<br>*P* < 0.01<br><br>*n analyzed n = 10 in each group | **HAM-$D_{21}$***<br>Response, n (%)<br>20 Hz rTMS: 6 (60)<br>5 Hz rTMS G2: 6 (60)<br>Sham: 1 (10)<br>*P* = 0.01<br><br>Remission, n (%)<br>20 Hz rTMS: 5 (50)<br>5 Hz rTMS G2: 5 (50)<br>Sham G3: 0 (0)<br>*P* = NR |

One good six week study compared high frequency rTMS plus low frequency rTMS (n = 25 patients) to sham rTMS stimulation (n = 25 patients).[17] Failure was not required in the current

episode. The number of treatments depended on the presence of at least partial response. Patients entering the rTMS were severely depressed (mean HAM-$D_{17}$ of 22.5), while the control group was only moderately depressed (mean HAM-$D_{17}$ of 19.8). Sixteen percent of each group had bipolar disorder. rTMS patients had better outcomes than patients receiving sham stimulation on each response measure. Compared to control, rTMS patients had a greater improvement in HAM-D scores (-10.2 vs. -1.1, $P < 0.001$), greater response rate (52% vs. 8%, $P = 0.001$), and a greater remission rate (40% vs. 0%, $P = 0.001$).

Another good quality trial compared three groups: one with high frequency rTMS (n = 20 patients), one with low frequency rTMS (n = 20 patients), and one with sham stimulation (n = 20 patients) following 2 weeks of treatment.[74] The three groups had MADRS scores averaging between 35 and 38, consistent with severe depression. Both the high frequency and low frequency groups had 5 percent bipolar patients, while the control group had 20 percent. An ITT analysis favored the two rTMS groups. Both the high frequency (-5.25) and low frequency (-5.5) groups had greater decrease in MADRS severity than the sham group (-0.35, $P < 0.005$ for each comparison with control). Using a definition of response as > 20 percent in MADRS score, the two active groups tended to have greater rates of response (40 percent and 35 percent, respectively) compared to the sham stimulation group (10%) ($P = 0.07$ for both comparisons). Using the more standard definition of response as a 50 percent decrease, only one patient (in the low frequency group) responded by study end.

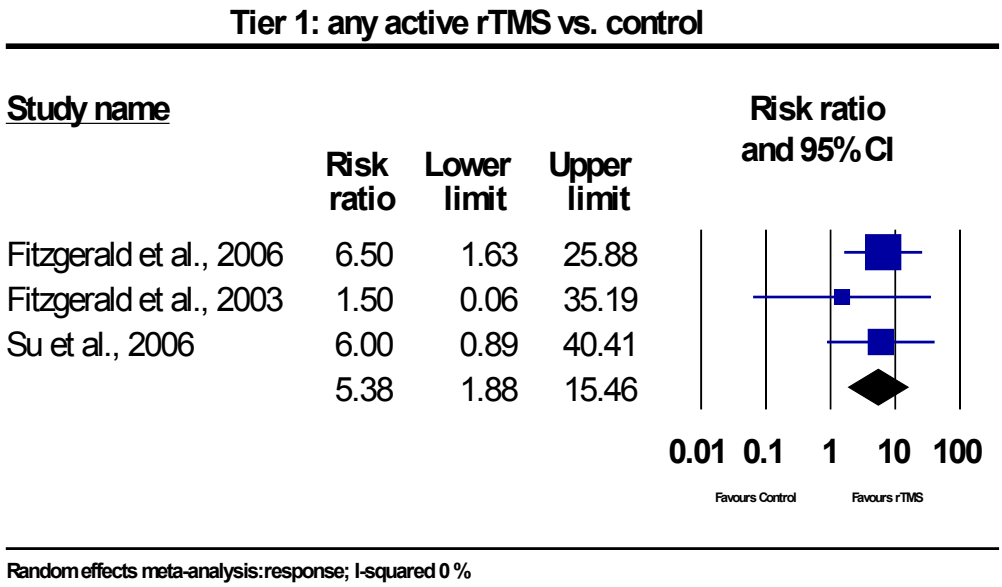
A two week study compared three groups: those receiving high frequency rTMS (20Hz) (n = 11 patients), those receiving "lower" high frequency rTMS (5Hz) (n = 11 patients), and those receiving sham rTMS treatment (n = 10 patients).[75] Study patients entered were severely depressed (mean HAM-$D_{21}$ severity for 20Hz group 23.2, 5Hz group 26.5, and sham group 22.7) The 20Hz high frequency group had 10 percent bipolar patients, while the other two group each had 20 percent with a bipolar depression. A treatment completer analysis showed that patients in the active groups had a greater decrease in HAM-$D_{21}$ severity (-13.4 and -14.2 respectively) than the control group (-3.7, $P < 0.01$ for each comparison). Similarly, response favored the two rTMS groups (60% for each vs. 10% for the sham stimulations comparison, $P = 0.01$ for both). Finally, both rTMS treatments had greater remission rates (50%) than the sham control group, which had no remitters ($P$ = not reported).

A fourth trial compared 1 week of low frequency rTMS (a group of 36, consisting of 18 who received low frequency rTMS and 18 who received high frequency rTMS) with one week of sham rTMS stimulation (involving a subgroup of 15 patients from the above group of 36 who received control treatment 8 weeks after having received rTMS).[73] Patients entered treatment severely depressed (mean HAM-$D_{21}$ severity 23.19 in rTMS group, 24.53 in sham group). No difference in decrease in HAM-$D_{21}$ item severity was identified (-5.69 in active group, -3.40 in control group, $P$ = not reported).

*Meta-analytic Synthesis of Tier 1 MDD/Bipolar Mix Outcomes.* We were able to quantitatively synthesize outcomes from among the four studies within an MDD/Bipolar mix Tier 1 population.[17,73-75] For changes in depressive severity involving the three studies using HAM-D as an outcome, patients receiving rTMS on average had approximately a 7 point greater decrease relative to sham control (- 7.15, 95% CI, -9.29 to -5.01). Because sample sizes were small and responses to placebo varied in the small control groups, the heterogeneity was high and our estimates are uncertain with respect to the magnitude of changes on the HAM-D. Given this uncertainty, we are not including the forest plot.

Response rates similarly favored rTMS, which increases the likelihood of response by more than a factor of five (relative risk 5.38, 95% CI, 1.88 to 15.46) (Figure 7), producing a Number Needed to Treat of 3 (95% CI, 1-14).

Figure 7. Relative risk meta-analysis of response rates comparing rTMS vs. sham in MDD/Bipolar mix TRD population (Tier 1)




With only two studies reporting remission for all Tiers combined (which only included Tier 1 and Tier 2 for this outcome), we did not quantitatively synthesize this information; both studies providing this outcome indicated greater absolutes remission rates for rTMS vs. sham (40% vs. 0%[17] and 50% vs. 0%[75]).

*Tier 1 MDD and MDD/BP Combined.* Meta-analyses combining TRD studies (Tier 1) from both MDD and MDD/Bipolar mix populations continued to support the benefit of rTMS over sham control. The weighted mean difference in HAM-D depressive severity was -5.56 (95% CI, -6.90 to -4.22) (Figure 8). The pooled relative risk indicated that patients receiving rTMS were over three times as likely to respond than those receiving placebo (pooled relative risk 3.26, 95% CI, 1.83 to 5.81) (Figure 9), which translates into a Number Needed to Treat of 5 (95% CI, 4—9). Remission rates also favored rTMS. The pooled relative risk for remission was 6.40 (95% CI, 2.07 to 19.76), which translates to a Number Needed to Treat of 4 (95% CI, 3—6) (Figure 10).

MDD/Bipolar mix point estimates tended to be slightly higher than those for MDD only, but confidence intervals overlapped, suggesting no clear difference. Indeed, combining the two populations did not affect the direction nor did it substantially impact the magnitude of the results, and the combined results were consistent with what was reported for the Tier 1 syntheses separately.

**Tier 2: Patients with one or more failures.** Consideration of Tier 2 provided four additional studies, three MDD only studies (reported in four articles)[69-72] and one additional MDD/Bipolar mix study.[76]

Figure 8. Effect size meta-analysis comparing rTMS vs. control on the HAM-D in an MDD and MDD/Bipolar mix TRD population (Tier 1)

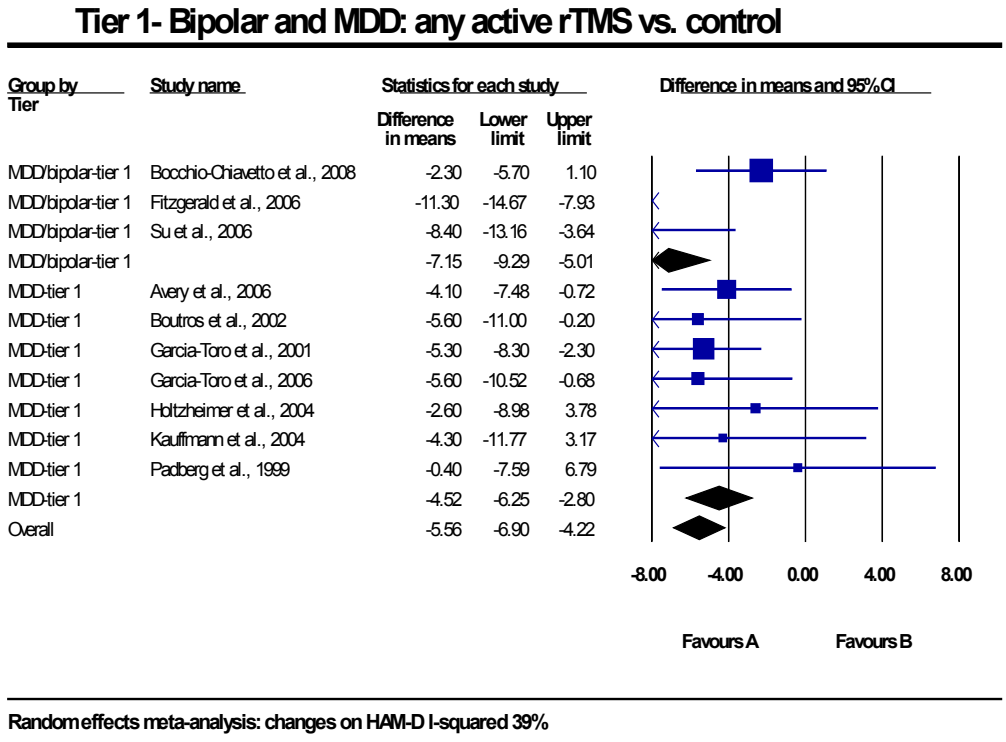



Figure 9. Relative risk meta-analysis of response rates comparing rTMs vs. sham in an MDD and MDD/Bipolar mix TRD population (Tier 1)

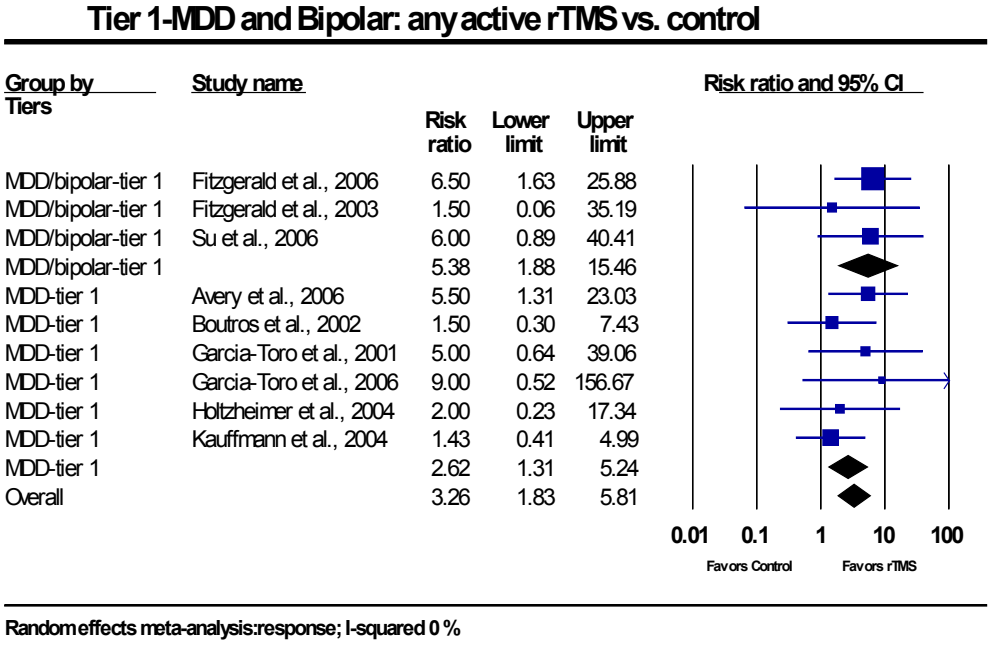
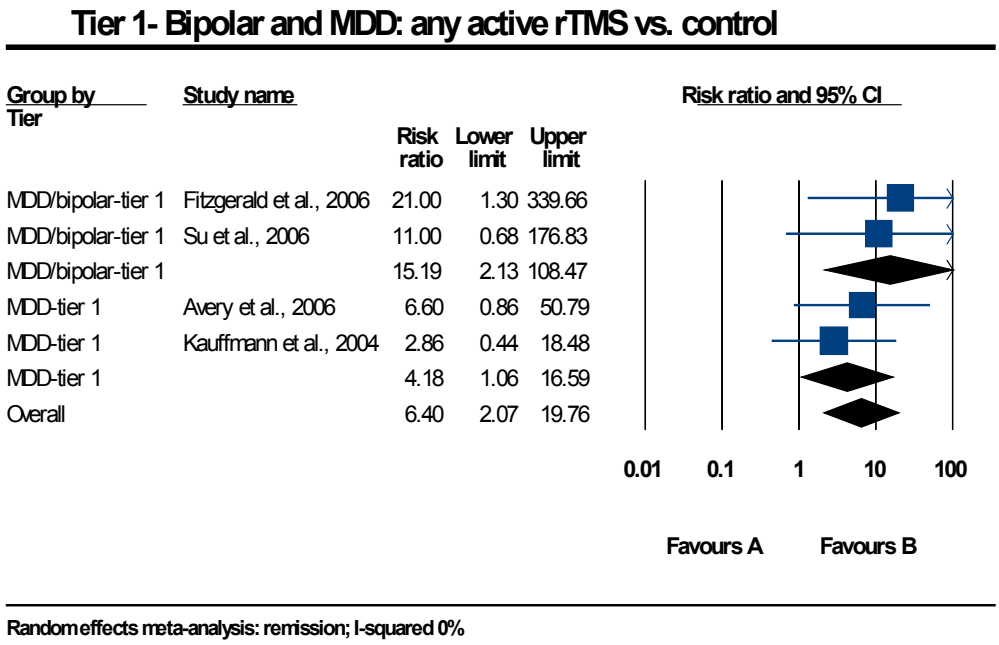


Figure 10. Relative risk meta-analysis of remission rates comparing rTMS vs. sham in an MDD and MDD/Bipolar mix TRD population (Tier 1)




*MDD only.* Consideration of Tier 2 study populations added three additional studies reported in four articles.[69-72] These trials, all switch studies, evaluated rTMS vs. sham stimulation in those with one or more treatment failures (Table 23).

The largest trial, and the only good quality report, was a four week study comparing high frequency rTMS (n = 165 patients) to sham stimulation (n = 160 patients).[72] Patients were required to have failed at least one but not more than four adequate antidepressant treatments in this or the most recent episode OR have failed to tolerate four adequate lifetime medication trials. The groups participating were severely depressed (mean HAM-$D_{17}$ approximately 23). A modified-ITT analysis involving 301 patients at six weeks favored rTMS, which showed a greater decrease in depressive severity (mean HAM-$D_{17}$ decrease of 5.5 vs. 3.3, $P = 0.005$) and a greater response rate (24.5% vs. 13.7%, $P < 0.05$), while there was a trend toward greater remission rates with rTMS (15.5% vs. 8.9%, $P = 0.065$).

A second trial compared 2 weeks of rTMS among four groups: high frequency rTMS (n = 10 patients), low frequency left-sided rTMS (n = 10 patients), low frequency rTMS (n = 10 patients), and sham control (n = 15 patients).[71] All patients had been referred for ECT after having failed an adequate course of an antidepressant medication. The groups involved were severely depressed (mean HAM-$D_{21}$ item ranged between 27 and 28 for each group). It was unclear whether the analysis conducted was ITT or treatment completers. For each outcome, the high frequency rTMS and the low frequency rTMS groups appeared to produce better outcomes than the low frequency left sided RTMS and sham groups. The two former groups produced a greater decrease in depressive severity than the latter two groups (-12.7 and -12.1 vs. 0 and -0.7, respectively; $P = 0.0001$ for each of the two former groups compared to each of the two latter groups). Response rates (50% and 50% vs. 0% and 0%, $P$ = not reported) and remission rates (30% and 10% vs. 0% and 0%; $P$ = not reported) also appeared higher in the same two groups.

Table 23. rTMS vs. sham, Tier 2 MDD only

| **Author, Year**<br>**Endpoint**<br>**Episode Failure**<br>**Quality** | **Intervention and sample size**<br>**Study Details** | **Population characteristics** | **Change in depressive symptoms** | **Response**<br>**Remission** |
|---|---|---|---|---|
| Manes et al., 2001[69] and Moser et al., 2002[70]<br><br>1 week, all reported patients included in analysis<br><br>Failure not required in current episode<br><br>Fair | rTMS (n = 10)<br>High frequency, 5 sessions<br><br>Sham (n = 10)<br><br>Treatment strategy<br>Switch<br><br>Definitions<br>Response definition: 50% reduction in HAM-D and no longer met DSM criteria for major or minor depression<br><br>Remission definition: HAM-D < 8 | Diagnosis<br>Major Depression,%<br>rTMS: 80<br>Sham: 100<br><br>Dysthymia,%<br>rTMS: 20<br>Sham: 0<br><br>**Mean number of failed anti-depressant trials:**<br>rTMS: 4 (2.3)<br>Sham: 4 (1.2)<br><br>Baseline Depression<br>HAM-D NR, mean (SD)<br>rTMS: 22.7 (5.2)<br>Sham: 22.7 (7.1) | HAM-D NR<br>Change, mean (SD)<br>rTMS: -9<br>Sham: -6.5<br>*P* >0.66 | HAM-D NR<br>Response, n (%)<br>rTMS: 3 (30)<br>Sham: 3 (30)<br>*P* = NS<br><br>Remission, n (%)<br>rTMS: 2 (20)<br>Sham: 2 (20)<br>*P* = NR |
| Stern et al., 2007[71]<br><br>2 weeks, all reported patients included in analysis<br><br>Required failure in the current episode<br><br>Fair | **rTMS -1(n = 10)**<br>High frequency,10 sessions<br><br>rTMS -2(n = 10)<br>Low frequency (1Hz), Left-DLPFC, 10 sessions<br><br>rTMS-3 (n = 10)<br>Low frequency, 10 sessions<br><br>Sham (n = 15)<br><br>Treatment strategy<br>Switch<br><br>Definitions<br>Remission definition<br>HAM-$D_{21}$ ≤ 10 | **Mean number of failed anti-depressant trials:**<br>rTMS-1: NR<br>rTMS-2: NR<br>rTMS-3: NR<br><br>Baseline Depression<br>HAM-$D_{21}$, mean (SD)<br>rTMS-1: 27.8 (3.2)<br>rTMS-2: 27.6 (3.9)<br>rTMS-3: 27.9 (3.8)<br>Sham: 27.4 (2.9) | **HAM-$D_{21}$**<br>Change, mean (SD)<br>rTMS-1: -12.7<br>rTMS-2: 0.0<br>rTMS-3: -12.1<br>Sham: -0.7<br>*P* = 0.0001 | **HAM-$D_{21}$**<br>Response, n (%)<br>rTMS-1: 5 (50)<br>rTMS-2: 0 (0)<br>rTMS-3: 5 (50)<br>Sham: 0 (0)<br>*P* = NR<br><br>Remission, n (%)<br>rTMS-1: 3 (30)<br>rTMS -2: 0 (0)<br>rTMS -3: 1 (10)<br>Sham: 0 (0)<br>*P* = NR |
| O'Reardon, 2007[72]<br><br>6 weeks; at week 4, patients not responding left study with LOCF, m-ITT<br><br>Did not require failure in the current episode<br><br>Good | rTMS (n = 165)<br>High frequency, up to 30 sessions<br>**Sham (n = 160)**<br><br>Treatment strategy<br>Switch<br><br>Definitions<br>Remission definition: HAM-$D_{17}$ ≤ 7 | **Mean number of failed anti-depressant trials:**<br>rTMS: 1.6<br>Sham: 1.6<br><br>**Baseline Depression**<br>HAM-$D_{17}$, mean (SD)<br>rTMS: 22.6 (3.3)<br>Sham: 22.9 (3.5) | **HAM-$D_{17}$***<br>Change, mean (SD)<br>rTMS:-5.5<br>Sham:-3.3<br>*P* = 0.005<br><br>*Results based on<br>rTMS: n = 155<br>Sham: n = 146 | **HAM-$D_{17}$***<br>Response, n (%)<br>rTMS: 38 (24.5)<br>Sham : 20 (13.7)<br>*P* < 0.05<br><br>Remission, n (%)<br>rTMS: 24 (15.5)<br>Sham: 13 (8.9)<br>*P* = 0.065 |

DLPFC, dorsolateral prefrontal cortex; DSM, Diagnostic and Statistical Manual; HAM-D, Hamilton Depression Scale; HAM-$D_{21}$, 21 item Hamilton Depression Scale; Hz, hertz; LOCF, last observation carried forward; m-ITT, modified intent-to-treat; n, number; NR, not reported; *P*, p-value; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation.

A brief one week trial compared high frequency rTMS (n = 10 patients) to sham stimulation (n = 10 patients).[69,70] Enrolled patients had moderate to severe depression (mean HAM-D severity approximately 23 in each group). Whether the analysis conducted was ITT or treatment completer was not clear. Results demonstrated no difference between the rTMS and sham groups in the decrease of depressive severity (-9 vs. -6.5, P > 0.66), the rate of response (30 percent in each), or the rate of remission (20 percent in each).

*MDD/Bipolar.* Consideration of Tier 2 added one MDD/Bipolar mix study, a 2 week switch study comparing high frequency rTMS (n = 10 patients) to sham rTMS treatment (n = 10 patients).[76] This study is summarized in Table 24, with a detailed description provided in the evidence tables (Appendix E).All patients had failed at least one adequate antidepressant trial during the current episode except one, who had previously received ECT and had proven treatment resistant to antidepressants in the past). Patients entered into the study with a severe degree of depression (approximately 37 on the HAM-$D_{25}$ item scale in each group). As with the Tier 1 group, the rTMS group had a mean HAM-$D_{25}$ decrease of 14 compared to a decrease of 0.2 in the control group ($P < 0.01$). Response rates also favored rTMS (10 percent vs. 0 percent, $P = 0.09$).

Table 24. rTMS vs. sham, Tier 2 MDD/BP

| **Author, Year**<br>**Endpoint**<br>**Episode**<br>**Failure**<br>**Quality** | **Intervention and sample size**<br>**Study Details** | **Population characteristics** | **Change in depressive symptoms** | **Response**<br>**Remission** |
|---|---|---|---|---|
| Berman et al., 2000[76]<br><br>2 weeks<br><br>Did not require failure in the current episode<br><br>Fair | **rTMS (n = 10)**<br>High frequency, 10 sessions<br><br>Sham (n = 10)<br><br>**Treatment strategy**<br>Switch | **Diagnosis**<br>Bipolar (%)<br>rTMS: 0<br>Sham: 10<br><br>**Mean number of failed anti-depressant trials:**<br>rTMS: 5<br>Sham: 3.5<br>(plus 1 failed augmentation medication each)<br><br>**Baseline Depression**<br>HAM-$D_{25}$, mean<br>rTMS: 37.1<br>Sham: 37.3 | **HAM-$D_{25}$**<br>Change, mean* (SEM)<br>rTMS: -14.0 (3.7)<br>Sham: -0.2 (4.1)<br><br>$P < 0.01$ *adjusted mean decreases based on best fit slopes | **HAM-$D_{25}$**<br>Response, n (%)<br>rTMS: 1 (10)<br>Sham: 0 (0)<br>$P = 0.09$ |

HAM-$D_{25}$, 25 item Hamilton Depression Scale; n, number; rTMS, repetitive transcranial magnetic stimulation; SEM, standard error of measurements.

**Tier 3: Patients with probable treatment resistant depression.** Three trials comparing rTMS with sham stimulation were identified in Tier 3 (Table 25).

*MDD only.* There were no eligible studies.

*MDD/Bipolar*. Three small studies compared rTMS vs. a sham control; these studies are summarized in Table 25 and described in detail in the evidence tables (Appendix E). Two studies reported significantly better outcomes for rTMS, while the third identified a trend in this direction. Results did not vary by strategy. Study duration did not appear to affect outcomes.

Table 25. rTMS vs. sham, Tier 3 MDD/BP

| **Author, year**<br>**Study Design**<br>**Primary endpoint(s)**<br>**Quality**<br>**Tier** | **Intervention and Sample Size**<br>**Study Details** | **Population Characteristics** | **Change in Depressive Symptoms** | **Response**<br>**Remission** |
|---|---|---|---|---|
| Bortolomasi et al. 2006[77]<br><br>1 week, all reported patients included in analysis<br><br>Did not require failure in the current episode<br><br>Tier 3—"drug resistance" not defined<br><br>Fair | **G1: rTMS (n = 12)**<br>High frequency, 5 sessions<br><br>G2: Sham (n = 7)<br><br>**Treatment strategy**<br>Augmentation | **Diagnosis**<br>Bipolar (%)<br>G1: 16.7<br>G2: 14.3<br><br>**Mean number of failed anti-depressant trials:**<br>G1: NR<br>G2: NR<br><br>**Baseline Depression:**<br>HAM-$D_{24}$<br>G1: 25.17<br>G2: NR | **HAM-$D_{24}$**<br>Change, mean (SD)<br>G1: -13.84<br>G2: NR<br>*P* = NR (sig) | **HAM-$D_{24}$**<br>Response, n (%)<br>NR<br>Remission, n (%)<br>NR |
| George et al. 1997[79]<br><br>Crossover, 2 weeks<br><br>Tier 3—all patients had 1+ implied current episode failures<br><br>Fair | **G1: rTMS (n = 12)**<br>High frequency, 10 sessions<br><br>**G2: Sham (n = 12)**<br><br>**Treatment strategy**<br>Mixed-within group difference<br><br>Patients discontinued their (failed) ADs with the exception of 3 patients who were partial responders | **Diagnosis**<br>Bipolar (%)<br>Overall: 8.3<br><br>**Mean number of failed anti-depressant trials:**<br>G1: NR<br>G2: NR<br><br>**Baseline Depression:**<br>HAM-$D_{21}$<br>Overall: 28.5 (4.2) | **HAM-$D_{21}$**<br>Change, mean (SD)<br>G1: -5.25<br>G2: +3.33<br>*P* < 0.03 | **HAM-$D_{21}$**<br>Response, n (%)<br>NR<br>Remission, n (%)<br>NR |
| Moller, 2006[84]<br><br>Crossover, within 1 week of completing 1 week of txt<br><br>Did not require failure in the current episode.<br><br>Tier 3—TRD not defined<br><br>Fair | G1: rTMS (n = 10)<br>High frequency, 5 sessions<br>**G2: Sham (n = 10)**<br><br>**Treatment strategy**<br>Augmentation | **Diagnosis**<br>Bipolar (%)<br>Overall: 20<br><br>**Mean number of failed anti-depressant trials:**<br>G1: NR<br>G2: NR<br><br>**Baseline Depression:**<br>HAM-$D_{17}$<br>Median (range)<br>G1: 20 (13-37)<br>G2: 16 (7-31) | **HAM-$D_{17}$**<br>Change (median)<br>G1: -7<br>G2: -1<br>*P* = 0.075 | **HAM-$D_{17}$**<br>Response, n (%)<br>NR<br>Remission, n (%)<br>NR |

ADs, antidepressants; G, group; HAM-$D_{17}$, 17 item Hamilton Depression Scale; HAM-$D_{21}$, 21 item Hamilton Depression Scale; HAM-$D_{24}$, 24 item Hamilton Depression Scale; n, number; NR, not reported; *P*, p-value; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation; sig, significant; TRD, treatment-resistant depression; txt, treatment.

With the exception of a control arm in one study, all groups were severely depressed. All studies used high frequency rTMS and none required treatment failure in the current episode.

One study compared 5 sessions per week of high frequency rTMS (n = 12 patients) to sham stimulation (n = 7 patients).[77] The authors indicated that patients needed to meet criteria for "drug resistance", but this definition was not provided. Patients enrolled were depressed (mean

HAM-$D_{24}$ for rTMS group = 25.17). Those receiving rTMS had a greater decrease in mean HAM-$D_{24}$ severity than those in the control group (the text states that the difference is statistically significant, but it does not report the test).

The other augment trial was a small randomized crossover study that compared patients (n = 10) receiving one week of high frequency and sham stimulation.[78] Patients were referred to the study because their depression was "drug resistant", and the authors note that "various antidepressants had previously been tried without adequate success". On average, patients entering the study were moderately-to-severely depressed (median HAM-$D_{17}$ for sham = 16 [moderate]) and for rTMS = 20 [severe]). Outcomes suggested benefit for rTMS as measured by mean change in depressive severity (-7 vs. -1), but in this small sample this difference was insignificant ($P = 0.075$).

A third trial tested a mixed strategy that also used a cross-over design. The study (n = 12 patients) compared two week outcomes for patients who received, in randomized order, two weeks of high frequency rTMS and two weeks of sham rTMS stimulation.[79] All patients still met criteria for a major depressive episode despite treatment with an antidepressant, suggesting failure in the current episode. Patients entering the trial were severely depressed (mean HAM-$D_{21}$ score = 28.5). Results from an ITT analysis favored active treatment; the rTMS group had a greater mean change in depressive severity (-5.25 vs. + 3.33, $P < 0.03$).

**Tiers 1-3 combined**. Nineteen studies comparing rTMS with sham rTMS stimulation were identified.[17,61-77,79] The majority of studies for this comparison found rTMS resulted in significantly greater efficacy as measure by change in depressive severity, response, and remission. Other studies did not report of tests of statistical significance or were underpowered to detect differences between groups. Of the studies included, twelve fell into Tier 1, four into Tier 2, and three into Tier 3. Eleven of the studies included only patients with MDD[61-72] while eight of the studies allowed a small number of patients with BP to participate (all studies, ≤ 20% sample had BP).[17,73-77,79] Differences in efficacy by tier and inclusion of patients with BP disorder were assessed via stratified meta-analyses.

*Meta-analytic Synthesis of MDD Outcomes (Tiers 1, 2, and 3 Combined).* Meta-analyses combining studies from all Tiers involved (only Tier 1 and Tier 2 studies, as there were no Tier 3 studies identified) supported the benefit of rTMS over sham control and were consistent with Tier 1 analyses. The weighted mean difference in HAM-D depressive severity was -3.57 (95% CI, -4.60 to -2.54) (Figure 11). The pooled relative risk indicated that patients receiving rTMS were approximately twice as likely to respond as those receiving placebo (pooled relative risk 2.06, 95% CI, 1.40 to 3.02) (Figure 12), which translates into a Number Needed to Treat of 6 (95% CI, 4—10). Remission rates also favored rTMS. The pooled relative risk for remission was 2.025 (95% CI, 1.180 to 3.477), which translates to a Number Needed to Treat of 11 (95% CI, 7—29) (Figure 13).

Combining these three Tiers for MDD only populations provided a more conservative point estimate and a narrower confidence interval for each of the three outcomes than the quantitative syntheses for Tier 1 MDD only.

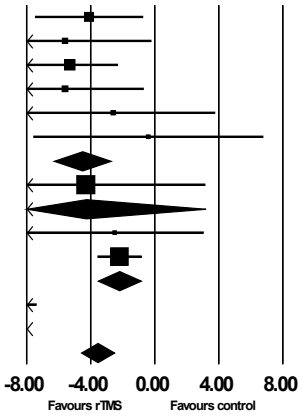
*Meta-analytic Synthesis of MDD/Bipolar Mix Outcomes (Tiers 1, 2, and 3 Combined.)* Combining this data with Tier 1 results continued to support benefit for rTMS. For changes in depressive severity as measured by the mean HAM-D difference, patients receiving rTMS on average had a decrease of nearly nine points relative to sham control (-8.71, 95% CI, -10.34 to -7.08). Because sample sizes were small and responses to placebo varied in the small control

groups, the heterogeneity was high and our estimates are uncertain with respect to the magnitude of changes on the HAM-D. Given this uncertainty, we are not including the forest plot.

Figure 11. Effect size meta-analysis comparing rTMS vs. control on the HAM-D for TRD-related studies (all Tiers) in MDD-only population

## Tier 1 & tier 2: any active rTMS vs. control

| Group by Tier | Study name | Statistics for each study: Difference in means | Lower limit | Upper limit | p-Value |
|---|---|---|---|---|---|
| tier 1-high | Avery et al., 2006 | -4.10 | -7.48 | -0.72 | 0.02 |
| tier 1-high | Boutros et al., 2002 | -5.60 | -11.00 | -0.20 | 0.04 |
| tier 1-high | Garcia-Toro et al., 2001 | -5.30 | -8.30 | -2.30 | 0.00 |
| tier 1-high | Garcia-Toro et al., 2006 | -5.60 | -10.52 | -0.68 | 0.03 |
| tier 1-high | Holtzheimer et al., 2004 | -2.60 | -8.98 | 3.78 | 0.42 |
| tier 1-high | Padberg et al., 1999 | -0.40 | -7.59 | 6.79 | 0.91 |
| tier 1-high | | -4.53 | -6.31 | -2.76 | 0.00 |
| tier 1-low | Kauffmann et al., 2004 | -4.30 | -11.77 | 3.17 | 0.26 |
| tier 1-low | | -4.30 | -11.77 | 3.17 | 0.26 |
| tier 2-high | Manes et al., 2001 | -2.50 | -8.07 | 3.07 | 0.38 |
| tier 2-high | O'Reardon et al., 2007 | -2.20 | -3.58 | -0.82 | 0.00 |
| tier 2-high | | -2.22 | -3.56 | -0.87 | 0.00 |
| tier 2-low/high | Stern et al., 2007 | -11.70 | -16.03 | -7.37 | 0.00 |
| tier 2-low/high | | -11.70 | -16.03 | -7.37 | 0.00 |
| Overall | | -3.57 | -4.60 | -2.54 | 0.00 |

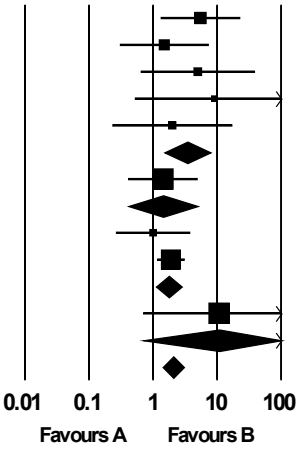



**Random effects meta-analysis: changes on HAM-D; I-squared 57%**

Figure 12. Relative risk meta-analysis of response rates comparing rTMS vs. sham for TRD-related studies (all Tiers) in MDD-only population

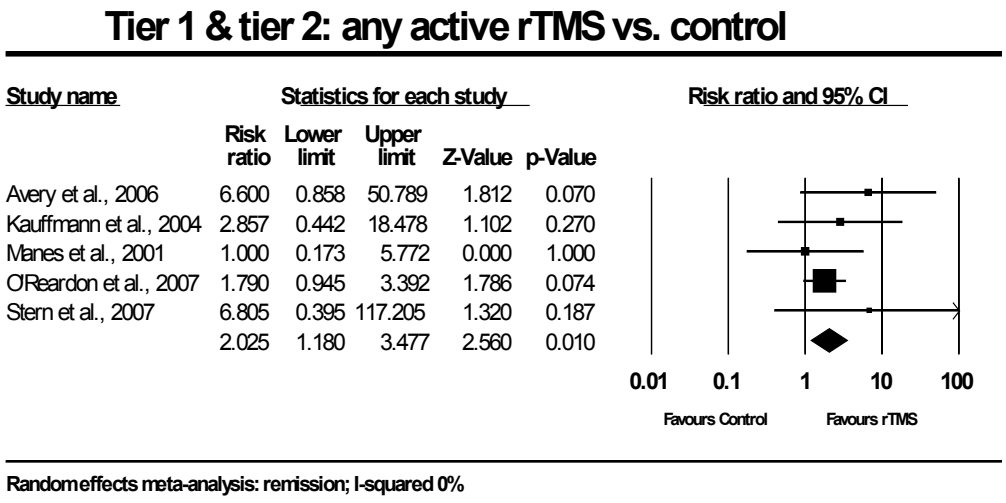
## Tier 1 & tier 2: any active rTMS vs. control

| Group by Tier | Study name | Statistics for each study: Risk ratio | Lower limit | Upper limit | p-Value |
|---|---|---|---|---|---|
| tier 1-high | Avery et al., 2006 | 5.50 | 1.31 | 23.03 | 0.02 |
| tier 1-high | Boutros et al., 2002 | 1.50 | 0.30 | 7.43 | 0.62 |
| tier 1-high | Garcia-Toro et al., 2001 | 5.00 | 0.64 | 39.06 | 0.12 |
| tier 1-high | Garcia-Toro et al., 2006 | 9.00 | 0.52 | 156.67 | 0.13 |
| tier 1-high | Holtzheimer et al., 2004 | 2.00 | 0.23 | 17.34 | 0.53 |
| tier 1-high | | 3.43 | 1.49 | 7.86 | 0.00 |
| tier 1-low | Kauffmann et al., 2004 | 1.43 | 0.41 | 4.99 | 0.58 |
| tier 1-low | | 1.43 | 0.41 | 4.99 | 0.58 |
| tier 2-high | Manes et al., 2001 | 1.00 | 0.26 | 3.81 | 1.00 |
| tier 2-high | O'Reardon et al., 2007 | 1.90 | 1.16 | 3.12 | 0.01 |
| tier 2-high | | 1.76 | 1.10 | 2.80 | 0.02 |
| tier 2-high/low | Stern et al., 2007 | 10.76 | 0.70 | 166.21 | 0.09 |
| tier 2-high/low | | 10.76 | 0.70 | 166.21 | 0.09 |
| Overall | | 2.06 | 1.40 | 3.02 | 0.00 |



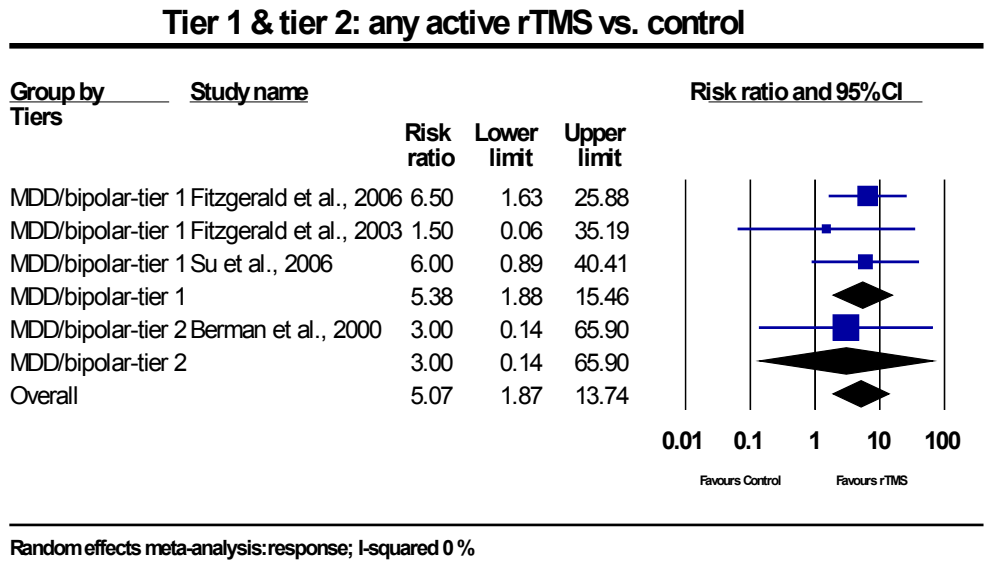

**Random effects meta-analysis: remission; I-squared 0%**

Figure 13. Relative risk meta-analysis of remission rates comparing rTMS vs. sham for TRD-related studies (all Tiers) in MDD-only population




Response rates also favored rTMS, with rTMS groups being over 5 times as likely to achieve response (random effects relative risk 5.07, 95% CI, 1.87 to 13.74) (Figure 14), leading to a Number Needed to Treat of 3 (95% CI, 1-14). With only two studies reporting remission for all Tiers combined (which only included Tier 1 and Tier 2 for this outcome), we did not quantitatively synthesize this information; both studies indicated greater remission rates for rTMS over sham.

Figure 14. Relative risk meta-analysis of response rates comparing rTMS vs. sham for TRD-related studies (all Tiers) in an MDD/Bipolar mix population




Compared to the meta-analytic synthesis of Tier 1 MDD/Bipolar mix studies, the combination of Tiers 1-3 produced nearly identical point estimates for change in depressive severity and response rate and narrower confidence intervals.

*Meta-analytic synthesis of MDD and MDD/bipolar mix outcomes (Tiers 1, 2, and 3 Combined)*. Meta-analyses combining studies from all Tiers involved and including both MDD

and MDD/Bipolar mix populations continued to support the benefit of rTMS over sham control and were consistent with Tier 1 combined analyses. The weighted mean difference in HAM-D depressive severity was -4.98 (95% CI, -5.86 to -4.11). Because sample sizes of individual studies were small and responses to placebo varied in the small control groups, the heterogeneity was high and our estimates are uncertain with respect to the magnitude of changes on the HAM-D. Given this uncertainty, we are not including the forest plot.

The pooled relative risk indicated that patients receiving rTMS were over twice as likely to respond than those receiving placebo (pooled relative risk 2.31, 95% CI, 1.62 to 3.30) (Figure 15), which translates into a Number Needed to Treat of 5 (95% CI, 4—10). Remission rates also favored rTMS. The pooled relative risk for remission was 2.33 (95% CI, 1.39 to 3.93), which translates to a number needed to treat of 8 (95% CI, 6—15) (Figure 16).

Figure 15. Relative risk meta-analysis of response rates comparing rTMS vs. sham for TRD-related studies (all Tiers) in MDD and MDD/Bipolar mix populations

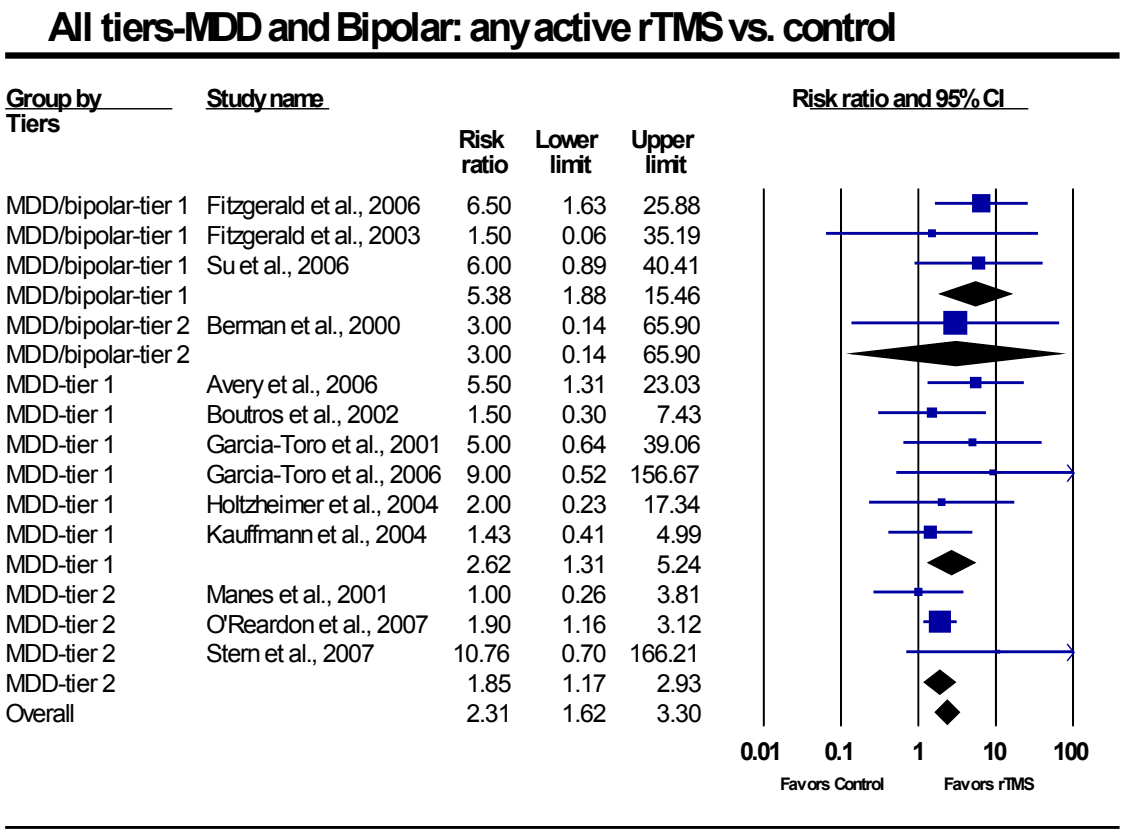
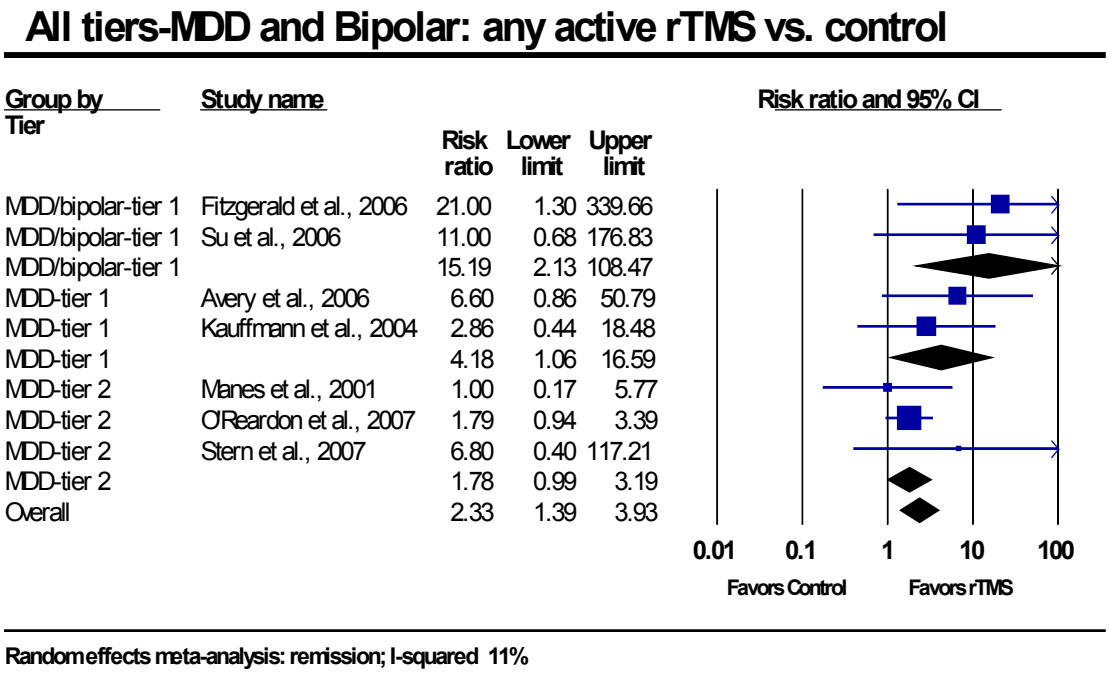


Figure 16. Relative risk meta-analysis of remission rates comparing rTMS vs. sham for TRD-related studies (all Tiers) in MDD and MDD/Bipolar mix populations




Compared to Tier 1 syntheses of MDD and MDD/Bipolar populations combined, the consideration of all three tiers provided more conservative point estimates and narrower confidence intervals for each outcome.

Indeed, the meta-analytic results for MDD and MDD/Bipolar mix for all tiers combined were most nearly identical to results for the Tier 1 MDD only group, our main population of interest. For the two comparable outcomes, point estimates were similar with narrower confidence intervals. For change in depressive severity, the all tier MDD and MDD/Bipolar mix results were -4.98 (95% CI, -5.86 to -4.11), compared to -4.52 (95% CI, -6.25 to -2.80) for Tier 1 MDD only. Rates of response were similarly equivalent (2.31, 95% CI, 1.62 to 3.30 for all Tiers MDD and MDD/Bipolar combined vs. 2.62, 95% CI, 1.31 to 5.24 for Tier 1 MDD only).

*Summary of key variables*. Consideration of all Tiers together for the combined MDD and MDD/Bipolar mix populations provided results consistent with those from Tier 1 alone combined but with more conservative point estimates and narrower confidence intervals, suggesting that results from analyses of studies from all tiers reflect what can be expected in TRD (Tier 1) populations. This finding of all tier evidence reflecting what was found with Tier 1 alone held whether the population included was MDD only, or MDD/Bipolar mix, respectively.

Results from Tiers 1-3 for MDD only were in the same direction as and of similar magnitude to those for the Tiers 1-3 MDD/Bipolar mix populations. For each outcome, point estimates for the MDD/Bipolar mix group were higher with wider confidence intervals, but they were not significantly different from the MDD only group. When these results were combined, confidence intervals were either equivalent or narrower than when the diagnostic samples were split, suggesting that combining MDD and MDD/Bipolar presentations was reasonable.

Only three studies required an antidepressant failure in current episode;[61,71,73] there was no clear variation in treatment efficacy between these studies and those not requiring a current episode failure.

At baseline almost all study populations (13 of 19) had severe depression, while[17,63-65,67,68,71-76,79] five study populations had moderate-to-severe depression[61,62,69,70,77,84] and one study population severity was not reported.[66] With little variation in depression severity, we were unable to detect any differences by this variable.

In this comparison, there were eleven studies using an augmentation strategy,[17,61-63,65,68,73-75,77,84] five using a switch strategy,[64,69-72,76] and three using a mixed strategy with within group differences.[66,67,79] We were unable to detect clear differences by treatment characteristics (i.e. pharmacotherapy strategy, rTMS frequency, or treatment duration) through qualitative analysis due to other potentially confounding variables resulting from study design or participant characteristics.

## Vagus Nerve Stimulation Plus Treatment as Usual vs. Treatment as Usual

**Tier 1: Patients with two or more failures.** One trial comparing VNS plus treatment as usual with treatment as usual was identified in Tier 1 (Table 26).

Table 26. VNS vs. sham

| Author, Year<br>Endpoint<br>Episode<br>Failure<br>Quality | Intervention and Sample Size<br>Study Details | Population Characteristics | Change in Depressive Symptoms | Response Remission |
|---|---|---|---|---|
| | | **VNS vs. Sham** | | |
| Rush et al., 2005[80]<br><br>10 weeks, m-ITT/per medication protocol<br><br>Required failure in the current episode<br><br>Good | VNS (n = 119)<br>10 weeks of VNS therapy with continued medications.<br><br>Sham (n = 116)<br><br>Treatment strategy<br>Augmentation | **Diagnosis**<br>Bipolar (%)<br>VNS: 11.7<br>Sham: 9.1<br><br>**Number of failed anti-depressant trials (% ≥ 4):**<br>ECT: 46.4%<br>rTMS: 40.0%<br><br>**Baseline Depression**<br>HAM-$D_{24}$, mean (SD)<br>VNS: 28.8 (5.3)<br>Sham: 29.7 (5.2) | HAM-$D_{24}$*<br>% Change, mean (SD)<br>VNS: -16.3 (28.1)<br>Sham: -15.3 (25.5)<br>*P* = 0.639<br><br>*based on G1 n = 112, G2 n = 110 | HAM-$D_{24}$*<br>Response, n (%)<br>VNS: 17 (15.2)<br>Sham: 11 (10.0)<br>*P* = 0.25 |

G, group; HAM-$D_{24}$, 24 item Hamilton Depression Scale; *P*, p-value; SD, standard deviation; VNS, Vagus Nerve Stimulation.

*MDD Only.* There were no eligible studies.

*MDD/Bipolar* . One good ten week study compared VNS (n = 119 patients) to a control group (n = 116 patients.[80] This study is summarized in Table 26 with a detailed description provided in the evidence tables (Appendix E).[‡‡‡] The control group had the surgical procedure to implant the VNS device, but they did not have the device turned on for the sessions. Patients were required to have had an unsatisfactory response to at least two adequate trials of antidepressant medication, but not more than six failures, for the current episode. Over 40% of the sample had failed 4 or more prior antidepressant trials, indicating a high degree of treatment

[‡‡‡] Appendixes are cited in this report and provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/___/___.pdf.

resistance. The two groups entering into this study were severely depressed, with a mean HAM-D24 score of 28.8 in the VNS group and 29.7 in the control. In a modified-ITT analysis that excluded those noncompliant with the medication protocol, the results did not demonstrate a statistically significant difference between the two groups for the primary outcome ($HAM\text{-}D_{24}$). No differences were found in the percent change in depressive severity (-16.3% for VNS vs. -15.3% for control, $P = 0.639$) or the response rates (15.2% vs. 10.0%, $P = 0.25$). Of note, response rates for a secondary outcome, the 30-item Inventory for Depressive Symptomatology-self report, favored VNS (17.0% vs. 7.3%, $P = 0.032$).

**Tier 2: Patients with one or more failures.** There were no eligible studies.

**Tier 3: Patients with probable treatment resistant depression.** There were no eligible studies.

**Tiers 1-3: Combined results.** Only one study comparing VNS to sham stimulation was identified.[80] This study is described in the section above.

## Psychotherapy Plus Treatment as Usual vs. Treatment as Usual

**Tier 1: Patients with two or more failures.** There were no eligible studies.

**Tier 2: Patients with one or more failures.** One trial comparing psychothrerapy with a control group were identified in Tier (Table 27).

*MDD Only* .Three tier 2 studies[59,60,81,82] comparing psychotherapy to a control group were identified (Table 27). All indicated improvement with CBT added to current care. Only one of these studies received a good quality rating.[59,60] Two studies used augmentation,[59,60,81] while the third used an unlimited strategy (augmenting UC with CBT);[82] the type of treatment strategy produced no clear variation in outcome. The presence of treatment failure in the current episode did not clearly influence outcome. The duration of the trials (all 16-20 weeks) did not vary. Groups in all studies were moderately depressed.

One good 20 week RCT (described in two articles) compared 16 sessions of cognitive therapy and clinical management (equivalent to UC; n = 80 patients) to clinical management alone (n = 78 patients).[59,60] In each case, clinical management consisted of a visit with a psychiatrist every 4 weeks with minor medication adjustments to an antidepressant medication regimen allowed. Patient's entered into study having residual depressive symptoms ($HAM\text{-}D_{17} \geq 8$) despite having received greater than 4 weeks of adequate antidepressant treatment. Depression in both groups was mild (mean $HAM\text{-}D_{17}$ for the two groups was 12.1-12.2). In an ITT analysis, there was no difference in the mean decrease in depressive severity (CBT plus clinical management -3.4 vs. clinical management alone—2.8, $P$ = NS). Remission was defined more stringently as a $HAM\text{-}D_{17}$ score $\leq 7$ at *two* consecutive visits 4 weeks apart; using this definition, remission rates were greater in the CBT plus clinical management (24 percent vs. 13 percent, $P = 0.03$).

One trial compared a four month treatment of Cognitive Behavioral Therapy (CBT) plus UC (n = 14 patients) to UC (n = 11 patients).[82] Mean depressive severity at baseline as measured by the BDI was 31.1 for CBT plus UC vs. 26.8 for UC, consistent with depression that was moderate-to-severe. UC in each group resulted in unlimited medication strategy. In an ITT analysis, the CBT plus UC group reduced depressive severity as measured by the BDI by an average of 11.2 points more than the UC group (95% CI, -19.3 TO -3.1). Also, the CBT plus UC group had 8 patients meeting response criterion, compared to none in the UC group.

Table 27. Psychotherapy vs. control, Tier 2 studies

| Author, Year<br>Endpoint<br>Episode Failure<br>Quality | Intervention and Sample Size<br>Study Details | Population Characteristics | Change in Depressive Symptoms | Response Remission |
|---|---|---|---|---|
| Harley, 2008[81]<br>16 weeks, completers analysis<br>Failure not required in current episode<br>Fair | Dialectical Behavioral Therapy Skills Training (n = 13)<br>16 sessions<br>Waitlist Control (n = 11)<br>Treatment strategy<br>Augmentation<br>Definitions<br>Remission definition: HAM-$D_{17}$ score ≤ 7 | **Mean number of failed anti-depressant trials:**<br>DBT: NR<br>rTMS: NR<br>Baseline Depression<br>HAM-$D_{17}$, mean (SD)<br>CBT: 16.15 (4.47)<br>Control: 18.64 (4.72) | HAM-$D_{17}$*<br>Change, mean (SD)<br>CBT: -5.6<br>Control: -1.78<br>*P < 0.05*<br>* results based on completers (G1: n = 10, G2: n = 9) | HAM-$D_{17}$<br>Remission (%)<br>CBT: 3 (23.1)<br>Control: 0 (0)<br>*P* = NR |
| Paykel, 1999[59] and Scott, 2000[60]<br>20 weeks<br>Required failure in the current episode<br>Good | Cognitive Therapy plus Clinical Management (n = 80)<br>16 sessions<br>Clinical Management (n = 78)<br>Treatment strategy<br>Primarily augmentation with minor medication dose adjustments were allowed.<br>Definitions<br>Remission definition: HAM-$D_{17}$ score ≤ 7 at 2 consectutive ratings 4 wks apart | **Mean number of failed anti-depressant trials:**<br>DBT: NR<br>rTMS: NR<br>Baseline Depression<br>HAM-$D_{17}$, mean (SD)<br>CBT: 12.2 (2.9)<br>Control: 12.1 (2.7) | HAM-$D_{17}$<br>Change, mean (SD)<br>CBT: -3.4<br>Control: -2.8<br>*P* = NS | HAM-$D_{17}$<br>Remission, n (%)<br>CBT: 19 (24)<br>Control: 10 (13)<br>Hazard Ratio for remission 2.42 (95% CI: 1.08 to 5.45), *P* = 0.03 |
| Wiles et al. 2008[82]<br>4 mos<br>Required failure in the current episode<br>Fair | CBT (n = 14)<br>12-20 sessions<br>Usual care (n = 11)<br>No restrictions<br>Treatment Strategy<br>Unlimited | **Mean number of failed anti-depressant trials:**<br>DBT: NR<br>rTMS: NR<br>Baseline Depression<br>BDI, mean (SD)<br>CBT: 31.1 (8.5)<br>Usual care: 26.8 (6.8) | BDI<br>CBT scores decreased by an average of 11.2 points more than Usual Care (95%CI, -19.3 to -3.1) | BDI<br>Response, n (%)<br>CBT: 8 (57.1)<br>Usual care: 0 (0.0)<br>*P* = NR |

BDI, Beck Depression Inventory; CBT, cognitive behavioral therapy; CI, confidence interval; DBT, Dialectical Behavioral Therapy; G, group; HAM-$D_{17}$, 17 item Hamilton Depression Scale; mos, months; n, number; NR, not reported; NS, not significant; *P*, p-value; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation; wks, weeks.

One four month trial compared a distinct form of CBT that involves both group and individual treatments called Dialectical Behavioral Therapy (DBT) (n = 13 patients) to a wait list control (n = 11).[81] The two participating groups had moderate depressive severity at study enrollment (HAM-$D_{17}$ scores averaged 16.15 for DBT group and 18.64 for waitlist control).. In a treatment completer analysis at 4 months, the DBT group (n = 10) had greater decrease in depressive severity than the waitlist group (n = 9) (-5.6 vs. -1.78, $P < 0.05$) and were more likely to achieve remission (23.1 percent vs. 0 percent).

We did not quantitatively synthesize these results.

*MDD/Bipolar.* There were no eligible studies.

**Tier 3: Patients with probable treatment resistant depression.** There were no eligible studies.

**Tiers 1-3: Combined results.** Three Tier 2 studies comparing psychotherapy to a control group were identified. Almost all efficacy measures indicated greater efficacy for the psychotherapy group. In one study the greater comparative decrease in depressive severity for the psychotherapy group did not reach statistical significance, however, in this same study, greater efficacy was seen for the psychotherapy group in remission.[59,60] One population characteristic which may be related to the lack of significant difference seen in the change of depressive severity is the severity of depression at baseline for study populations. On average patients in this study had mild depressive severity, limiting the impact that the intervention could have had on depression severity.[59,60] Participants in the other two studies had moderate and moderate-to-severe depression and did see significant difference between group on this measure.[81,82] With only three studies identified for this comparison, it is difficult to determine how study design, participant, or treatment characteristics may have affected treatment efficacy. All three studies fell into Tier 2 and two of the trials[59,60,82] required a failure in the current episode. All patients had MDD. Two of the studies used augmentation strategies[59,60,81] while the other did not limit the pharmacotherapy strategies of participants.[82] Duration and method of psychotherapeutic interventions were similar across studies.

# Key Question 1b: Comparisons Involving Pharmacologic Interventions for Acute Phase Treatment—Overview of Comparisons

In this section, we assess how nonpharmacologic treatments compare with pharmacological treatments in efficacy or effectiveness in treating acute-phase depressive symptoms in patients with TRD; these comparisons can help place nonpharmacologic treatment for TRD within the context of pharmacologic outcomes. First, we review the literature that directly compares nonpharmacologic and pharmacologic interventions for TRD, using the same approach we did in KQ 1a: categorizing first by intervention comparison, next by Tier, and then by MDD vs. the MDD/Bipolar mix, while considering the role of the same key elements on treatment outcome.

For nonpharmacologic vs. pharmacologic comparisons, we identified 4 studies. One study compared ECT vs. pharmacotherapy, one study compared rTMS plus pharmacotherapy vs. pharmacotherapy, and two compared CBT vs. pharmacotherapy. Only one of these studies involved a TRD (Tier 1) population; enrolling an MDD/Bipolar mix sample, it provided data showing that switching to ECT provided a greater decrease in depressive severity than switching to a new pharmacotherapy.[58]

Considering Tier 2 studies added two trials comparing CBT vs. pharmacotherapy, both in MDD only populations. These two studies involved moderately depressed groups and provided data showing that CBT was no different than medication treatments for a variety of treatment strategies.[85,86] We could not make any conclusions about the impact of Tier definition, diagnosis, depressive severity, treatment strategy, treatment characteristics, or treatment failure in the current episode.

For pharmacologic vs. pharmacologic treatments, we identified nine trials that used a variety of pharmacologic treatment strategies to treat TRD. All involved patients that were severely depressed. Response rates for the pharmacologic options did not clearly differ from rates seen with rTMS, nor did they appear different from the two studies reporting CBT outcomes vs. medications, although they did appear to have poorer outcomes than one study comparing ECT

vs. medications. Finally, mean remission rates for pharmacologic options were similar to the reported single study nonpharmacologic rate for rTMS (in a study vs. sham).

# Key Question 1b: Comparisons Involving Pharmacologic Interventions for Acute Phase Treatment—Overview of Nonpharmacologic vs. Pharmacologic Treatments

Only four studies providing nonpharmacologic vs. pharmacologic treatments were available (Table 28).[58,85-87] Having such a limited database prevented a consideration of the effect on outcome of which Tier of evidence was used, whether the population was MDD vs. MDD/Bipolar mix, the degree of depressive severity, the type of treatment strategy, the type of treatment characteristics, and whether the treatment failure was in the current episode.

Strength of evidence assessments were made for three outcomes: change in depressive severity, response rates, and remission rates. We first will present the strength of evidence for Tier 1 studies alone, and then present strength of evidence for all three tiers considered together. When possible, within each comparison we report results by treatment strategy since this is a fundamental aspect of the antidepressant therapy.

Table 28. Number of good and fair quality studies by comparison, tier, and diagnostic mix

| | Diagnosis | |
|---|---|---|
| **Comparison, Tier** | **MDD only** | **MDD and Bipolar Disorder** |
| **ECT vs. pharmacotherapy** | | |
| Tier 1 (≥ 2 failures) | 0 | 1 |
| **rTMS + pharmacotherapy vs. pharmacothereapy** | | |
| Tier 2 (≥ 1 failures) | 1 | 0 |
| **Psychotherapy vs. pharmacotherapy** | | |
| Tier 2 (≥ 1 failures) | 2 | 0 |

A single MDD/Bipolar mix study[58] suggested better outcome for ECT compared with pharmacologic treatment; another found better outcomes for rTMS and pharmacotherapy combined compared with pharmacotherapy alone.[87] Two studies found no difference between CBT and pharmacologic options.[85,86]

## Strength of Evidence: Tier 1 Only

Only one study providing nonpharmacologic vs. pharmacologic treatments was available.[58]. Having such limitations prevented consideration of the effect on outcome whether the population was MDD vs. MDD/Bipolar mix, the degree of depressive severity, the type of treatment strategy, the type of treatment characteristics, and whether the treatment failure was in the current episode.

Data were available to allow strength of evidence assessments for two outcomes: change in depressive severity and response rates (Table 29). This single trial provided low strength of evidence that ECT produced better outcomes than medications in Tier 1 MDD/Bipolar mix population, the study did not address remission rates.[58]

Table 29. ECT vs. pharmacotherapy: strength of evidence—Tier 1 only

| Comparison | Number of Studies; Subjects | Risk of bias Design/ Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| Change in depressive severity | 1; 39 | Medium/ High<br>RCT<br>1 fair | Unknown | Direct | Imprecise | 1 trial using a switch strategy found ECT resulted in significantly greater improvements symptom severity compared to pharmacotherapy (paroxetine).<br>Low |
| Response | 1; 39 | Medium/ High<br>RCT<br>1 fair | Unknown | Direct | Imprecise | 1 trial using a switch strategy found ECT produced significantly greater response rates compared to pharmacotherapy (paroxetine).<br>Low |
| Remission | 0; 0 | — | — | — | — | — |

ECT, electroconvulsive therapy; RCT, randomized controlled trial.

## Strength of Evidence: Considering all Tiers

Considering evidence from all tiers added three new studies, both providing data for two new comparison (CBT vs. pharmacotherapy and rTMS plus pharmacotherapy vs. pharmacotherapy). These two additional studies provided low strength of evidence that CBT was not different from medication treatments for a variety of treatment strategies in Tier 2 MDD for each of the three outcomes Table 30).[85,86] In the single study comparing rTMS plus pharmacotherapy to pharmacotherapy alone, rTMS produced greater a decrease in depression severity (Table 30).[87] The data did not affect the finding of a low strength of evidence that ECT produced a greater decrease in depression severity and better response rates than medications in Tier 1 MDD/Bipolar mix population (Table 29).[58]

# Key Question 1b: Comparisons Involving Pharmacologic Interventions for Acute Phase Treatment—Key Points of Nonpharmacologic vs. Pharmacologic Treatments

Only four trials provided a direct comparison of nonpharmacologic and pharmacologic treatment for TRD. The limited number of comparisons prevented any firm conclusions regarding the effect on outcome of the tier level of evidence used, whether the population was MDD vs. MDD/Bipolar mix, the degree of depressive severity, the type of treatment strategy, the type of treatment characteristics, or whether the treatment failure was in the current episode.

Table 30. rTMS + pharmacotherapy vs. pharmacotherapy: strength of evidence—Tiers 1-3

| Comparison | Number of Studies; Subjects | Risk of bias Design/ Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| Change in depressive severity | 1; 41 | Medium /High<br><br>RCT<br><br>1 fair | Unknown | Indirect | Imprecise | 1 fair trial found patients starting new AD (combination strategy) and rTMS had a significantly greater decrease in depressive severity compared to pharmacotherapy and sham (as measured by effect size).<br>Low |
| Response | 0; 0 | — | — | — | — | — |
| Remission | 0; 0 | — | — | — | — | — |

AD, antidepressants; rTMS, repetitive transcranial magnetic stimulation.

Table 31. Psychotherapy vs. pharmacotherapy: strength of evidence—Tiers 1-3

| Comparison | Number of Studies; Subjects | Risk of bias Design/ Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| Change in depressive severity | 2; 317 | Medium<br><br>RCT<br><br>2 fair | Unknown | Direct | Precise | 1 fair trial using both switching and augmenting strategies found no significant differences in change in depressive severity between CBT and pharmacotherapy.<br>1 small fair trial using a mixed strategy and conducting a treatment completer analysis did not report statistical significance<br>Low |
| Response | 1; 304 | Medium<br><br>RCT<br><br>1 good | Unknown | Direct | Precise | 1 good trial using both switching and augmenting strategies found no differences in response rates between CBT and pharmacotherapy.<br>Low |
| Remission | 1; 304 | Medium<br><br>RCT<br><br>1 good | Unknown | Direct | Precise | 1 good trial using both switching and augmenting strategies found no differences between CBT and pharmacotherapy<br>Low |

CBT, cognitive behavioral therapy; NR, not reported; RCT, randomized controlled trial.

## ECT vs. Pharmacotherapy

One study comparing ECT with pharmacotherapy found a greater change in depressive severity and a higher rate of response for participants in the ECT group.

## rTMS Plus Pharmacotherapy vs. Pharmacotherapy

One study comparing rTMS plus pharmacotherapy with sham treatment and pharmacotherapy (paroxetine) found a greater change in depressive severity for participants in the active rTMS group.

## CBT vs. Pharmacotherapy

One study comparing CBT with pharmacotherapy found no differences in change in depressive severity, rate of response, or rate of remission between groups. A second study, with a small sample (N = 13), showed a difference in change in depressive severity but did not report the test of statistical significance.

# Key Question 1b: Comparisons Involving Pharmacologic Interventions for Acute Phase Treatment—Detailed Analysis of Nonpharmacologic vs. Pharmacologic Treatments

## ECT vs. pharmacotherapy

**Tier 1: Patients with two or more failures.** One study comparing ECT to pharmacotherapy in an MDD/Bipolar mix population was identified for tier 1 (Table 32), finding greater improvement in severity and response for patients receiving ECT vs. paroxetine.

*MDD only.* There were no eligible studies

*MDD/Bipolar.* One four week trial compared outcomes for right-sided unilateral ECT (n = 21 patients) with paroxetine (n = 18 patients).[58] All patients discontinued current antidepressant therapy while patients in the paroxetine group initiated pharmacotherapy. In the ECT group, 9.5 percent of patients (n = 2) had bipolar illness, while 16.7 percent (n = 3) had bipolar illness in the medication group. Patients were severely depressed (mean HAM-$D_{21}$ scores were 31.1 in ECT group (SD 4.9) and 32.8 (SD 5.4) in pharmacotherapy group). The ECT group experienced a greater decrease in depressive severity (-18.6 vs. -9.6, $P = 0.001$) and a greater response rate (71.4 percent vs. 27.8 percent, $P = 0.006$) than the paroxetine group.

**Tier 2: Patients with one or more failures.** There were no eligible studies.

**Tier 3: Patients with probable treatment resistant depression.** There were no eligible studies.

**Tiers 1-3: Combined.** Only one study comparing ECT to pharmacotherapy was identified;[58] this study is described in the section above.

Table 32. ECT vs. pharmacotherapy, Tier 1

| Author, year<br>Endpoint<br>Episode<br>Failure<br>Quality | Intervention and sample size<br>Study Details | Population characteristics | Change in depressive symptoms | Response<br>Remission |
|---|---|---|---|---|
| | | **ECT vs. Pharmacotherapy** | | |
| Folkerts et al. 1997[58]<br>End of study phase (2-4 weeks), per protocol analysis<br>Did not require failure in the current episode<br>Fair | **ECT (n = 21*)**<br>Right unilateral, mean txts = 7.2 sessions (2-3 weeks)<br>Pharmacotherapy - Paroxetine (n = 18*)<br>40mg (max 50mg/d, mean 44 mg/day)<br>*per protocol<br>**Treatment** Statregy<br>Switch | **Diagnosis**<br>Bipolar (%)<br>ECT: 9.5<br>Pharm: 16.7<br>**Mean number of failed anti-depressant trials:**<br>ECT: 4.9<br>Pharm: 4.3<br>**Baseline Depression**<br>HAM-$D_{21}$, mean (SD)<br>ECT: 31.1 (4.9)<br>Pharm: 32.6 (5.4) | **HAM-$D_{21}$**<br>Change, mean (SD)<br>ECT: -18.6<br>Pharm: -9.6<br>*P = 0.001* | **HAM-$D_{21}$**<br>Response, n (%)<br>ECT: 15 (71.4)<br>Pharm: 5 (27.8)<br>*P* = 0.006 |

ECT, electroconvulsive therapy; G, group; HAM-$D_{21}$, 21 item Hamilton Depression Scale; max, maximum, mg, milligram; mg/d, milligram per day; n, number; *P*, p-value; Pharm, Pharmacotherapy; SD, standard deviation; txt(s), treatment(s).

## Repetitive Transcranial Magnetic Stimulation Plus Pharmacotherapy vs. Sham Control Plus Pharmacotherapy

**Tier 1: Patients with two or more failures.** There were no eligible studies.

**Tier 2: Patients with one or more failures.** Only one study comparing rTMS plus pharmacotherapy to pharmacotherapy was identified. This study is presented in Table 33.

Table 33. Tier 2 studies comparing rTMS + pharmacotherapy vs. sham + pharmacotherapy on populations including bipolar disorder

| Author, year<br>Study Design<br>Primary Endpoint(s)<br>Quality | Intervention and sample size<br>Study Details | Population characteristics | Change in depressive symptoms | Response<br>Remission |
|---|---|---|---|---|
| | | **rTMS + escitalopram vs. sham + escitalopram** | | |
| Bretlau et al., 2008[87]<br>3 weeks, mITT<br>Required failure in the current episode.<br>Fair | G1: rTMS + escitalopram (n = 25)<br>High frequency, 15 sessions over 3 weeks plus 20 mg escitalopram<br>G2: Sham + Escitalopram (n = 24)<br>20 mg escitalopram<br>**Treatment Strategy**<br>Combination | **Mean number of failed anti-depressant trials (current episode):**<br>ECT: 2.8 (0.9)<br>Pharm: 2.5 (0.9)<br>**Baseline Depression:**<br>HAM-$D_{17}$, mean* (SD)<br>G1: 25.3 (3.0)<br>G2: 24.7 (3.2)<br>*based on<br>G1: n =22<br>G2: n = 23 | **HAM-$D_{17}$**<br>Change, mean* (SD)<br>G1: -8.9<br>G2: -5.6<br>Effect size:<br>0.78 (0.18-1.39)* | **HAM-$D_{17}$**<br>Response, n (%)<br>NR<br>Remission, n (%)<br>NR |

ECT, electroconvulsive therapy; G, group; mg, milligram; n, number; HAM-$D_{17}$, 17-item Hamilton Depression Scale; Pharm, _; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation.

*MDD only***.** There were no eligible studies

*MDD/Bipolar.* One 3-week study compared the combination of high frequency rTMS plus escitalopram (n = 25 patients) with sham rTMS plus escitalopram (n = 24 patients) in patients who had discontinued their previous antidepressant pharmacotherapy (failed within the current episode).[87] Those participating were moderately-to-severely depressed (mean HAM-$D_{17}$ was 25.3 [SD 3.0] in rTMS group and 24.7 [SD 3.2] in the sham control). Authors conducted a modified ITT analysis. Mean depressive severity change was -8.9 in the rTMS escitalopram group and -5.6 in the sham alone group. This comparison favored rTMS plus pharmacotherapy over pharmacotherapy alone with the authors reporting an effect size of 0.78 (95% CI, 0.18 to 1.39).

**Tier 3: Patients with probable treatment resistant depression.** There were no eligible studies

## CBT vs. Pharmacotherapy

Two Tier 2 studies, both MDD only, compared psychotherapy vs. pharmacotherapy and are described in Table 34. Both studies required an antidepressant failure in the current episode and utilized mixed strategies with between group differences. One study compared augmenting to switching while the second study required patients randomized to psychotherapy discontinue medications and compared this group to those who continued their antidepressant medications. Studies were similar in duration thus no comparison by study duration was made.

**Tier 1: Patients with two or more failures.** There were no eligible studies

**Tier 2: Patients with one or more failures.** *MDD only.* One study utilized a randomization strategy that considered patient choice. Sixteen sessions of cognitive therapy were compared to medication treatment as either an augmentation strategy (each was added to citalopram treatment, respectively) or a switch strategy (changed to CT or a different medication treatment).[85] Patients entering all arms were of moderate severity (QIDS-SR mean 11 to 12). Using an ITT analysis, no differences in percentage change in depressive symptomatology were found when comparing CT to medication in either the augmentation switch (-29.5% vs. -28.3%, $P = 0.8302$) or switch (-15.6% vs. -17.2%, $P = 0.9040$) strategy comparisons. For patients who received augmentation to their citalopram, the response rate did not differ for those to whom CT was added (n = 65 patients) vs. those to whom medication was added (n = 117 patients) (35.4% vs. 28.2%, $P = 0.2493$). Similarly, the response rate did not differ between those who switched to CT (n = 36 patients) compared to those who switched to a different medication (n = 86 patients) (22.2% vs. 26.7%, $P = 0.8390$). As with change in severity and response, no differences between cognitive therapy and pharmacotherapy were found in remission between groups in the augmentation ($P = 0.7803$) or switch group comparisons ($P = 0.9032$).

One small study[86] randomized patients to either switch to 4 months of either CBT (n = 7) or a continue their current medication management (n = 6). Enrolled patients had moderate depressive severity (mean HAM-D score at baseline 18.6 for CBT [SD 3.3] and 18.3 [SD 3.9] for medication). A limited treatment completer's analysis of acute phase outcomes at 4 months suggested a greater decrease in severity for the CBT group (-7.6 points [n = 5 patients] vs. plus 1.5 points [n = 4 patients], statistical analysis not reported).

Table 34. Tier 2 studies comparing psychotherapy to pharmacotherapy in patients with major depressive disorder

| Author, year<br>Endpoint<br>Episode Failure<br>Quality | Intervention and Sample Size<br>Study Details | Population characteristics | Change in depressive symptoms | Response<br>Remission |
|---|---|---|---|---|
| Thase et al., 2007[85]<br>12-14 weeks<br>Required failure in the current episode<br>Fair | Augmentation - Cognitive Therapy (n = 65)<br>Continued citalopram and added CT (16 sessions in 12 weeks)<br>Augmentation - Medication (n = 117)<br>Citalopram plus buproprion SR or buspirone<br>Switch - Cognitive Therapy (n = 36)<br>Switch from citalopram to CT 16 sessions in 12 weeks<br>Switch - Medication (n = 86)<br>Switch from citalopram to sertraline, bupropion SR, or extended-release-XR<br>**Treatment strategy**<br>Mixed-between group differences<br>Definitions<br>Remission defined as QID-SR ≤ 5 | **Mean number of failed anti-depressant trials:**<br>Aug CT: NR<br>Aug Med: NR<br>Switch CT: NR<br>Switch Med: NR<br>**Baseline Depression**<br>QIDS-SR, mean (SD)<br>Aug CT: 11.9 (4.3)<br>Aug Med: 12.0 (4.6)<br>Switch CT: 11.2 (4.3)<br>Switch Med: 12.1 (4.6) | **QIDS-SR**<br>% Change, mean (SD)<br>Aug CT: -29.8 (40.5)<br>Aug Med: -28.3 (39.6) *P* = 0.8302<br>Switch CT: -15.6 (40.7)<br>Switch Med: -17.2 (46.2) *P* = 0.9040 | **QIDS-SR**<br>Response, n (%)<br>Aug CT: 23 (35.4)<br>Aug Med : 33 (28.2)<br>*P* = 0.2493<br>Switch CT: 8 (22.2)<br>Switch Med: 23 (26.7)<br>*P* = 0.8390<br>Remission, n (%)<br>Aug CT: 20 (30.8)<br>Aug Med: 39 (33.3)<br>*P* = 0.7803<br>Switch CT: 11 (30.6)<br>Switch Med: 23 (26.7)<br>*P* = 0.9032 |
| Moore et al. 1997[86]<br>4 months is closest to end of treatment, completers analysis<br>Required failure in the current episode<br>Fair | Cognitive Behavioral Therapy (n = 7)<br>minimum of 4 txts 1st month, 2 txts 2nd month and one per month following<br>Continued medication management (n = 6)<br>Continued medication dose within recognized therapeutic theshold<br>**Treatment Strategy**<br>Mixed- between group differences | **Mean number of failed anti-depressant trials:**<br>CBT: NR<br>Meds: NR<br>**Baseline Depression**<br>HAM-$D_{17}$*, mean (SD)<br>CBT: 18.6 (3.3)<br>Meds: 18.3 (3.9)<br>*Completers only (CBT n = 5, Meds n = 4) | **HAM-$D_{17}$**<br>Change*, mean (SD)<br>CBT: -7.6<br>Meds: +1.5<br>*Completers only<br>*P* = NR | **HAM-$D_{17}$**<br>Response:<br>NR at end of txt<br>Remission:<br>NR at end of txt |

CT, cognitive therapy; CBT, cognitive behavioral therapy; G, group; HAM-$D_{17}$, 17-item Hamilton Depression Scale; HAM-$D_{21}$, 21-item Hamilton Depression Scale; ITT, intent-to-treat; m-ITT, modified intent-to-treat; n, number; NR, not reported; *P*, p-value; QIDS-SR, Quick Inventory of Depressive Symptomatology-Self Report; SD, standard deviation; SR, sustained release; txt, treatment; XR, extended release.

*MDD/Bipolar.* There were no eligible studies.

**Tier 3: Patients with probable treatment resistant depression.** There were no eligible studies.

**Tiers 1-3: Combined results.** Only two studies were identified for this comparison.[85,86] While one study did not find difference between groups in treatment efficacy (i.e. change in

severity, response, and remission),[85] the second study showed a difference in change in depressive severity but did not report the results of a test of statistical significance.[86] Both studies were identified in Tier 2, required a failure in the current episode, included only patients with MDD, included samples with moderate depressive severity, and utilized similar treatment characteristics (i.e. both used cognitive behavioral therapy and were approximately four months in duration). The first study compared treatment arms that augmented with either psychotherapy or a new antidepressant medication and arms that switched to psychotherapy or a new antidepressant.[85] The second study compared switching to psychotherapy to continued medication management.[86]

## Key Question 1b: Pharmacologic Interventions for Acute Phase Treatment—Overview of Pharmacologic vs. Pharmacologic Treatments

All studies reviewed in this section are RCTs that involve Tier 1 TRD (≥ 2 failures of adequate antidepressant trials) and MDD-only patients. This synthesis allows an indirect comparison between what one might expect as a "next step" pharmacologic intervention relative to a next step nonpharmacologic intervention. Consequently, these studies may provide a reference for the degree of response (or remission) that one could expect from a next step pharmacologic treatment (relative to a next step nonpharmacologic treatment).

Some of these studies include a group that did not receive an active primary antidepressant treatment (e.g., olanzapine, which by itself is not used as an antidepressant); these arms will not be considered in the subsequent analyses. We focus instead on the same three outcomes addressed in previous sections—change in depressive severity, response rate, and remission rate. However, we will not formally assess strength of evidence as we did in the prior sections. Rather, we will present the available clinical response data that illustrate what is expected following an active antidepressant treatment. We will consider both responses seen after a change in pharmacologic treatment (either a switch or augmentation) and responses seen after maintenance on the same pharmacologic management without a change in treatment. Finally, also in contrast to our prior sections, we will not consider the role of MDD/Bipolar mix or Tier definition, as these variable are by definition fixed in this section, but we will attempt to consider the other key elements.

We identified nine Tier 1 MDD only studies involving moderately-to-severely depressed groups that compared pharmacologic treatment as a next treatment step (Table 35). We attempted to determine mean effect sizes, relative risks of response, and relative risks of remission for pharmacologic vs. control studies to allow a comparison with similar outcomes in the nonpharmacologic vs. control trials (KQ 1a, indirect). However, there were no comparable, common control groups not receiving a mood-related medication to allow such comparisons. Instead, we determined mean average outcomes for pharmacologic treatments. Although we were unable to statistically compare these outcomes, there was broad overlap in their decreases in depressive severity, relative risks of response, and relative risks of remission.

Table 35. Number of good and fair quality studies by comparison and definition of treatment resistance (tier)

| Comparison and Population Included | Definition of Treatment Resistance | | |
|---|---|---|---|
| | ≥ 2 failures | ≥ 1 failure | Probable |
| MDD only | | | |
| Pharmacotherapy vs. Pharmacotherapy | 9 | NA | NA |

MDD, major depressive disorder, NA, not applicable.

# Key Question 1b: Pharmacologic Interventions for Acute Phase Treatment—Key Points of Direct Comparisons

All studies included in the pharmacologic intervention vs. pharmacologic intervention were conducting in patients with TRD, MDD only who had two or more antidepressant failures. Nine studies were identified, six studies primarily tested switch strategies while three assessed augmentation. Five of the nine studies also included a maintenance arm, allowing further analysis of this strategy as well. To allow comparison to the nonpharmalogical interventions, weighted means were calculated for each strategy for the three outcomes of interest.

Regarding changes in depressive severity, mean changes in MADRS scores were similar across the three strategies (switch -11.2 [95% CI, -7.8 to -14.7], augmentation -12.2 [95% CI, -8.5 to -15.9], and maintenance -9.4 [95% CI, -4.9—13.9]). Consistent results were seen for response and remission rates (switch 35.3 percent [95% CI, 28.4 to 42.3] and 19.8 percent [95% CI, 14.3 to 25.5], augmentation 37.5 percent [95% CI, 27.1 to 47.4] and 25.3 percent [95% CI, 15.6 to 34.9], maintenance 29.9 percent (16.4 to 43.4] and 16.8 percent [95% CI, 7.8 to 25.8], respectively).

Only one study did not require a failure in the current episode[88] limiting further analysis by this variable. Though some variability in the depressive severity of populations was present, differences by severity were not apparent.

# Key Question 1b: Pharmacologic Interventions for Acute Phase Treatment—Detailed Analysis of Direct Comparisons

Nine studies providing relevant outcome measures for the effect of next step pharmacologic treatments in MDD patients with Tier 1 TRD were identified (Table 36).

Table 36. Tier 1 studies testing switch strategies in patients with TRD

| Author, year<br>Endpoint<br>Episode Failure<br>Quality | Intervention and Sample Size<br>Study Details | Population Characteristics | Change in Depressive Symptoms | Response<br>Remission |
|---|---|---|---|---|
| Switching strategies | | | | |
| Fava et al., 2006[89]<br>14 weeks<br>Required failure in the current episode<br>Good | **MIR (n = 114)**<br>Up to 60mg/d<br>**NOR (n = 121)**<br>Up to 200mg/day<br>**Treatment strategy**<br>G1: Switch<br>G2: Switch<br>**Definitions**<br>Remission defined as HAM-D-17 ≤ 7 | **Baseline Depression**<br>HAM-$D_{17}$, mean (SD)<br>MIR: 19.8 (7.0)<br>NOR: 18.6 (5.9) | **HAM-D-17**<br>Change: NR | **HAM-D-17**<br>Remission, n<br>MIR :14 (12.3)<br>NOR: 24 (19.8)<br>*P* = 0.27 |
| Mazeh et al., 2007[90]<br>6 weeks* only in the elderly<br>Required failure in the current episode<br>Fair | **PAR (n = 15)**<br>10-60mg/d, mean = 26mg/d<br>**VEN (n = 15)**<br>75-300mg/d, mean = 165mg/d<br>**Treatment strategy**<br>PAR: Switch<br>VEN: Switch<br>**Definitions**<br>Remission defined as HAM-$D_{21}$ ≤ 7 | **Baseline Depression**<br>HAM-$D_{21}$, mean (SD)<br>PAR: 30.1 (7.9)<br>VEN: 26.3 (5.9) | **HAM-D-21**<br>Change, mean (SD)<br>PAR: -12.5<br>VEN: -19.1<br>*P* < 0.0003 | **HAM-D-21**<br>Response, n (%)<br>PAR: 8 (53)<br>VEN: 12 (80)<br>*P* = NR<br>Remission, n (%)<br>G1: 5 (33)<br>G2: 9 (60)<br>*P* = NR |
| Poirier and Boyer, 1999[91]<br>4 weeks<br>Required failure in the current episode<br>Fair | **VEN (n = 61)**<br>37.5mg/twice day, increased to 200 - 300 mg/day<br>**PAR (n = 62)**<br>initiated at 20mg/day and increased to 30—40 mg/day<br>**Treatment strategy**<br>Switch<br>**Definitions**<br>Remission defined as HAM-D-17 < 10 | **Baseline Depression**<br>HAM-$D_{17}$, mean (SD)<br>VEN: 24.6 (3.9)<br>PAR : 24.5 (4.1) | **HAM-D-17**<br>Change*, mean (SD)<br>VEN: -11.1 (8.5)<br>PAR: -10.2 (6.8)<br>*P* = 0.55<br>ITT, *P* = 0.70<br>*N observed (G1: 52, G2: 55) | **HAM-D-17**<br>Response, n<br>VEN: 27 (44.3)<br>PAR: 18 (29.0)<br>ITT, *P* = 0.07<br>Remission, n<br>VEN: 22 (36.1)<br>PAR: 11 (17.7)<br>ITT, *P* = 0.02 |

HAM-$D_{17}$, 17 item Hamilton Depression Scale; HAM-$D_{21}$, 21-item Hamilton Depression Scale; ER, extended release; ITT, intent-to-treat; LD, low-dose; MADRS, Montgomery-Asberg Depression Rating Scale; mg/d, milligrams per day; MIR, mirzapine; n, number; NOR, nortiptyline; NR, not reported; NS, not sufficient; OLA-FLU, olanzapine/fluoxetine; SD, standard deviation; TRAN, tranylcypromine; VEN, venlafaxine; wk, week.

Table 36. Tier 1 studies testing switch strategies in patients with TRD (continued)

| Author, year<br>Endpoint<br>Episode Failure<br>Quality | Intervention and Sample Size<br>Study Details | Population Characteristics | Change in Depressive Symptoms | Response<br>Remission |
|---|---|---|---|---|
| McGrath et al., 2006[92]<br><br>12 weeks<br><br>Required failure in the current episode<br><br>Good | **TRAN (n = 58)**<br>10mg/d for 2wk, weekly increases of 10 mg/d until intolerance or 60 mg/d maximum<br><br>**VEN ER plus MIR (n = 51)**<br>Venlafaxine - 37.5mg/d week 1, 75mg/d week 2, 150mg/d weeks 3-5, 225 mg/d weeks 6-8, 300mg/d thereafter<br>Mirtazapine—15mg/d weeks 1-2, 30mg/d next 8 weeks, 45mg/d thereafter<br><br>**Treatment strategy**<br>Switch<br><br>**Definitions**<br>Remission defined as HAM-D-21 ≤ 7 | **Baseline Depression**<br>HAM-$D_{17}$, mean (SD)<br>TRAN: 19.6 (7.6)<br><br>VEN-MIR: 19.7 (5.5) | **HAM-D-17**<br>Change: NR | **HAM-D-17**<br>Remission, n (%)<br>TRAN: 4 (6.9)<br>VEN-MIR: 7 (13.7)<br>*P* = NS |
| Corya et al., 2006[93]<br><br>12 weeks<br><br>Required failure in the current episode<br><br>Fair | **OLA-FLU (n = 243)**<br>Combined 4 groups<br><br>**OLA (n = 62)**<br>6 or 12mg/d<br><br>**FLU (n = 60)**<br>25 or 50mg/d<br><br>**VEN (n = 59)**<br>75-375mg/d<br><br>**LD OLA plus FLU (n = 59)**<br>1mg/d OLA, 5mg FLU<br><br>Treatment strategy<br>OLA-FLU: Switch/Combination<br>OLA: Not of interest<br>FLU: Switch<br>VEN: Maintenance<br>LD OLA-FLU: Switch<br><br>**Definitions**<br>Remission defined as MADRS ≤ 8 at two consecutive visits | **Baseline Depression**<br>MADRS, mean (SD)<br><br>Overall: 30.0 (6.8) | **MADRS**<br>Change, mean (SD)<br>OLA-FLU: -14.06 (0.59)<br>OLA: -7.71 (1.17)<br>FLU: -11.70 (1.14)<br>VEN: -13.73 (1.16)<br>LD OLA-FLU: -11.97 (1.13)<br><br>OLA-FLU vs.OLA *P* < 0.001<br>all others NS | **MADRS**<br>Response, n (%)<br>OLA-FLU: 100 (43.3)<br>OLA: 15 (25.4)<br>FLU: 19 (33.9)<br>VEN: 29 (50.0)<br>LD OLA-FLU: 20 (36.4)<br>OLA-FLU vs. OLA, *P* = 0.017<br>All others NS<br><br>Remission, n (%)<br>OLA-FLU: 69 (29.9)<br>OLA: 8 (13.8)<br>FLU: 10 (17.9)<br>VEN: 13 (22.4)<br>LD OLA-FLU: 11 (20.0)<br>OLA-FLU vs. OLA, *P* = 0.013.<br><br>All others NS |

Table 36. Tier 1 studies testing switch strategies in patients with TRD (continued)

| Author, year<br>Endpoint<br>Episode Failure<br>Quality | Intervention and Sample Size<br>Study Details | Population Characteristics | Change in Depressive Symptoms | Response<br>Remission |
|---|---|---|---|---|
| Shelton et al., 2005[94]<br><br>8 weeks<br><br>Did not require failure in the current episode<br><br>Good | **OLA-FLU combination (n = 146)**<br>6 mg/d OLA plus 25mg/d FLU<br>OR<br>12mg/d OLA plus 50mg/d FLU<br><br>**OLA (n = 144)**<br>6-12mg/d<br><br>**FLU (n = 142)**<br>25 to 50 mg/d<br><br>**NOR (n = 68)**<br>Max dose 175mg/d<br><br>**Treatment strategy**<br>G1: Switch/Combination<br>G2: Not of interest<br>G3: Switch<br><br>NOR: Maintenance | **Baseline Depression**<br>MADRS, mean (SD)<br>OLA-FLU: 28.5 (7.5)<br>OLA: 28.4 (7.3)<br>FLU: 28.4 (7.3)<br>NOR: 28.8 (6.5) | **MADRS**<br>Change, mean (SE)<br>OLA-FLU: -8.71 (0.70)<br>OLA: -6.95 (0.71)<br>FLU: -8.51 (0.70)<br>NOR: -7.46 (0.98<br><br>FLU vs.OLA-FLU, *P* = 0.841<br><br>G2 vs. G1, *P* = 0.77 | **MADRS**<br>Response, n (%)<br>OLA-FLU: 40 (27.5)<br>OLA: 27 (19.3)<br>FLU: 41 (28.9)<br>NOR: 20 (30.3)<br>*P* = 0.18<br><br>Remission, n (%)<br>OLA-FLU: 24 (16.9)<br>OLA: 18 (12.9)<br>FLU: 18 (13.3)<br>NOR: 12 (18.2)<br>*P* = 0.62 |

**Tier 1: Patients with two or more failures.** Nine studies were identified for this population. Six of the studies used switch strategies and three tested an augmentation strategy.

*Switching strategies.* Five studies testing a switch strategy were identified and are described in Table 36. Two studies compared strategies of augmenting Fluoxetine with Olanzapine, continuing fluoxetine alone, or switching to Olanzapine alone. One study compared the twelve week outcomes for patients who failed venlafaxine treatment and were randomized to one of five groups: a combination of Olanzapine either 6 or 12 mg/day)/Fluoxetine (either 25 or 50 mg/day) (n = 243 patients, pooled from 4 groups), Olanzapine alone (either 6 or 12 mg/day) (n = 62 patients), Fluoxetine alone (either 25 or 50 mg/day) (n = 60 patients), a "pseudo placebo" low-dose combination of Olanzapine (1mg/day) and Fluoxetine (5 mg/day)(n = 59 patients), or continuing with Venlafaxine alone 75-375 mg/day) (n = 59 patients).[93] Only one treatment failure was required in the current episode (failure to respond to Venlafaxine). Baseline depressive-severity for the overall sample was in the moderate-to-severe range (MADRS 30.0). An ITT analysis favored the Olanzapine/ Fluoxetine combination vs. Fluoxetine alone in all depression outcome comparisons, but showed no difference between any of the other groups. The combination was better than Fluoxetine alone for greater change in depressive severity (-14.06 vs. -7.71, $P < 0.001$; other severity changes ranged from -11.7 to -13.73), greater response rate (43.3% vs. 25.4%, $P = 0.017$; other response rates ranged from 33.9 percent to 50.0 percent) and greater remission rate (29.9% vs. 13.8%, $P = 0.013$; other remission rates ranged from 17.9 percent to 22.4 percent).

A second Olanzapine/Fluoxetine switch study compared the eight-week outcomes for four groups following nortriptyline treatment failure: a combination of Olanzapine (6 mg/day or 12 mg/day)/Fluoxetine (25 mg/day or 50 mg/day) (n = 146 patients), Olanzapine alone (6-12 mg/day)(n = 144 patients), Fluoxetine alone (25-50 mg/day) (n = 142 patients), and continuing on Nortriptyline alone (50-175 mg/day) (n = 68 patients).[94] Only one treatment failure was

required to be in the current episode (failure to respond to Nortriptyline). Baseline depressive severity for each group averaged between 28 and 29 on the MADRS, consistent with moderate-to-severe depressive severity. A mixed-effects model repeated-measures regression showed no difference between the four groups in decrease in depressive severity (-8.71, -6.95, -8.51, and -7.46, respectively, *P* = NS), response rates (27.5%, 19.3%, 28.9%, and 30.3%, respectively, *P* = 0.18), or remission rates (16.9%, 12.9%, 13.3%, 18.2%, respectively, *P* = 0.62).

Two studies compared switching to Venlafaxine with switching to Paroxetine. One four-week study compared Venlafaxine, 200-300 mg/day (n = 61 patients) to Paroxetine, 30-40 mg/day (n = 62 patients).[95] Patients had failed two adequate treatments other than Venlafaxine or Paroxetine in the current episode. Enrolled patients were severely depressed at study entry (mean HAM-D 24-25). The authors conducted an ITT analysis. The change in depressive severity did not differ between the two groups. However, the response rate tended to favor Venlafaxine (44.3% vs. 29.0%, *P* = 0.07), and the remission rate supported Venlafaxine over Paroxetine (36.1% vs. 17.7%, *P* = 0.02).

A separate six-week study involving the same comparison assessed 6 week outcomes for patients 65 years and older who were randomized to receive Venlafaxine (75 mg to 300 mg/day, mean daily dose 165 mg/day; n = 15 patients) or Paroxetine (10-60 mg/day, mean 26 mg/day; n = 15 patients).[90] Patients had failed two adequate trials during the current episode and were severely depressed at study entry (mean HAM-D 26-30). In an ITT analysis, the decrease in depressive severity after 6 weeks was greater for Venlafaxine than Paroxetine (-19.1 vs. -12.5, *P* < 0.0003). Differences between response rates (80% vs. 53%, *P* = NR) and remission rates (60% vs. 33%) in this small sample was less clear.

One study compared 12 week outcomes for patients who had failed three adequate antidepressant treatments in the current episode for patients randomized to Tranylcypromine (10 mg to 60 mg/day) (n = 58 patients) or a combination of Venlafaxine ER (37.5 mg to 300 mg/day) plus Mirtazapine (15—45 mg/day) (n = 51 patients).[96] Patients were severely depressed at study entry (mean HAM-D 19-20). Outcomes tended to favor the Venlafaxine/Mirtazapine combination, but not to a statistically significant degree. In an ITT analysis, response rates (as measured by the QIDS-SR) did not significantly differ (12.1 percent with Tranylcypromine vs. 23.5 percent with Venlafaxine plus Mirtazapine), nor did the remission rates measured by HAM-D (6.9% vs. 13.7%).

A good 12-14 week quality study compared switching to Mirtazapine (up to 60 mg/day; n = 114 patients) or Nortriptyline (up to 200mg/day; n = 121 patients) in a group of patients who had failed two adequate antidepressant treatments in the current episode.[89] Enrolled patients were severely depressed at baseline (mean HAM-D 18-20). Response rates as measured by the QIDS-SR did not different significantly (13.4 percent for Mirtazapine vs. 16.5 percent for Nortriptyline). Similarly, remission rates did not differ significantly between the Mirtazapine and Nortriptyline groups (12.3% vs. 19.8%, *P* = 0.27).

*Augmenting strategies.* Three studies tested augmenting strategies and are described in Table 37. One study compared outcomes at six weeks for patients who were assigned to patients who had not responded to a seven-week Nortriptyline trial and were assigned to augment Nortriptyline with either Lithium (dose not clarified; n = 18 patients) or placebo (n = 17 patients).[97] Prior to their Nortriptyline trial, they had failed at least one but no more than 5 medication trials during the current episode. Patients were moderately depressed at study entry (mean HAM-D 17-18). In an ITT analysis, change in depressive severity did not differ between groups (-2.9 for Lithium augmentation vs. -3.6 for placebo, *P* = 0.72). Similarly, response rates

Table 37. Tier 1 studies testing augmentation strategies in patients with TRD

| Author, year<br>Endpoint<br>Episode Failure<br>Quality | Intervention and sample size<br>Study Details | Population characteristics | Change in depressive symptoms | Response<br>Remission |
|---|---|---|---|---|
| **Augmentation strategies** | | | | |
| Nierenberg et al., 2003[97]<br>6 weeks<br>Required failure in the current episode<br>Fair | **LITH Augmentation (n = 18)**<br>Dosing strategy NR<br>Placebo (n = 17)<br>All patients continued nortriptyline<br>**Treatment strategy**<br>LITH: Augmentation<br>PLACEBO: Maintain | **Baseline Depression**<br>HAM-$D_{21}$, mean (SD)<br>LITH: 18.8<br>PLACEBO: 19.8 | **HAM-D-21**<br>Change, mean (SD)<br>LITH: -2.9<br>PLACEBO: -3.6<br>*P* = NR | **HAM-D-21**<br>Response, n (%)<br>LITH: 2 (11.1)<br>PLACEBO: 3 (17.6)<br>*P* = NS |
| Shelton et al., 2001[88]<br>8 weeks<br>Required failure in the current episode<br>Fair | **OLA+PLA (n = 8)**<br>5-20mg/d<br>**FLU+PLA (n = 10)**<br>20-60mg/d<br>**OLA+FLU (n = 10)** same dose as above<br>**Treatment strategy**<br>OLA+PLA: Not of interest<br>FLU+PLA: Maintain<br>OLA: Augmentation | **Baseline Depression**<br>MADRS: NR | **MADRS**<br>Change, mean (SD)<br>OLA+PLA: -2.8<br>FLU+PLA: -1.2<br>OLA+FLU:-13.6 | **MADRS**<br>Response, n (%)<br>OLA+PLA: 0 (0)<br>FLU+PLA: 1 (10)<br>OLA+FLU G3: 6 (60)<br>OLA-FLU vs. OLA+PLA, *P* = 0.03<br>OLA+FLU vs. FLU+PLA, *P* = 0.11 |
| Thase et al., 2007[98]<br>8 weeks<br>Required failure in the current episode<br>Fair | **OLA+FLU (n = 200)**<br>OLA 6, 12, or 18 mg/day plus 50 mg/day FLU<br>OLA (n = 206)<br>6, 12, or 18 mg/day<br>FLU (n = 200)<br>50 mg/day<br>**Treatment strategy**<br>OLA-FLU: Augmentation<br>OLA: Not of interest (Switch)<br>FLU: Maintain | Baseline Depression<br>MADRS, mean (SD)<br>OLA+FLU: 30.0 (6.7)<br>OLA: 29.9 (6.4)<br>FLU: 29.9 (6.7) | **MADRS**<br>Change, mean (SD)<br>OLA+FLU: -12.6 (10.3)<br>OLA: -9.2 (9.7)<br>FLU: -8.9 (9.0)<br>OLA+FLU vs. OLA, *P* < 0.001<br>OLA+FLU vs. FLU, *P* < 0.001 | **MADRS**<br>Response, n (%)<br>OLA+FLU G1: 80 (40.4)<br>OLA G2: 60 (29.6)<br>FLU G3: 51 (25.9)<br>OLA+FLU vs. FLU, *P* = 0.028<br>OLA+FLU vs. FLU, *P* = 0.003<br>Remission, n (%)<br>OLA+FLU: 54 (27.3)<br>OLA: 34 (16.7)<br>FLU : 29 (14.7)<br>OLA+FLU vs.FLU, *P* = 0.012<br>OLA-FLU vs. FLU, *P* = 0.003 |

FLU+PLA, fluoxetine plus placebo; HAM-$D_{21}$, 21-item Hamilton Depression Scale; LITH, lithium, n, number; MADRS, Montgomery-Asberg Depression Rating Scale; mg/d, milligrams per day; NR, not reported; OLA, olanzapine; OLA-FLU, olanzapine+fluoxetine; OLA+PLA, olanzapine plus placebo; SD, standard deviation;

did not differ significantly for Lithium augmentation vs. placebo augmentation (11.1% vs. 17.6%, *P* = NS).

A second study, which consisted of two parallel, concurrent trials, compared the 8 week outcomes of an Olanzapine/ Fluoxetine combination (6, 12, or 18 mg Olanzapine plus 50 mg/day of Fluoxetine; n = 200 patients) Olanzapine (6, 12, or 18 mg/day; n = 199 patients) or Fluoxetine (50 mg/day; n = 206 patients).[98]

The pooled analyses are reported here. Treatment failure was in the current episode. Patients entering the study were moderately-to-severely depressed (MADRS score of approximately 30). ITT analyses at study end showed favored the combination treatment relative to the other two groups in each instance. The combination produced greater difference between groups in the decrease in depressive severity (-10.8 vs. -10.1 in Olanzapine only, and vs. -9.4 in Fluoxetine only, $P < 0.001$ in each instance); a greater response rate (40.4% vs. 25.9%, [$P = 0.003$] and vs. 29.6 percent [$P = 0.028$], respectively); and a greater remission rate (27.3% vs. 14.7% [$P = 0.003$] and vs. 16.7% [$P = 0.012$], respectively).

A third study compared outcomes at 8 weeks for patients who had failed to respond to two different classes of antidepressants and had additionally failed a trial of Fluoxetine in the current episode. These patients were assigned to either switch to Olanzapine (5 to 20 mg/day; n = 8 patients), add Olanzapine (5 to 20 mg/day) to Fluoxetine (20 to 60 mg/day) (n = 10 patients) or continue with Fluoxetine (50 mg/day) with placebo added (n = 10 patients).[99] Baseline mean depressive severity was not reported. The Olanzapine/Fluoxetine augmentation group had a greater decrease in HAM-$D_{21}$ items severity than either the Olanzapine switch group (-11.7 vs. -5.9, $P = 0.03$) or the Fluoxetine continuation group (-11.7 vs. -3.8, $P = 0.07$). The Olanzapine/Fluoxetine augmentation group also had a greater response rate than the Olanzapine switch group (60% vs. 0%, $P = 0.03$) and a trend towards greater response than the Fluoxetine continuation group (60% vs. 10%, $P = 0.11$).

**Synthesis of MDD outcomes (Tier 1 only).** To provide information reporting average outcomes in pharmacologic trials of TRD, we calculated weighted means for the change in depressive severity, response rate, and remission rate (Table 38).

Table 38. Mean clinical outcomes for TRD (Tier 1) patients in pharmacologic studies

| | Switching | Augmentation | Maintenance |
|---|---|---|---|
| Mean change HAM-D | -10.6 (-4.9 to -16.4) | No data | No data |
| Mean change MADRS | -11.2 (-7.8 to -14.7) | -12.2 (-8.5 to -15.9) | -9.4 (-4.9 to -13.9) |
| Mean response rates (HAM-D and MADRS) | 35.3% (28.4 to 42.3) | 37.5% (27.1 to 47.4) | 29.9% (16.4 to 43.4) |
| Mean remission rates (HAM-D and MADRS) | 19.8% (14.3 to 25.2) | 25.3% (15.6 to 34.9) | 16.8% (7.8 to 25.8) |

HAM-D, Hamilton Rating Scale for Depression; MADRS, Montgomery-Asberg Depression Rating Scale.

We quantitatively synthesized weighted means of the changes in depressive severity for studies involving the two interviewer-administered instruments, the HAM-D and MADRS. For patients switched to a new medication, the mean average change in HAM-D was -10.6 points, and the mean average change in studies using the MADRS was—11.2. For patients receiving medication augmentation, the mean average change in depressive severity was–12.2 on the MADRS. We also identified five measures of depressive severity change in patients who continued on their same medication without a change in treatment. The four measured by MADRS showed a decrease ranging from 7.46 to 13.73 points, with a mean change MADRS was—9.4, slightly less than switching or maintaining. The one measured by HAM-D showed a decrease of 3.6 points.

For changes in response rates, we identified 13 measures involving switch strategies. These response rates varied greatly, with response rates ranging from 12.1 to 80 percent. The two highest response rates were from a study restricted to an elderly population,[90] a sample distinct from the others. Four of these rates were assessed using the QIDS-SR, a self-report measure different from the HAM-D and MADRS which could add in heterogeneity. Excluding these six

rates left 7 measures and a maximal response rate of 44.3 percent. A weighted mean response rate for these 7 measures was 35.3 percent. Considering augmentation strategies provided 5 more measures, ranging from 11.1 percent to 43.3 percent. A quantitative synthesis of these rates suggests an average response rate of 37.5 percent for TRD patients following an active next-step pharmacologic treatment. For those who maintained on their pharmacologic treatment, we identified three measures of response rates, which ranged from 10 to 50 percent. The weighted mean average response rate for maintenance treatment was 29.9 percent.

Finally, for changes in remission rates, we identified 13 measures involving switch strategies, 11 of which were not restricted to the elderly population. These 11 remission rates ranged from 6.9 percent to 36.1 percent, with a weighted mean average remission rate of 19.8 percent for TRD patients following an active next-step pharmacologic treatment. Three studies with augmentation arms provided four augmentation measures of remission rates, ranging from 16.9 percent to 29.9 percent, with a weighted mean average remission rate of 25.3 percent. For those who maintained on their pharmacologic treatment, three measures of remission showed rates varying from 14.7 percent to 22.4 percent, with a weighted mean average remission rate of 16.8 percent.

# Key Question 2: Efficacy or Effectiveness for Maintaining Remission or Treating Patients With Unresponsive or Recurrent Disease: Overview

As with KQ 1, KQ 2 addressed direct or indirect comparisons of the four nonpharmacological interventions (ECT, rTMS, VNS, and either CBT or IPT). Unlike KQ 1, however, we did not include studies that compared pharmacologic interventions. In the detailed analysis section below, first we present the studies by comparison, then by tier, and then by whether the population involves MDD only patients or an MDD/Bipolar mix. Information is presented for the three tiers used in KQ 1 (Tier 1, two or more failures; Tier 2, one or more failures, but not including the studies in Tier 1; and Tier 3, "probable" treatment resistance). Again, only studies with quality ratings of good or fair are featured.

We identified a total of nine studies addressing maintenance of remission using nonpharmacologic interventions (Table 39). Two Tier 1 studies, reported in three articles, compared rTMS vs. sham in an MDD only population, with both indicating rTMS was superior to sham in preventing relapse.[67,68,83] However, these trials included very few patients in the relapse prevention phase.

Tier 2 evidence added three trials

Table 39. Number of studies by comparison and definition of treatment resistance (tier)

| Comparison and Population Included | Studies Included | |
|---|---|---|
| | MDD Only | MDD and Bipolar Disorder |
| Tier 1: ≥ 2 failures | | |
| rTMS vs. sham | 2 | 0 |
| Tier2: ≥ 1 failure | | |
| rTMS vs. sham | 2 additional | 1 additional |
| **Tier 3: Probable** | | |
| rTMS vs. sham | 0 | 1 additional |
| ECT vs. rTMS | 1 additional | 1 additional |
| CBT vs. usual | 1 additional | 0 |

CBT, cognitive behavioral therapy; control, usual care control group; ECT, electroconvulsive therapy; MDD, major depressive disorder; rTMS, repetitive transcranial magnetic stimulation; vs., versus.

comparing rTMS vs. sham. Two of these trials involved MDD only patients (four articles).[71,72,100,101] One study involved an MDD/Bipolar mix population.[76] All three trials supported benefit of rTMS over sham in maintaining remission.

Tier 3 evidence added four studies. One study compared rTMS vs. sham in an MDD/Bipolar mix population, finding benefit again for rTMS over sham.[77] Two studies provided the only head-to-head comparison available, comparing ECT vs. rTMS, one in an MDD only population that was reported in two articles[102,103] and one in an MDD/Bipolar mix population that was reported in three articles.[54-56] Both indicated no difference in maintaining remission at 6 month follow-up.

Most studies either allowed patients to continue antidepressants throughout the trial or required that they be given an antidepressant following the active nonpharmacological treatment. The duration of follow-up for assessing maintenance of remission ranged from 2 weeks to nearly 1 year. The method for assessing maintenance of remission varied among trials. Some trials followed (or randomized) only patients who had achieved a response or remission during active treatment and then measured relapse during a post-treatment period. Other trials followed all randomized participants during a post-treatment period regardless of response or remission with initial treatment. These trials generally reported the number of patients in remission at the end of treatment and at the end of follow-up, which provides an indirect measure of maintenance of remission.

## Strength of Evidence Tier One Only

There were no Tier 1 direct (head-to-head) comparisons available. The single comparison involving a Tier 1 TRD population was rTMS vs. sham; two studies provided insufficient evidence (Table 40). While these studies found that relapse rates do not differ significantly between rTMS and sham, too few patients were followed during the continuation phases of these studies to allow for meaningful conclusions.

Table 40. Maintenance of remission in TRD (those with two or more failures) for rTMS vs. sham; number of studies and subjects, and strength of the evidence domains

| | Number of studies; subjects* | Risk of bias Design/ Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| rTMS vs. sham | 2; 20 | High<br>2 RCTs<br>Fair | Inconsistent | Direct | Imprecise | 2 current trials found significant differences in maintenance of remission<br>Insufficient |

* Number of subjects reflects only those followed past acute treatment

RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation.

## Strength of Evidence: Considering all Tiers

Considering evidence from all tiers added seven studies (Table 41). Four studies were added to the rTMS vs. sham comparison. The remaining three studies added two new comparisons, with two studies comparing ECT vs. rTMS (the only direct nonpharmacologic evidence available for this key question), and one study comparing CBT vs. usual care.

Table 41. Maintenance of remission in TRD-related studies (all tiers) by intervention type; number of studies and subjects, and strength of the evidence domains

| | Number of studies; subjects* | Risk of bias Design Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| ECT vs. rTMS | 2; 57 | Medium<br>2 RCTs<br>2 fair | Consistent | Direct | Imprecise | 2 fair trials found no significant difference in ECT vs. rTMS in maintenance of remission<br>Low |
| rTMS vs. sham | 6; 132 | Medium<br>6 RCTs<br>6 fair | Inconsistent | Direct | Imprecise | rTMS may maintain remission better than sham (depending on length of followup); significance NR for 4 trials)<br>Low |
| CBT vs. usual care | 1; 158 | Medium<br>1 RCT<br>1 fair | Unknown | Direct | Unknown | CBT had lower risk of relapse than usual care at all followup points but significance of results varied by how relapse was defined and was not reported for all follow-up measures.<br>Low |

* Number of subjects reflects only those followed past acute treatment

CBT, cognitive behavioral therapy; control, usual care control group; ECT, electroconvulsive therapy; NA, not applicable; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; vs., versus.

Two Tier 3 studies directly compared ECT vs. rTMS, providing low strength of evidence that relapse rates are similar between ECT and rTMS. One study, reported in two articles, was in an MDD only population,[102,103] while the other, also reported in three articles, was in an MDD/Bipolar mix population.[54-56]

The four studies added to the rTMS vs. sham comparison changed the strength of evidence from insufficient to a low strength of evidence favoring rTMS in maintaining remission.[71,72,76,77,100,101] Data from the six total studies are inconsistent and involve a small number of participants followed after the active treatment phase.

Consideration of all tiers added a second new comparison, with a single Tier 3 MDD only study comparing CBT vs. usual care. This one fair randomized controlled trial provides low strength evidence for CBT compared with usual care. This relatively large study found CBT to maintain remission longer than usual care, although the statistical significance of the findings varied as a function of patient symptomatology. Additional studies could influence this conclusion and no direct comparisons have been made.

# Key Question 2: Efficacy or Effectiveness for Maintaining Remission or Treating Patients with Unresponsive or Recurrent Disease: Key Points

Only limited evidence addressed maintenance of remission among MDD patients. These included the following interventions: ECT (two studies), rTMS (eight studies, including ECT in two head-to-head trials), and CBT (1 study). No studies assessing maintenance of remission directly compared ECT, rTMS, VNS, and CBT in patients in a TRD (Tier 1) population. No evidence was identified for VNS. The only evidence for TRD (Tier 1) compared rTMS vs. sham.

## Electroconvulsive Therapy vs. Repetitive Transcranial Magnetic Stimulation

TRD (Tier 1) data was available for this comparison, but two trials provided direct Tier 3 evidence. One trial in an MDD only population, reported in two articles, found no statistically significant differences in relapse rates at 3 and 6 months after treatment ended.[102,103] A second trial in an MDD/Bipolar mix population, reported in three articles, provided similar results indicating no statistically significant differences in relapse rates between ECT and rTMS.[54-56]

## Repetitive Transcranial Magnetic Stimulation vs. Sham

Two Tier 1 studies found no statistically significant differences in relapse rates between rTMS and sham at 20 weeks[68] and 6 months.[67,83]

Three Tier 2 studies provided data supporting the benefit for rTMS vs. sham in maintaining remission. One MDD-only study found greater improvement in symptoms for rTMS patients than for the control patients at 2 weeks post-treatment.[71] Only the high-frequency rTMS delivered to the left dorsolateral prefrontal cortex and the low-frequency rTMS delivered to the right left dorsolateral prefrontal cortex were more effective than the sham intervention. A second study, also in an MDD-only population, found a trend towards lower relapse rates for rTMS compared with sham, but statistically significant differences were not reported.[72,100,101] One study involving an MDD/Bipolar mix population found that the one patient who responded after rTMS maintained response at 2 month follow-up.[76]

One Tier 3 study, involving an MDD/Bipolar mix population, showed benefit for rTMS vs. sham for 3 weeks after treatment ended, but the benefit had disappeared at three month follow-up.[77]

## Cognitive Behavioral Therapy vs. Usual Care

No TRD (Tier 1) evidence was available for this comparison. One relatively large study (150 patients) reported in three articles involved a Tier 3, MDD only population, supported the benefit of CBT vs. usual care in maintaining remission.[59,60,104] The study compared 20 weeks of CBT with usual care (clinical management involving psychiatrist visits and antidepressant medications) and measured remission rates over a total of 68 weeks. Patients treated with CBT had a lower risk of relapse than sham-treated patients (hazard ratio 0.54; 95% CI, 0.32-0.93; $P$ = 0.02).

# Key Question 2: Efficacy or Effectiveness for Maintaining Remission or Treating Patients With Unresponsive or Recurrent Disease: Detailed Analysis

## Electroconvulsive Therapy vs. Repetitive Transcranial Magnetic Stimulation

**Tier 1: Patients with two or more failures.** *MDD only.* No trial addressed maintenance of remission with ECT vs. rTMS therapy in an MDD Only population.

*MDD/bipolar.* No trial addressed maintenance of remission with ECT vs. rTMS therapy in an MDD-Bipolar mix population.

**Tier 2: Patients with one or more failures.** *MDD only.* No trial addressed maintenance of remission with ECT vs. rTMS therapy in an MDD Only population.

*MDD/bipolar.* No trial addressed maintenance of remission with ECT vs. rTMS therapy in an MDD-Bipolar mix population.

**Tier 3: Patients with probable treatment resistance.** *MDD only.* In the RCT of ECT versus rTMS,[102,103] 43 participants entered treatment, but only 41 continued in the 6-month follow-up to assess relapse rates (Table 42). In 20 participants, ECT was delivered according to a protocol with intensity 2.5 times the threshold energy and charge titrated up every second or third treatment to maintain a seizure length of 25 seconds or longer. Twenty-one participants received 20 sessions of high frequency at 90 percent motor threshold and 1,200 pulses per second. Prior to beginning treatment, the mean HAM-$D_{17}$ scores (standard deviation) for patients were 28.4 (9.3) in the ECT group and 25.8 (6.1) in the rTMS group. At the beginning of follow-up (i.e., end of treatment), mean HAM-$D_{17}$ scores were 7.9 (4.5) in the ECT group and 7.8 (3.7) in the rTMS group. These scores remained relatively stable at 3 and 6 months after treatment ended. At 3 months, 2 of 20 (10%) ECT-treated participants and 1 of 21 (5%) rTMS-treated participants relapsed. At 6 months, the figures were 4 of 20 (20%) and 4 of 21 (19%), respectively. Relapse rates were not statistically significantly different between these groups.

Table 42. Maintenance of remission in patients with major depressive disorder and probable treatment resistance by intervention; ECT vs. rTMS

| Author, year<br>Design<br>Quality | Intervention, Sample Size, and Study Details | Maintenance of Remission |
|---|---|---|
| Dannon et al., 2002;[102] extension of Grunhaus et al., 2000[103]<br>RCT<br>Fair | **ECT plus antidepressant post ECT (n = 20)**<br>35% bilateral, means sessions = 10.25 (3.1)<br>rTMS plus antidepressant post rTMS (n = 21)<br>High frequency, 20 sessions<br>Definitions<br>Response: HAM-$D_{17}$ reduction ≥ 50% and final GAS < 60<br>Relapse: return of depressive symptoms with HAM-$D_{17}$ ≥ 16<br>Measured at end of treatment (response) and 3 and 6 months post-treatment (relapse) | HAM-$D_{17}$<br>End of treatment (baseline), mean (SD)<br>ECT: 7.9 (4.5)<br>rTMS: 7.8 (3.7)<br>*P* = NS<br>3-month post-treatment, mean (SD)<br>ECT: 7.7 (5.0)<br>rTMS: 6.4 (4.9)<br>*P* = NS<br>6-month post-treatment, mean (SD)<br>ECT: 8.4 (5.6)<br>rTMS: 7.9 (7.1)<br>*P* = NS<br>3-month relapse, number (%)<br>ECT: 2 (10)<br>rTMS: 1 (5)<br>*P* = NS<br>6-month relapse, number (%)<br>ECT: 4 (20)<br>rTMS: 4 (19)<br>*P* = NS |

ECT, Electroconvulsive therapy; GAS, Global Assessment Scale; HAM-$D_{17}$, Hamilton Rating Scale for Depression, 17-item instrument; n, number; NS, not significant; pps, pulses per second; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation.

*MDD/bipolar.* One RCT compared 6-month remission rates for ECT and rTMS in 46 patients referred for ECT to treat a major depressive episode (Table 43).[54-56] Patients were not required to be treatment resistant, although on average patients had failed more than two previous adequate courses of medication—mean number (standard deviation) of failed treatments: ECT, 2.5 (1.4); rTMS, 2.4 (1). A small percentage of included participants had diagnoses of bipolar depression (9 percent or psychosis (15 percent). Patients continued their usual medical care and psychotropic medications, with no changes in medication allowed during their active treatment. ECT was administered twice weekly. The number of ECT treatments was based on response, as determined by the referring physicians. High frequency rTMS was administered for 15 consecutive weekday sessions. At the end of treatment, HAM-$D_{17}$ scores were statistically significantly lower for the ECT group than for the rTMS group ($P = 0.002$), and the ECT group had a greater percentage of patients in remission (59.1 percent vs. 16.7 percent, respectively; $P = 0.006$). After 6 months of follow-up, HAM-$D_{17}$ scores and remission rates were similar for the ECT and rTMS patients.

Table 43. Maintenance of remission in patients with major depression or bipolar disorder and probable treatment resistance by intervention; ECT vs. rTMS

| Author, year<br>Design<br>Quality | Intervention and Sample Size<br>Study Details | Maintenance of Remission |
|---|---|---|
| Eranti et al., 2007;[55] with McLoughlin et al., 2007[54] and Knapp et al., 2008[56]<br><br>RCT<br><br>Fair | ECT (n = 22; n = 12 for 6-month follow-up)<br>82% bilateral mean sessions 6.3 (2.5)<br><br>rTMS (n = 24; n = 4 for 6-month follow up)<br>High frequency, 15 sessions<br><br>Definitions<br>Remission: HAM-$D_{17}$ ≤ 8<br>Response: HAM-$D_{17}$ reduction ≥ 50%<br>Measured at end of treatment and 6-months after baseline (maintenance of remission) | HAM-$D_{17}$<br>Baseline, mean (SD)<br>ECT: 24.8 (5.0)<br>rTMS: 23.9 (7.0)<br>*P* = NS<br><br>End of treatment, mean (SD)<br>ECT: 10.7 (NR)<br>rTMS: 18.5 (NR)<br>*P* = 0.002<br><br>6-month (from baseline), mean (SD)<br>ECT: 13.8 (NR)<br>rTMS: 13.5 (NR)<br>*P* = NS<br><br>End of treatment remission, n (%)<br>ECT: 13 (59.1)<br>rTMS: 4 (16.7)<br>*P* = 0.006<br><br>6-month remission, n (%)<br>ECT: 6 (50)<br>rTMS: 2 (50)<br>*P* = NR |

ECT, Electroconvulsive Therapy; HAM-$D_{17}$, Hamilton Rating Scale for Depression, 17-item instrument; MT, motor threshold; n, number; NR, not reported; NS, not significant; pps, pulses per second; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation.

## Repetitive Transcranial Magnetic Stimulation vs. Sham

**Tier 1: Patients with two or more failures.** *MDD only.* No studies assessing maintenance of remission directly compared ECT, rTMS VNS, and CBT in patients in this group. No sham-controlled studies addressed this population for ECT, VNS, or CBT. Two rTMS RCTs using a sham procedure as control addressed maintenance of remission (longer-term relapse rates) in an MDD population (Table 44).[67,68,83]

Subjects in both rTMS trials were allowed to remain on psychotropic medications. The slightly larger and more recently conducted trial (n = 68)[67,83] compared 15 sessions of high-frequency rTMS at 110 percent motor threshold with 1600 pulses per session with a similarly delivered sham rTMS. At the end of treatment, responders could enter a 6-month follow-up to assess relapse. The smaller trial (n = 21)[68] compared 10 sessions of high-frequency rTMS at 80 percent motor threshold with 800 pulses per session with a similarly delivered sham rTMS. At the end of treatment responders could enter a 20-week follow-up.

In both trials, significantly more rTMS-treated than sham-treated participants were classified as responders: respectively, 30.6 percent vs. 6.1 percent; $P = 0.008$);[67,83] 50 percent vs. 22 percent; ($P < 0.05$)[68]. Of the small number of responders in these trials followed for maintenance of response, more than 50 percent relapsed; no statistically significant differences in relapse rates were observed between the rTMS and sham groups.

Table 44. Maintenance of remission in patients with major depressive disorder failing two or more previous treatments by intervention type; rTMS vs. sham

| Author, year<br>Design<br>Quality | Intervention and Sample Size<br>Study Details | Results on HAM-D Instruments |
|---|---|---|
| Avery et al., 2006[67,83]<br><br>RCT<br><br>Fair | rTMS (n = 35, 11 for relapse follow up)<br>High frequency, 15 sessions over 4 weeks<br><br>Sham (n = 33, 2 for relapse follow up)<br><br>Treatment strategy<br>Mixed-within group differences 31% of rTMS group and 27% of control group continued taking medications<br><br>Definitions<br>Remission definition: HAM-$D_{21}$ < 10 Response: HAM-$D_{17}$ reduction ≥ 50% Remission: HAM-$D_{17}$ < 8<br><br>Relapse: not defined<br><br>Measured at end of treatment (visit 16) and reassessed 1 week later (visit 17);<br><br>Response could enter 6-month follow up | $HAMD_{17}$<br>6-month relapse, n (%)<br>rTMS: 6 (54.5); 1 lost to follow-up<br>Sham: 1 (50); 1 lost to follow-up<br>*P* = NR |
| Boutros et al., 2002[68]<br><br>RCT<br><br>Fair | rTMS (n = 12, 6 for follow-up phase)<br>High frequency, 10 sessions<br><br>Sham (n = 9, 1 for follow-up phase)<br><br>Treatment strategy<br>Augmentation<br><br>Definitions<br>Response1 definition: >30% decrease in HAM-$D_{25}$<br><br>Response2 definition: ≥50% decrease in HAM-$D_{25}$**calculated from table<br><br>Relapse: HAM-$D_{25}$ ≥ baseline score ± 10%<br><br>Relapse measured up until 20 weeks | **HAM-$D_{25}$**<br>20-week relapse, n (%)<br>rTMS: 4 (66.6); 1 lost to follow-up<br>Sham: 1/1 (100)<br>*P* = NS |

HAM-D, Hamilton Rating Scale for Depression, 17-item instrument or 25-item instrument; n, number; NR, not reported; NS, not significant; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation.

*MDD/bipolar.* No trial addressed maintenance of remission with rTMS vs. sham therapy in this mixed MDD-bipolar population.

**Tier 2: Patients with one or more failures.** *MDD only.* No studies assessing maintenance of remission directly compared ECT, rTMS, VNS, and CBT. Two trials were relevant for this topic in this patient population (Table 45).

Two RCTs compared rTMS with sham rTMS and assessed maintenance of remission following active treatment.[71,72,100,101] One trial randomized 30 participants to 10 sessions of three different rTMS strategies (10 subjects in each group) and 15 participants to 10 sessions of similar sham strategies (5 subjects in each groups). The three treatment groups were high frequency delivered to the left dorsolateral prefrontal cortex (left high), low frequency delivered to the left dorsolateral prefrontal cortex (left low), and low frequency delivered to the right left dorsolateral prefrontal cortex (right low). At the end of treatment, the left high and right low treatment groups had similar reductions in HAM-$D_{21}$ scores, and these differences were statistically significantly greater than the left low and sham groups ($P < 0.001$). These differences remained after 2 weeks

Table 45. Maintenance of remission in patients with major depressive disorder failing one or more previous treatments by intervention; rTMS vs. sham

| Author, Year<br>Design<br>Quality | Intervention and Sample Size<br>Study Details | Maintenance of Remission |
|---|---|---|
| Stern et al., 2007[71]<br><br>RCT<br><br>Fair | **High rTMS (n = 10)**<br>10 sessions<br><br>Low-left rTMS (n = 10)<br>Low frequency (1Hz), Left-DLPFC, 10 sessions<br><br>Low rTMS (n = 10)<br>10 sessions<br><br>Sham (n = 15)<br><br>Treatment strategy<br>Switch<br><br>Definitions<br>Remission definition HAM-$D_{21}$ ≤ 10<br><br>Response and remission measured at end of treatment (2 weeks) and after 1 and 2 weeks follow-up | **HAM-$D_{21}$**<br>End of treatment score, mean (SD)<br>Left high rTMS: 15.1 (6)<br>Left low rTMS: 27.6 (5.9)<br>Right low rTMS: 15.8 (4.8)<br>Sham: 26.7 (3.6)<br>*P* < 0.001<br><br>2-week follow-up score, mean (SD)<br>Left high rTMS: 13.4 (5.6)<br>Left low rTMS: 26.6 (3)<br>Right low rTMS: 14.9 (5.9)<br>Sham: 26.8 (2.3)<br>*P* < 0.001<br><br>End of treatment response, n (%)<br>Left high rTMS: 5 (50)<br>Left low rTMS: 0 (0)<br>Right low rTMS: 5 (50)<br>Sham: 0/ (0%)<br>*P* = NR<br><br>2-week follow-up response, n (%)<br>Left high rTMS: 4 (40)<br>Left low rTMS: 0 (0)<br>Right low rTMS: 6 (60)<br>Sham: 0 (0)<br>*P* = NR<br><br>End of treatment remission, n (%)<br>Left high rTMS: 3 (33.3)<br>Left low rTMS: 0 (0)<br>Right low rTMS: 1 (10)<br>Sham: 0 (0)<br>*P* = NR<br><br>2-week follow-up remission, n (%)<br>Left high rTMS: 4 (40)<br>Left low rTMS: 0 (0)<br>Right low rTMS: 3 (33.3)<br>Sham: 0 (0)<br>*P* = NR |

DLPFC, dorsolateral prefrontal cortex; HAM-D, Hamilton Rating Scale for Depression, 21-item instrument; Hz, hertz; MT, motor threshold; n, number; NR, not reported; NS, not significant; pps, pulses per second; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation.

Table 45. Maintenance of remission in patients with major depressive disorder failing one or more previous treatments by intervention; rTMS vs. sham (continued)

| Author, Year<br>Design<br>Quality | Intervention and Sample Size<br>Study Details | Results |
|---|---|---|
| O'Reardon, 2007,[72] Janicak, 2007,[100] Solvason, 2007[101]<br><br>RCT<br><br>Fair | rTMS (n = 44 for follow-up phase)<br>High frequency, up to 30 sessions; rescue add-on permitted for symptom breakthrough (deterioration of CGI-S by 1 point over 2-week interval) during continuation<br><br>Sham (n = 23 for follow-up phase)<br><br>Treatment strategy<br>Acute treatment switch; continuation rescue was augment to current pharmacotherapy<br><br>Definitions<br>Relapse defined as recurrence of the full syndrome of major depression per DSM-IV over ≥ 2 weeks: HAM-$D_{17}$ ≥ 20; CGI-S ≥ 4 | HAM-$D_{17}$<br>Remission<br>Score at week 4, mean (SD)<br>rTMS: -14.6 (6.16)<br>Sham: -14.4 (6.11)<br><br>Relapse Rates:<br>Continuation at week 4, n (%)<br>rTMS: 1 (2.3)<br>Sham: 1 (4.4)<br>*P* = NR<br><br>Continuation at week 24, n (%)<br>rTMS: 3 (7.8)<br>Sham: 3 (15.0)<br><br>Continuation at week 4, n (%)<br>rTMS: 19(43.2)<br>Sham: 10(43.5)<br>*P* = NR<br><br>Continuation week 24, n (%)<br>rTMS: 18 (40.9)<br>Sham: 10(43.5)<br>*P* = NR |

of followup; no left low- or sham-treated participants were in remission after 2 weeks, whereas 40 percent of left high- and 33 percent of right low-treated participants were in remission after 2 weeks (*P* = NR).

Another trial followed 67 patients over 24-weeks to assess the durability of acute response to high frequency rTMS or sham.[72,100,101] The acute phase of this trial was a switch strategy that randomized 155 severely depressed participants to active rTMS and 146 severely depressed participants to sham rTMS.[72] After 6 weeks of acute treatment, 44 active rTMS-treated patients and 23 sham rTMS-treated patients were classified as responders. These patients entered a 3-week taper phase, and then began 24-weeks of open-label continuation follow-up.[100] Open-label rTMS was permitted as rescue augmentation to the current antidepressant regimen for symptom breakthrough. Relapse was defined as recurrence of the full syndrome of major depression per DSM-IV criteria observed over at least 2 weeks. After 24 weeks, 3 (15%) active rTMS-treated participants relapsed and 3 (7.8%) sham-treated participants relapsed (*P* = NR).

*MDD/bipolar.* One RCT compared rTMS with a sham procedure in 20 patients who had failed at least one adequate pharmacological trial during the current or previous episode (Table 46).[76] The majority of included patients (80 percent) had failed two or more medication trials during the current episode. The inclusion criteria allowed patients to have comorbid psychiatric diagnoses provided that the onset occurred after the development of major depression and that the symptoms of major depression were more prominent. This resulted in the inclusion of one patient (assigned to sham) with a bipolar II, depressed diagnosis; the remainder had unipolar major depression. Patients assigned to active treatment (n = 10) received 10 sessions of high frequency rTMS applied to the left dorsolateral prefrontal cortex. Patients assigned to the sham intervention (n = 10) received 10 sessions using the same device with the coil angled 30° to 45°

off the scalp and the bottom of the coil elevated 0.5 centimeter from the scalp. Response was defined by a 25-item HAM-D score ≤15 and a reduction in this score of 50 percent or more from baseline. At the end of treatment, one rTMS-treated patient (10%) and no sham-treated patients were categorized as responders (*P* = 0.09). The rTMS responder remained a responder during 2 months of follow-up.

Table 46. Maintenance of remission in patients with major depression or bipolar disorder who failed one or more previous treatments by intervention; rTMS vs. sham

| Author, year<br>Design<br>Quality | Intervention and Sample Size<br>Study Details | Maintenance of Remission |
|---|---|---|
| Berman et al., 2000[76]<br><br>RCT<br><br>Fair | **rTMS (n = 10; 1 for follow up phase)**<br>High frequency, 10 sessions<br><br>**Sham (n = 10; 0 for follow-up phase)**<br><br>Definitions<br>Response: HAM-$D_{25}$ ≤ 15 and reduction from baseline ≥ 50%<br><br>Response measured at end of treatment (2 weeks) and up to 2 months after treatment | **HAM-$D_{25}$**<br>End of treatment score, mean (SD)<br>rTMS: 24.6 (NR)<br>Sham: 36.4 (NR)<br>*P* < 0.01<br><br>End of treatment response, n (%)<br>rTMS: 1 (10)<br>Sham: 0 (0)<br>*P* = 0.09<br><br>2-month maintained response, n (%)<br>rTMS: 1 (100)<br>Sham: 0 (100)<br>*P* = NR |

HAM-D, Hamilton Rating Scale for Depression, 25-item instrument; n, number; NR, not reported; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation.

**Tier 3: Patients with probable treatment resistance.** *MDD only.* No trial addressed maintenance of remission with rTMS vs. sham therapy in an MDD Only population.

*MDD/bipolar.* One RCT compared rTMS with a sham procedure in 19 patients with unspecified drug resistance (Table 47).[77] The majority of patients had unipolar major depression, although 16 percent had bipolar depression. Patients assigned to active treatment (n = 12) received 5 sessions of high frequency rTMS applied daily to the left prefrontal cortex for 5 days. Patients assigned to the sham intervention (n = 7) received 5 similar sessions with the coil placed perpendicular to the scalp surface without direct contact. Depression severity was measured by the 24-item HAM-D and the 21-item BDI. At the end of treatment, rTMS-treated patients had significantly lower HAM-D and BDI scores than sham-treated patients (*P* < 0.001). This statistically significant difference was maintained through week-4 (3 weeks after end of treatment), but patients reverted to the previous depressed mood at week-12 (*P* = NS).

Table 47. Maintenance of remission in patients with major depression or bipolar disorder who failed one or more previous treatments by intervention; rTMS vs. sham

| Author, year<br>Design<br>Quality | Intervention, Sample Size, and Study Details | Maintenance of Remission |
|---|---|---|
| Bortolomasi et al., 2006 [77]<br>RCT<br>Fair | **rTMS (n = 12)**<br>High frequency, 90% MT, 800 pps, 5 sessions<br>**Sham (n = 7)**<br>Similar stimulation patterns with coil perpindicular to but not touching scalp<br>Definitions<br>Outcome = change in HAM-$D_{24}$ and $BDI_{21}$ | **HAM-$D_{24}$**<br>Baseline score, mean (SD)<br>rTMS: 25.17 (NR)<br>Sham: NR (NR)<br>*P* = NR<br>End of treatment (at week 1), mean (SD)<br>rTMS: 11.33 (NR)<br>Sham: 18.29 (NR)<br>*P* < 0.001<br>At week 4, mean (SD)<br>rTMS: 11.42 (NR)<br>Sham: 19.14 (NR)<br>*P* < 0.001<br>At week 12, (NR)<br>Both groups reverted to depressed mood<br>*P* = NS<br>**$BDI_{21}$**<br>Results similar to HAM-$D_{24}$ |

$BDI_{21}$, Beck Depression Inventory, 21-item instrument; HAM-$D_{24}$, Hamilton Rating Scale for Depression, 24-item instrument; MT, motor threshold; n, number; NR, not reported; pps, pulses per second; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation.

## Cognitive Behavioral Therapy vs. Usual Care

**Tier 1: Patients with two or more failures.** *MDD only*. No trial addressed maintenance of remission with CBT vs. usual care in an MDD Only population.

*MDD/bipolar.* No trial addressed maintenance of remission with CBT vs. usual care in an MDD-Bipolar mix population.

**Tier 2: Patients with one or more failures.** *MDD only.* No trial addressed maintenance of remission with CBT vs. usual care in an MDD Only population.

*MDD/bipolar.* No trial addressed maintenance of remission with CBT vs. usual care in an MDD-Bipolar mix population.

**Tier 3: Patients with probable treatment resistance.** *MDD only.* One trial, lasting 68 weeks and involving 158 participants, compared relapse rates for CBT and sham treatment (Table 48).[59,60,104] All participants received usual clinical management and antidepressant drug continuation throughout the study. In the CBT group, 80 participants received 16 sessions over a 20-week period, plus two booster sessions approximately 6 to 14 weeks later. The sham group was seen by a psychiatrist every 4 weeks during the first 20 weeks and then every 8 weeks thereafter. The relapse outcome was defined by two criteria. The first criterion was meeting the criteria from the Diagnostic and Statistical Manual, Version 3, Revised (DSM-III-R) for major depression for 1 month or more, with a HAM-$D_{17}$ score of 17 or higher on two successive visits 1 week apart. The second criterion, which was applied only during the follow-up phase, was persistent symptoms for 2 months or more with a HAM-$D_{17}$ score of 17 or higher at both visits. At the end of treatment (i.e., 20 weeks) and at 44 weeks, relapse rates were similar between CBT- and sham-treated participants. At the end of 68 weeks, significantly more sham-treated

participants than CBT-treated participants had relapsed. Based on the combined definition of major depression with persistent symptoms, 29 percent of CBT-treated participants and 47 percent of sham-treated participants had relapsed by 68 weeks (hazard ratio for relapse 0.54; 95% CI, 0.32-0.93; *P* = 0.02).

Table 48. Maintenance of remission in patients with major depressive disorder and probable treatment resistance by intervention; CBT vs. usual care

| Author, year<br>Design<br>Quality | Intervention, Sample Size, and Study Details | Maintenance of Remission |
|---|---|---|
| Paykel et al., 1999[59] Scott et al., 2000;[60] Scott et al., 2003[104]<br><br>RCT<br><br>Fair | **CBT plus clinical management (n = 80)**<br>16 session during 20 weeks<br><br>Sham plus clinical management (n = 78)<br><br>Definitions<br>Relapse:<br>1. HAM-$D_{17}$ ≥ 17 on 2 successive visits 1 week apart, OR<br>2. at follow-up for ≥ 2 months | **Relapse Rates,** number (%)<br>Major depression alone<br>At 20 weeks<br>CBT: 9 (11)<br>Sham: 14 (18)<br>*P* = NR<br><br>At 44 weeks<br>CBT: 15 (19)<br>Sham: 25 (31)<br>*P* = NR<br><br>At 68 weeks<br>CBT: 18 (22)<br>Sham: 29 (36)<br>*P* = 0.08<br>Hazard Ratio 0.58 (95% CI, 0.37-1.07)<br><br>**Relapse Rates,** number (%)<br>Major depression plus symptoms<br>At 20 weeks<br>CBT: 8 (10)<br>Sham: 14 (18)<br>*P* = NR<br><br>At 44 weeks<br>CBT: 19 (24)<br>Sham: 31 (40)<br>*P* = NR<br><br>At 68 weeks<br>CBT: 23 (29)<br>Sham: 37 (47)<br>*P* = 0.02<br>Hazard Ratio 0.54 (95% CI, 0.32-0.93) |

CBT, cognitive behavioral therapy; CI, confidence interval; HAM-$D_{17}$, Hamilton Rating Scale for Depression, 17-item instrument.

*MDD/bipolar.* No trial addressed maintenance of remission with CBT vs. usual care in an MDD-Bipolar mix population.

# Key Question 3: Efficacy or Effectiveness for Treating Treatment-Resistant Depression for Particular Symptom Subtypes

## Overview

This KQ focused on the comparative benefit of treatment for patients with TRD and an accompanying symptom subtype. Specifically of interest were symptom groups such as

psychosis, catatonia, or melancholy, subtypes that can accompany depression and which are often used to inform clinical interventions. We identified no studies that address this question in TRD (Tier 1) patients. However, a consideration of evidence from all tiers identified one relevant trial Tier 3 trial, reported in two articles.[103,105] The study was a head-to-head comparison of ECT and rTMS in psychotic and nonpsychotic patients with TRD.

## Strength of Evidence: Tier One Only

There is no evidence.

## Strength of Evidence: Considering all Tiers

The strength of evidence of the use of ECT vs. rTMS is insufficient. There is only one study identified, and treatment bias substantially affected validity in this trial: the ECT arm was allowed to continue psychotropic medications (including antipsychotics), while the rTMS group was not and the amount of treatments varied between the groups Table 49.

Table 49. Impact of on patients with TRD and an accompanying symptom sub-type; number of studies and subjects, and strength of the evidence domains—considering all tiers

| | Number of studies; subjects | Risk of bias Design/ Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| ECT vs. rTMS | 1; 40 | High<br>1 RCT<br>1 Fair | Unknown | Direct | Imprecise | Mixed results. ECT was significantly more effective than rTMS in the overall population and in psychotic patients, but differential use of antipsychotic with ECT (and not rTMS) substantially biased the results; ECT and rTMS were equally effective in nonpsychotic patients.<br>Insufficient. |

ECT, electroconvulsive therapy; RCT, randomized controlled trials; rTMS, repetitive transcranial magnetic stimulation; vs., versus.

## Key Points

In the one study available on this topic,[103,105] the effect of ECT was statistically significantly greater than that of rTMS. The presence of psychotic symptoms may have influenced the effect of these two interventions: psychotic patients appeared to have better outcomes with ECT than with rTMS. In nonpsychotic patients, the effect of the two interventions was similar. However, the two groups were being treated with different drugs at baseline; ECT patients were allowed any medication, including antipsychotics, at a stable rate but the rTMS patients were limited to clonazepam.

## Detailed Analysis

**ECT vs. rTMS.** There were no eligible studies in tier one or two. In tier three there were no eligible studies in a MDD only population and one study (3 articles) in a MDD/bipolar mix population.

**Tier 1.** There were no eligible studies.

**Tier 2.** There were no eligible studies.

**Tier 3.**There were no eligible studies in a MDD only population and one study (2 articles) in a MDD/bipolar mix population.

*MDD only.* There were no eligible studies.

*MDD/bipolar mix*. The study was undertaken with 40 inpatients and outpatients who had been referred for ECT (Table 50); detailed information is available in Evidence Table XX in Appendix E. The investigators randomized patients to either ECT or rTMS. Of those receiving ECT, 10 had TRD only and 10 had TRD and psychosis; of those receiving rTMS, 11 had TRD only and 9 had TRD and psychosis. The primary comparison was the change in HAM-D score at 2 weeks and end of treatment (approximately 4 weeks), with higher scores better than lower scores.

Table 50. Study assessing response to ECT or rTMS in patients with treatment-resistant depression (major depressive disorder) and psychosis

| Author, year<br>Design<br>Quality<br>Tier | Intervention and Sample Size<br>Study Details | Maintenance of Remission |
|---|---|---|
| Grunhaus et al., 2000[103]<br>RCT<br>Fair<br>Tier 3 | ECT (10 psychotic patients, 10 nonpyschotic patients 0)<br>Unilateral, could be switched if not responding after 6th txt<br><br>rTMS (11 psychotic patients, 9 nonpsychotic patients)<br>High Frequency rTMS<br>90% MT, 400-1200 pps, 20 sessions | **HAM-$D_{17}$**<br>Mean change in score, psychotic vs. nonpychotic patients:<br>At week 2<br>ECT: 16.0 vs. 5..5<br>rTMS: 5.3 vs. 7.7<br>ECT vs. rTMS in psychotic patients $P < 0.01$<br><br>Mean change in score, for psychotic vs. nonpsychotic patients:<br>At week 4<br>ECT: 23.1 vs. 9.3<br>rTMS: 7.9 vs. 12.5<br>ECT vs. rTMS in psychotics $P < 0.01$<br><br>Responders, % of psychotic vs. nonpsychotic:<br>ECT: 100 vs. 60<br>rTMS: 22 vs.63 |

ECT, electroconvulsive therapy; HAM-$D_{17}$, Hamilton Rating Scale for Depression, 17-item instrument; MT, motor threshold; pps, pilses per session; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; vs., versus.

Overall, patients responded better to ECT than to rTMS ($P < 0.05$). With regard to psychotic versus nonpsychotic patients, the study reported two important findings. First, in nonpsychotic patients, ECT and rTMS were equally effective. HAM-$D_{17}$ scores at the end of treatment for ECT and rTMS were 13.9 and 11.0 ($P$ = NS), respectively. Second, in psychotic patients, ECT appeared to be more effective than rTMS; HAM-$D_{17}$ scores at the end of treatment were 8.4 and 20.8 ($P = 0.01$), respectively

This study has limitations for our key question by virtue of treatment bias restricted applicability to our population of interest. The ECT group had been allowed to continue on any psychotropic medication, including antipsychotic medications, at a stable dose, while the rTMS

group had all their psychotropic medications discontinued although they were prescribed clonazepam (a benzodiazepine derivative with anticonvulsant, muscle relaxant, and anxiolytic properties) to reduce anxiety, limit insomnia, and help prevent seizures. This variation introduced a treatment, or co-intervention, bias. In this sample, 25 patients had been treated unsuccessfully two or more times and 15 patients either had been treated unsuccessfully only one time or had had no failures; nonetheless, all had been referred for ECT, and so we classified them as tier 3 (probable treatment resistance).

# Key Question 4: Organization of Safety, Adverse Events and Adherence

Key Question 4 contains information addressing safety, adverse events and adherence in the use of non-pharmacological treatments to treat TRD. The following section is split into four segments, each comparing the effects of the four nonpharmacologic interventions (ECT, rTMS, VNS, CBT/IPT) vs. each other (head to head comparisons) or vs. control interventions (e.g., sham procedures) but focusing on a different outcome. KQ 4a addresses the impact on cognitive functioning. KQ 4b examines specific adverse events (other than cognitive functioning) that were assessed systematically. The next two segments use two measures of study withdrawals. KQ 4c examines general tolerability to the treatments by using withdrawals specifically due to adverse events. The final segment, KQ 4d, examines adherence by examining withdrawals for any reason (overall withdrawals), as only a few studies measured adherence as an outcome.

## Key Question 4a: Cognitive Functioning—Overview

This KQ concerns the issue of whether the four nonpharmacologic interventions (ECT, rTMS, VNS, CBT/IPT) compared with each other (head to head comparisons) or against control interventions (e.g., sham procedures) have different effects on cognitive functioning. Cognitive functioning is measured in several domains, such as the Mini-Mental Status Examination (MMSE) and various intelligence, learning, or memory tests such as the Rey Auditory Verbal Learning Test (RALVT), the Wechsler Adult Intelligence Scale (WAIS), and the Cambridge Examination for Mental Disorders of the Elderly (and the cognitive, self-contained part of the Cambridge instrument denoted CAMCOG). Appendix xx[§§§] lists the major instruments used to detect or diagnose cognitive impairments across a wide range of faculties.

We included 9 studies of either good or fair quality; of these, four compared ECT to rTMS, and six evaluated rTMS against a sham procedure (Table 50). None had cognitive functioning as a primary outcome of interest. All tested cognitive functioning effects in the acute phase of treatment and did not address long term or cumulative effects of the interventions. In the detailed

[§§§] Appendixes are cited in this report and provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/___/___.pdf.

analysis section below, we consider first the studies involving only patients with MDD and then the mixed MDD-bipolar populations. For studies that did not report sufficient information to determine if the population was MDD only or a mixed MDD-bipolar population, we placed them in the mixed MDD-bipolar section. Information is presented for the three tiers used throughout this report: Tier 1, two or more treatment failures; Tier 2, one or more treatment failures; and Tier 3, "probable" treatment resistance.

Table 51. Number of good and fair quality studies by TRD tier and diagnostic mix

| Comparison and Population Included | Diagnosis | |
|---|---|---|
| | MDD only | MDD and Bipolar Disorder |
| **Tier 1: ≥ 2 failures** | | |
| ECT vs. rTMS | 2 | 0 |
| rTMS vs. Sham | 3 | 1 |
| Tier 2: ≥ 1 failure | | |
| rTMS vs. Sham | 1 additional | 0 |
| Tier 3: Probable | | |
| ECT vs. rTMS | 0 | 2 additional |

ECT, electroconvulsive therapy; MDD, major depressive disorder; rTMS, repetitive transcranial magnetic stimulation, vs., versus.

When considering only studies conducted in Tier 1 patients with MDD there were two head-to-head trials of ECT versus rTMS[51,106] and three rTMS versus sham studies (four articles).[61,64,67,83] There was one Tier 1 study, conducted in an MDD/bipolar population that compared rTMS to sham.[74]

Additional eligible studies were found in Tiers 2 and 3. One study (two articles) was conducted in Tier 2 patients with MDD.[69,70] Two head-to-head studies (four articles) in Tier 3 compare ECT with rTMS.[54-56,107]

## Strength of Evidence: Tier 1 Only

Table 52 shows the evidence for studies that are limited to Tier 1, patients that have previously failed two or more previous treatments for depression. The two studies that compare ECT versus rTMS, one a RCT and the other a cohort study showing some evidence that ECT has a negative impact on cognitive functioning compared rTMS during the acute phase of treatment. In the four studies that populate Tier 1 on comparisons of rTMS versus sham there appears to be some evidence that rTMS improves cognitive functioning compared to sham in cognitive functioning during acute phase treatment.

## Strength of Evidence: Considering all Tiers

Table 53 provides the strength of evidence considering all tiers which adds three additional studies. Adding two additional studies to the head-to-head comparisons provides evidence of low strength that ECT has a negative impact on cognitive functioning compared to rTMS. This impact appears to be transient. The addition of one rTMS versus sham study for a total of five rTMS-versus-sham comparisons found either no difference or greater improvements in cognitive functioning for subjects treated with rTMS compared to those treated with sham. Thus, there is some evidence that rTMS improves cognitive functioning (low strength of evidence). The total number of subjects represented and samples sizes of included studies were small and outcome measures and intervention regimens (duration and frequency of treatments) were fairly heterogeneous; future research could likely change these conclusions.

Table 52. Impact of ECT and rTMS on cognitive functioning in patients; number of studies and subjects; strength of evidence limited to Tier 1

| Comparison | Number of Studies; Subjects | Risk of bias Design Quality | Consistency | Directness | Precision | Results Strength of Evidence |
|---|---|---|---|---|---|---|
| ECT vs. rTMS | 2; 72 | Medium<br>1 RCT, and 1 prospective cohort study<br>Both fair | Inconsistent | Direct | Imprecise | Some evidence suggests that ECT has a deleterious impact on cognitive functioning compared to rTMS<br>Low |
| rTMS vs. sham | 4; 161 | Medium<br>4 RCTs,<br>1 good, 3 fair | Inconsistent | Direct | Imprecise | Some evidence that rTMS improves cognitive functioning compared to sham<br>Low |

ECT, electroconvul some addosive therapy; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; vs., versus.

Table 53. Impact of ECT and rTMS on cognitive functioning in patients; number of studies and subjects; strength of evidence

| Comparison | Number of Studies; Subjects | Risk of bias Design Quality | Consistency | Directness | Precision | Results Strength of Evidence |
|---|---|---|---|---|---|---|
| ECT vs. rTMS | 4; 146 | Medium<br>2 RCTs<br>1 fair, 1 good, 2 prospective cohort studies<br>Both fair | Inconsistent | Direct | Imprecise | Some evidence suggests that ECT has a deleterious impact on cognitive functioning compared to rTMS; this appears to be transient<br>Low |
| rTMS vs. sham | 5; 181 | Medium<br>5 RCTs<br>1 good, 4 fair | Inconsistent | Direct | Imprecise | Some evidence that rTMS improves cognitive functioning compared to sham<br>Low |

ECT, electroconvulsive therapy; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; vs., versus.

# Key Question 4a: Cognitive Functioning—Key Points

Limited evidence addressed the impact of these procedures on cognitive functioning; no evidence was available for VNS or CBT/IPT.

Overall, we included nine studies (13 articles) that examined cognitive functioning during acute phase treatment.[51,54-56,61,64,67,69,70,74 ,83,106,107] Six studies were limited to patients with MDD only, two comparing ECT with rTMS,[51,106] and four comparing rTMS with sham.[61,64,67,69,70,83] Three studies (five articles) included a mixed (20 percent or less bipolar) population; two studies (four articles) compared ECT with rTMS.[54-56,107] and one study compared rTMS to sham.[74]

Included studies are small; samples had a mean of 35 participants per study and ranged from 15[64] to 68[67,83] participants per study. Overall, cognitive functioning impacts did not differ much

between treatment groups. Some tests did show a statistically significant difference but not necessarily a clinically meaningful one.[61,64,67,69,70,83]

Any negative cognitive functioning impact that did occur with ECT faded away relatively quickly. Differences tended to dissipate to insignificance between end of treatment assessments and subsequent assessments (mean 8.8 days,[107] 2 weeks,[107] and 6 months[54-56]).

# Key Question 4a: Cognitive functioning—Detailed Analysis

## ECT vs. rTMS

There were two studies, one RCT and a prospective cohort study in tier one. There were no eligible studies in tier two. In tier three there were no MDD studies. In an MDD/bipolar mix population there was one RCT and one prospective cohort study.

**Tier 1.** There were two studies, one RCT and a prospective cohort study.

*MDD only.* Two studies shown in Table 54, provided data on the head-to-head comparison of ECT versus rTMS.[51,106] One was an RCT that compared right unilateral ECT for 2 weeks in 20 patients with high frequency rTMS in 22 patients[51]. At the end of treatment at 2 weeks and after a 2-week followup, for a total of 4 weeks, the groups did not differ on cognitive tests that included the Weschler Adult Intelligence Scale, Weschler Memory Scale, and the Rivermead Behavioral Memory Test.

The other was a prospective cohort study of 30 subjects.[106] The study used the Rey Auditory Verbal Learning Test (*RAVLT*), Memory for Persons Test, Autobiographical Memory Interview, Four card task, and the Squire Subjective Memory Questionnaire (SSMQ), to test cognitive functioning. Several of the cognitive tests showed a statistically significant difference between the ECT and rTMS groups, with ECT having a deleterious effect on cognitive functioning compared to rTMS.

Two sections of the RAVLT showed significant differences in post-treatment measures: recall after interference (ECT 3.9 vs. rTMS 1.8; $P < 0.01$), recall after delay (ECT 4.2 vs. rTMS 2.4; $P < 0.05$). Differences were also found in retrograde memory function. The ECT group made significantly more errors than those in the rTMS group in recognizing words learned before treatment (ECT 5.0 vs. rTMS 1.1, $P = 0.025$). After treatment, ECT recipients also recalled significantly fewer items (0.4) from the visual card task administered before treatment than did the rTMS group (1.4, $P = 0.012$). Subjective memory, measured using the SSMQ, improved in the rTMS group from -16.8 to 3.8 and stayed similar in the ECT subjects, changing from -20.7 to -15.2 at endpoint ($P < 0.05$ for rTMS vs. ECT).

*MDD/Bipolar mix.* There were no eligible studies.

**Tier 2.** There were no eligible studies.

**Tier 3.** There were no MDD studies and in an MDD/bipolar mix there was one RCT and one prospective cohort study (Table 55).

*MDD only.* There were no eligible studies.

*MDD/Bipolar mix***.** One RCT and one prospective cohort study provide head-to-head evidence comparing rTMS with ECT for mixed MDD-bipolar populations, as shown in table five.[54-56,107] The RCT compared high frequency rTMS (n = 22, for 15 sessions) versus ECT (n = 24, mean number of sessions 6.3, range 2-10, based on physicians' opinion).[54-56] The primary cognitive tests included the MMSE and Cambridge Examination for Mental Disorders in the Elderly (CAMCOG). There were no statistically significant differences in MMSE scores or total

Table 54. Cognitive functioning results from rTMS studies with ECT in tier one patients with major depressive disorder

| Author, Year<br>Design<br>Endpoint<br>Quality<br>Tier | Intervention and Sample Size<br>Study Details | Outcomes |
|---|---|---|
| Rosa et al, 2006[51]<br>RCT<br>Primary endpoint was after up to 4 weeks of active treatment<br>Fair | **ECT (n = 20)**<br>Right unilateral txt for 2 weeks, if pt not responding bilateral txt commenced<br>rTMS (n = 22)<br>High frequency, 100% MT, 2500 pps, 10-20 sessions (2-4 weeks) | WAIS-R, subsections of WMS (digit span) and RBMT; ECT vs. rTMS: no significant differences |
| Schulze-Rauschenbach et al. 2005[106]<br>Prospective cohort,<br>Outcomes measured 8.8 days on average after last treatment<br>Fair | **ECT (n = 14)**<br>Right unilateral txt for 2 weeks<br>rTMS (n = 16)<br>High frequency, 100% MT, 400-600 pps, mean 10.8 sessions (Range NR) | *Learning and Anterograde Memory with AVLT*<br>Recall after interference:<br>Before treatment<br>ECT: 2.8 (2.2) vs. rTMS: 3.2 (1.9)<br>1 week after treatment<br>ECT: 3.9 (1.9) vs. rTMS 1.8 (2.0) $P < 0.01$<br>Recall after delay:<br>Before treatment<br>ECT: 2.4 (1.8) vs. rTMS: 3.2 (1.6)<br>1 week after treatment<br>ECT: 4.2 (1.6) vs. rTMS: 2.4 (2.0) $P < 0.05$<br>Other AVLT subscales or the Memory for Persons Test (MPT):<br>No significant differences<br>*Retrograde memory with AVLT*<br>Recall:<br>No difference on recall or recognition hits<br>Recognition false alarms<br>One week after treatment:<br>ECT: 5.0 (3.0) vs. rTMS: 1.1 (1.1) $P < 0.05$<br>Four-card task - Free recall:<br>One week after treatment<br>ECT: 0.4 (0.5) vs. rTMS: 1.4 (1.2) $P < 0.05$<br>Subjective memory with SSMQ:<br>Before treatment:<br>ECT: -20.7 (19.0) vs. rTMS: -16.8 (16.9)<br>One week after treatment:<br>ECT: -15.2 (25.2) vs. rTMS: 3.8 (11.8) $P < 0.05$ |

AVLT, Auditory Verbal Learning Test; ECT, electroconvulsive therapy; MPT, memory persons test; MT, motor threshold; n, number; NR, not reported; pps, pulses per session; pt, patient; RBMT, Rivermead Behavioral Memory Test; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; txt, treatment; vs., versus; WAIS-R, Weschler Adult Intelligence Scale-Revised; WMS, Weschler Memory Scale.

Table 55. Cognitive functioning results in Tier 3 patients with major depressive disorder and bipolar disorder

| Author, year<br>Design<br>Quality | Intervention and Sample Size<br>Study Details | Outcomes |
|---|---|---|
| McLoughlin et al., 2007[54], Eranti et al., 2007[55] and Knapp et al., 2008[56]<br>RCT<br>Primary endpoint is end of treatment (at clinicians' discretion for ECT group, 3 weeks in rTMS)<br>Good | **ECT (n = 24)**<br>82% bilateral 18% unilateral<br>rTMS (n = 22)<br>High Frequency, 110% MT, 1000 pps, 15 sessions<br>Bipolar (%): 9 | CAMCOG Attention and orientation subscale (max = 17). Baseline ECT baseline 12.8 (3.2), end of treatment 13.9 (3.6), 6 mos 13.9 (3.5) rTMS baseline 14.7 (3.0) end of treatment 13.5 (3.3) FU 6 mos 13.4 (3.8), *P* = 0.004<br>No significant differences for rest of CAMCOG subscales (verbal fluency, anterograde memory, and retrograde memory)<br>Baseline MMSE n = ECT: 16 rTMS: 22<br>Baseline, mean (SD) ECT: 24.3(3.6)<br>rTMS: 25.7 (3.9)<br>Endpoint score, mean (SD)<br>End of treatment/6 month followup<br>ECT: 25.6 (3.9)/25.4 (5.3)<br>rTMS: 24.4 (5.3)/24.7 (4.8)<br>Change at end of treatment, mean:<br>ECT: 1.3 rTMS: -1.3 *P* < 0.08<br>No significant differences on the Columbia ECT Subjective Side Effects Schedule for self-reported cognitive side effects. |
| O'Connor et al. 2003[107]<br>Prospective cohort<br>Outcomes recorded at end of treatment and after 2 weeks of followup<br>Fair | **ECT (n = 14)** unilateral, three times per week for 2 to 3 weeks.<br>**rTMS (n = 14)** high frequency rTMS, 90% MT,1600 pps, 10 sessions<br>**Bipolar (%): NR*** | RAVLT, Acquisition, mean (SD). Baseline: ECT 43.78 (11.07) vs. rTMS 43.71 (12.09). End of treatment: ECT 29.14 (7.93) vs. rTMS 43.00 (10.09) *P* < 0.01. Two weeks later: ECT 46.92 (10.80) vs. rTMS 44.07 (10.43) *P* > 0.05. RAVLT, Retention, (15-item word list after a 20-minute delay interval), mean (SD).Baseline ECT 8.07 (4.49) words vs. rTMS 9.76 (3.08) End of treatment ECT 2.14 (1.99) vs. rTMS 8.23 (2.80) Two weeks later, ECT 8.92 (4.14) vs. rTMS 8.31 (4.07).<br>TNET. Baseline: ECT 64.30 (19.40) vs. rTMS 55.63 (18.12). End of treatment: ECT 39.10 (13.21) vs. rTMS 57.81(18.33).Two weeks later: ECT 59.20 (20.67) vs. rTMS 61.54 (19.12). |

CAMCOG, Cambridge Examination for Mental Disorders in the Elderly; ECT, electroconvulsive therapy; MMSE, mini-mental state examination; mos, months; MT, motor threshold; n, number; NR, not reported; pps, pulses per second; RAVLT, Rey Auditory Verbal Learning Test; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation; TNET, Transient News Events Test; vs., versus.

*In studies that did not report the percentage of bipolar patients and/or did not report it as an exclusion criteria we considered the study to have a mixed population consisting of patients with MDD and bipolar depression.

CAMCOG scores between the ECT group and the rTMS group. In addition, most of the CAMCOG subscales (verbal fluency, anterograde memory, and retrograde memory) showed no significant differences; but subjects treated with ECT did statistically significantly better than those treated with rTMS on the attention and orientation subscale (respectively, an increase of 1.1 from baseline vs. a decline of 1.2 from baseline; $P = 0.004$.

The prospective cohort study compared rTMS with ECT in 28 subjects (14 in each arm).[107] Testing was completed at baseline, at the end of treatment, and 2 weeks later. Cognitive function measures included the RAVLT and a revised version of the Transient News Events Test (TNET). The study reported transient negative cognitive side effects, most of which dissipate in the days after treatment, resulting from ECT in comparison with rTMS. New learning was assessed with the RAVLT for both acquisition and retention. For acquisition, at baseline ECT patients scored 43.78 and rTMS patients 43.71. At the end of treatment, ECT patients scored 29.14 and rTMS patients 43.00 ($P < 0.01$). At 2-week followup, scores for ECT subjects improved (46.92) and

those for rTMS patients remained stable (44.07), and differences were no longer significant. For retention, measured by the free recall portion of the RAVLT, a similar pattern was found. In addition, findings were similar for retrograde memory (measured by the TNET).

## rTMS vs. sham

Within Tier 1, there were three RCTs identified in an MDD patient population and one RCT in an MDD/bipolar mix population. One additional RCT in an MDD patient population was identified when accounting for a Tier 2 definition. For MDD/bipolar patients there were no eligible studies in Tier 2. Within Tier 3 in an MDD only population there were no eligible studies.

**Tier 1.** There were three RCTs in MDD patients and one RCT in patients with MDD/bipolar mix (Table 56).

Table 56. Cognitive functioning results in Tier 1 patients with major depressive disorder

| **Author, Year**<br>**Design**<br>**Endpoint**<br>**Quality**<br>**Tier** | **Intervention and Sample Size**<br>**Study Details** | **Outcomes** |
|---|---|---|
| Avery et al. 2006[67,83]<br>RCT<br>Outcomes measured after 2 weeks (except GOAT)<br>Good | **rTMS (n = 35)**<br>High frequency, 110% MT, 1600 pps, 15 sessions<br>**Sham (n = 33)** | No significant differences in tests - RAVLT, Digit Symbol Test and Digit Span (from the WAIS-R), Trail Making Test Parts A and B, MMSE, COWAT, the color Stroop Test: or GOAT, 5 minutes after each rTMS session |
| Holtzheimer et al. 2004[64]<br>RCT<br>Outcomes measured after 2 weeks<br>Fair | **rTMS (n = 7 )**<br>High frequency, 110% MT, 1600 pps, 10 sessions<br>**Sham (n = 8)** | Verbal Memory<br>RAVLT, Trial 7, mean score (%):<br>rTMS: 12.7 (2.1)<br>Sham: 12.0 (2.3) $P < 0.05$ No significant differences between groups<br>Neuropsychological measures of attention, verbal memory, psychomotor speed, and mental flexibility. Outcome measures: RAVLT subscales, Digit Symbol Test, Digit Span, and the Stroop Test: |
| Padberg et al., 1999[61]<br>RCT<br>Outcomes measured after 1 week<br>Fair | **Fast TMS (n = 6)**<br>High frequency, 90% MT, 250 pps, 5 sessions<br>Slow rTMS (n = 6)<br>**Low frequency,**<br>90% MT, 250 pps, 5 sessions<br>**Sham (n = 6)** | No significant differences between groups Learning performance<br>Fast rTMS; improvement, $P = 0.032$<br>Slow rTMS: no difference $P$ = NS.<br>Verbal memory performance<br>Fast rTMS vs. slow rTMS: no difference; $P$ = NS |

COWAT, Controlled Oral Word Association Test; GOAT, Galveston Orientation and Amnesia Test; MMSE, mini-mental state examination; MT, motor threshold; n, number; pps, pulses per second; RAVLT, Rey Auditory Verbal Learning Test; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; WAIS-R, Weschler Adult Intelligence Scale-Revised.

*MDD only.* Three Tier 1 RCTs as shown in Table six evaluated rTMS against sham. The largest (n = 68) used high frequency rTMS for 15 sessions and took cognitive measurements at baseline and following the final treatment. None of the tests showed a statistically significant difference between the two groups.[67,83] The other two studies were smaller. One (n = 15) used high frequency rTMS for ten sessions.[64]Tests included the RAVLT, Digit Symbol Test, Digit Span, and Stroop Test. Subjects in the two groups performed equally well with exception of one measure of verbal memory, Trial 7 of RAVLT, in which subjects who received rTMS performed

slightly better (12.7) than sham subjects (12.0, $P < 0.05$). Subjects treated with rTMS had mean neuropsychological tests that were either improved or equal to baseline levels of functioning. The other (n = 18) randomized subjects to five sessions of high frequency rTMS, low frequency rTMS, or sham.[61] Verbal memory performance improved after fast rTMS. Treatment groups showed no overall difference in verbal memory performance ($P$ = NS).

*MDD/bipolar mix.* One trial (Table 57) compared low frequency rTMS with high frequency rTMS for 10 sessions and with a sham procedure (n = 20 for each arm).[74] Cognitive testing was completed at baseline and at the end of the study (6 weeks). Measures included the Personal Semantic Memory Schedule, the Autobiographical Memory Schedule, the WAIS-R (block design test, verbal paired associates recall and recognition subscale, and digit span subscale), Tower of London, and the Controlled Oral Word Association Test. When combining all patients that received at least one type of active treatment, they found significant improvement on verbal paired associates ($P < 0.001$), verbal fluency ($P < 0.001$), and digit span forwards ($P = 0.003$) subscales; the Personal Semantic Memory Schedule ($P = 0.02$); and the Autobiographical Memory Schedule ($P = 0.05$).

Table 57. Cognitive functioning results from rTMS studies with sham comparison in Tier 1 patients with major depressive disorder and bipolar disorder

| Author, Year<br>Design<br>Endpoint<br>Quality | Intervention and Sample Size<br>Study Details | Outcomes |
|---|---|---|
| Fitzgerald et al., 2003[74]<br>RCT<br>Outcomes measured at baseline and end of study (6 weeks)<br>Fair | rTMS (n = 20)<br>Low frequency, 100% MT, 300 pps, 10 sessions<br>rTMS (n = 20)<br>High frequency, 100% MT, 1000 pps, 10 sessions<br>Sham (n = 20)<br>Bipolar (%): 10 | Verbal Paired Associates<br>Low rTMS and High rTMS: significant improvement from baseline on: Verbal Paired Associates ($P < 0.001$), Verbal fluency ($P < 0.001$), Digit span forwards ($P = 0.003$), subscales of the WAIS-R; Personal Semantic Memory Schedule ($P = 0.02$), and Autobiographical Memory Schedule ($P = 0.05$) |

MT, motor threshold; n, number; P, p-value; pps, pulses per session; RCT, randomized controlled trials; rTMS, repetitive transcranial magnetic stimulation; WAIS-R; Wechsler Adult Intelligence Scale-Revised.

**Tier 2.** There was one RCT in an MDD patient population. Within an MDD/bipolar population there were no eligible studies (Table 58).

*MDD only.* One RCT (n = 20) (two articles) compared high frequency rTMS intervals, for five sessions, with a sham procedure, see Table eight.[69,70] Cognitive testing was completed at baseline and three days after the last (5$^{th}$) treatment. The rTMS group showed a significant improvement in Trail Making Test B test scores (baseline score: 87.22; endpoint: 58.59; $P < 0.05$), whereas scores for the sham group did not significantly change. The groups did not differ significantly on any other cognitive tests conducted (MMSE, Trail Making Test A, The Stroop Test, WAIS-R Digit Symbol; Controlled Oral Word Association (COWA), Boston Naming Test, Sentence Repetition, RAVLT, or Judgment of Line Orientation).

*MDD/Bipolar mix.* There were no eligible studies.

**Tier 3.** In MDD only and MDD/bipolar populations there were no eligible studies.

Table 58. Cognitive functioning results from rTMS study with sham comparison in Tier 2 patients with major depressive disorder

| Author, Year<br>Design<br>Endpoint<br>Quality | Intervention and Sample Size<br>Study Details | Outcomes |
|---|---|---|
| Manes et al., 2001[69] and Moser et al., 2002[70]<br>RCT<br>Outcomes measured a mean of 3 days following last treatment<br>Fair | **rTMS (n = 10)** High Frequency rTMS<br>80% MT, 800 pps, 5 sessions<br>Sham (n = 10) | No significant differences between groups in MMSE, Trail Making Test A, The Stroop Test, WAIS-R Digit Symbol; COWA, Boston Naming Test, Sentence Repetition, RAVLT (% of learned words recalled after delay), Judgment of Line Orientation<br>Trail Making Test B, seconds<br>Baseline<br>rTMS: 87.22<br>Sham: 103.67<br>Follow up<br>rTMS: 58,59<br>Sham: 100.64 |

COWA, Controlled Oral Word Association; MMSE, mini-mental state examination; MT, motor threshold; pps, pulses per session; RAVLT, Rey Auditory Verbal Learning Test; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; WAIS-R, Wechsler Adult Intelligence Scale-Revised.

# Key Question 4b: Specific Adverse Events—Overview

This part of KQ 4 concerns specific adverse events from one of the procedural interventions recorded using a systematic method. Results are presented for good or fair quality studies.

Overall, five studies (eight articles) presented in Table 59,[54-56,67,76,80,83,87] assessed adverse events during acute phase treatment using a systematic method of which only three studies (four articles) found any significant differences in adverse events.[67,76,83,87]

Table 59. Number of good and fair quality studies by TRD tier and diagnostic mix that measure adverse events systematically

| Comparison and Population Included | Diagnosis | |
|---|---|---|
| | MDD only | MDD and Bipolar Disorder |
| **Tier 1: ≥ 2 failures** | | |
| rTMS vs. Sham | 1 | 0 |
| VNS vs. Sham | 0 | 1 |
| **Tier 2: ≥ 1 failure** | | |
| rTMS vs. Sham | 0 | 2 additional |
| **Tier 3: Probable** | | |
| ECT vs. Rtms | 0 | 1 additional |

## Strength of Evidence: Tier 1 Only

Table 60 documents the strength of evidence concerning specific adverse events in both disease categories combined, limited to Tier 1 studies. It includes two studies, one (2 articles) that compare rTMS to sham[67,83] and one that compares VNS to sham.[80] These studies provide low strength of evidence that both rTMS and VNS compared to sham lead to a greater incidence of adverse events. The strength of evidence, though it is low and subject to change with the addition of more studies.

Table 60. Specific adverse events studies; strength of evidence limited to Tier 1

| Comparison | Number of Studies; Subjects | Risk of Bias Design Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| rTMS vs. sham | 1; 68 | High<br>1 RCT<br>Good | Unknown, single study (as most of the specific adverse events were assessed by a single study) | Direct | Imprecise | Some evidence suggests that rTMS results in more scalp pain at the stimulation site<br>Low |
| VNS vs. sham | 1; 235 | Medium<br>1 RCT<br>Fair | Unknown, single study | Direct | Imprecise | No significant differences in specific adverse events<br>Low |

CBT, cognitive behavioral; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation, VSN, vagus nerve stimulation.

One RCT comparing VNS with sham provided low strength of evidence that there were no significant differences overall in the systematic assessment of specific adverse events, although the reporting of particular events appears to be numerically higher in the VNS group.

## Strength of Evidence: Considering all Tiers

The addition of Tier 2 and 3 studies adds a new comparison, a direct comparison of ECT vs. rTMS from one study described in three articles,[54-56] and two rTMS vs. sham studies[76,87] and more evidence to analysis of adverse events that are collected systematically, as seen in Table 61. For the comparison of ECT versus rTMS it shows low strength of evidence that there are no differences in adverse events between the two comparisons in acute phase treatment.[54-56] Adding in the rTMS versus sham studies adds another adverse event, difficulty in starting urination that are experienced by the rTMS group versus sham.[76]

Overall, these studies were typically small and were generally focused on assessing efficacy, rather than adverse events. They were not powered to detect small differences in rare adverse events.

# Key Question 4b: Specific Adverse Events—Key Points

Evidence on adverse events is very limited; only five studies (eight articles)[54-56,67,76,80,83,87] reported specific adverse events using a systematic method; four of these found some differences in adverse events.[67,76,80,83,87] This section does not include studies assessing cognitive function; those are addressed in Key Question 4a. The single good quality RCT, a head to head comparison of ECT vs. rTMS, did not report any significant differences in specific adverse events.[54-56] Three of the studies compared rTMS versus sham procedures; of these, one used escitalopram (20 mg) in both groups. These three studies provide some evidence that rTMS results in more scalp pain at the stimulation site, difficulty concentrating, and difficulty starting urination than sham, but that there is no difference in headaches or seizures.

Table 61. Specific adverse events studies; number of studies and subjects; strength of evidence

| Comparison | Number of Studies; Subjects | Risk of Bias Design/ Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| ECT vs. rTMS | 1; 46 | Low<br>1 RCT/good | Unknown, single study | Direct | Imprecise | No significant differences in specific adverse events<br>Low |
| rTMS vs. sham | 3; 137 | Medium<br>3 RCTs/ 1 good; 2 fair | Unknown, single study (as most of the specific adverse events were assessed by a single study) | Direct | Imprecise | Significant differences in some specific adverse events, for rTMS scalp pain and difficulty starting urination<br>Low |
| VNS vs. sham | 1; 235 | Medium<br>1 RCT/fair | Unknown, single study | Direct | Imprecise | No significant differences in specific adverse events<br>Low |

CBT, cognitive behavioral; ECT, electroconvulsive therapy; rTMS, repetitive transcranial magnetic stimulation; VNS, vagus nerve stimulation

One RCT comparing VNS with sham reported no significant difference in specific adverse events and no clear difference in serious adverse events. This study did report an increased frequency of particular events with VNS treatment--including voice alteration, cough, dyspnea, dysphasia, and neck pain—whose clinical importance is unclear. Further, possible complications from the implantation procedure (including serious cardiac adverse events, such as asystole or bradycardia, or infection), were also reported.

# Key Question 4b: Specific Adverse Events—Detailed Analysis

## ECT vs. rTMS

There were no eligible studies in Tier 1 or 2. In Tier 3 there were no eligible studies in an MDD only population and one study (3 articles) in an MDD/bipolar mix population.

**Tier 1.** There were no eligible studies.

**Tier 2.** There were no eligible studies.

**Tier 3.**There were no eligible studies in an MDD only population and one study (3 articles) in an MDD/bipolar mix population.

*MDD only***.** There were no eligible studies.

*MDD/bipolar mix.* Table 62 shows one head to head RCT that compared ECT (n = 24) with rTMS (n = 22) and did not report any significant differences in specific adverse events.[54-56] The study used the Columbia ECT Subjective Side Effects Schedule, modified to include potential rTMS side effects (e.g., seizure induction, scalp discomfort, hearing loss) and any upredictable side effects. The study reported that the ECT group had lower overall scores for subjective side effect symptoms after treatment ($P$ = 0.02), but did not report any differences for specific side effects. Additionally there was one death in rTMS arm, however it was unrelated to treatment.

Table 62. Adverse events assessed systematically in tier three patients diagnosed with major depressive disorder and bipolar disorder

| Author, Year<br>Design<br>Quality | Intervention and Sample Size<br>Study Details | Adverse Events (Pain, Concentration, Sleep) |
|---|---|---|
| Eranti et al., 2007,[55]<br>McLoughlin et al., 2007,[54]<br>and Knapp et al., 2008[56]<br>RCT<br>ECT CSSES modified<br>Good | **ECT (n = 24)**<br>82% bilateral 18% unilateral<br>rTMS (n = 22)<br>High frequency rTMS 110% MT, 1000 pps, 15 sessions | CSSES<br>ECT Base 14.2 (4.7) end 6.7 (6.4) 6 months 7.1 (4.7)<br>rTMS 13.2 (5.8) end 9.7 (4.6) 6 months 8.9 (4.7)<br>Group effect *P* = 0.02<br>No treatment related major adverse events recorded during study i.e., seizure induction, anesthetic complications, mania |

CCSES, Columbia Subjective Side Effects Schedule; ECT, electroconvulsive therapy; RCT, randomized controlled trial; rTMS, repetitive thranscranial magnetic stimulation.

## rTMS vs. sham

Tier 1 consists of one RCT in patients with a diagnosis of MDD and no studies in patients with a mixed diagnosis of MDD/bipolar. In Tier 2 there were no eligible studies in the MDD only population and two RCTs in an MDD/bipolar population. Within Tier 3 no eligible studies were identified.

**Tier 1.** There was one RCT in patients with a diagnosis of MDD and no studies in patients with a mixed diagnosis of MDD/bipolar.

*MDD only.* One RCT (N = 68) comparing high frequency rTMS to sham used the SAFTEE (Systematic Assessment for Treatment Emergent Effects) instrument to measure adverse events, as seen in Table 63.[67,83] The results showed no significant differences between rTMS and sham. Additionally it was reported that zero seizures occurred in subjects in both groups. However the rTMS group experienced more occassions of scalp pain at the stimulation site at session one, 41 percent and session 15, 33 percent than the sham group, 0 and three percent, respectively.

Table 63. Adverse events assessed systematically comparing rTMS to sham in Tier 1 patients diagnosed with major depressive disorder

| Author, Year<br>Design<br>Quality | Intervention and Sample Size<br>Study Details | Adverse Events (Pain, Concentration, Sleep) |
|---|---|---|
| Avery et al. 2006[67,83]<br>RCT<br>SAFTEE Scores<br>Good | **rTMS (n = 35)**<br>High frequency, 110% MT, 1600 pps, 15 sessions<br>Sham (n = 33) | Scalp pain at the stimulation site,%<br>Session 1:<br>rTMS: 41 vs. Sham: 0 *P* < 0.05<br>Session 15:<br>rTMS: 33 vs. Sham: 3 *P* < 0.05<br>Seizures, n:<br>rTMS: 0 vs. Sham: 0<br>Changes in SAFTEE (from baseline in 128 individual scores for any emerging symptoms that suggest adverse effects):<br>rTMS vs. sham *P* = NR (Data = NR) |

MT, motor threshold; n, number; NR, not reported; pps, pulses per session; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; SAFTEE, Systematic Assessment for Treatment Emergent Effects; vs., versus.

*MDD/Bipolar mix.* There were no eligible studies.

**Tier 2.** No eligible studies in the MDD only population and two RCTs in an MDD/bipolar population

*MDD only.* There were no eligible studies.

*MDD/Bipolar mix.* Table 64 shows the two studies that compare rTMS to sham in Tier 2 patients diagnosed with MDD and bipolar. One RCT reported no difference in trouble concentrating between rTMS and sham groups.[76] This study also compared adverse events using a multiple symptoms "Side Effect Checklist." Adverse events recorded include poor memory, nausea or vomiting, constipation, drowsiness, blurred vision, increased appetite, dry mouth, decreased appétit, tremors and shakiness, nightmares, difficulty sitting still, trouble concentrating, irregular or pounding heartbeat, diarrhea, frequent need to urinate, rash, ringing in the ears, sweating, faintness or lightheadedness, poor coordination, and muscle stiffness. Only one adverse event showed a significant difference between comparisons. "Difficulty starting urination" was reported significantly more often among the rTMS patients (2.0 vs. 1.1 ($P = 0.03$) (Table 63).[76]

Table 64. Adverse events assessed systematically comparing rTMS to sham in Tier 2 patients diagnosed with major depressive disorder and bipolar disorder

| Author, Year<br>Design<br>Quality | Intervention and Sample Size<br>Study Details | Adverse Events (Pain, Concentration, Sleep) |
|---|---|---|
| Berman et al., 2000[76]<br>RCT<br>SECL<br>Fair | **rTMS (n = 10)** High Frequency rTMS 80% MT, 800 pps, 10 sessions<br>**Sham (n = 10)** | Headache, %:<br>rTMS: 60 vs. Sham: 50<br>*P* = NR<br>Difficulty starting urination"(ordinal scores from 0, none at all, to 3, severe):<br>rTMS 2.0 vs. Sham 1.1<br>*P* = 0.03<br>No significant difference between groups after correction for multiple comparisons (data NR) |
| Bretlau et al. 2008[87]<br>RCT<br>UKU Side Effect Scale<br>Fair | **rTMS plus escitalopram (n = 25)**<br>High frequency rTMS 100% MT, 1280 pps, 15 sessions over 3 weeks, plus 20 mg escitalopram<br>**Sham plus escitalopram (n = 24)**<br>20 mg escitalopram | UKU side effect scale, mean scores<br>Concentration difficulties:<br>At Week 3, mean<br>rTMS: 1.43 vs. Sham: 1.52<br>At Week 12, mean<br>rTMS: 0.71 vs. Sham: 1.22 *P* < 0.05<br>Tension/inner unrest, tremor, akathisia, nausea, diarrhea, sweating, diminished sexual desire, headache, memory impairment, dry mouth, palpitations, and micturia"<br>No significant difference between groups |

mg, milligram; MT, motor threshold; n, number; NR, not reported; pps, pulses per session; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; SECL, Side Effect Checklist; UKU, Udvaig for Kliniske Undersogsler; vs., versus.

The other Tier 2 RCT (N = 49) compared rTMS with sham along with escitalopram (20 mg) in both groups and used the Udvaig for Kliniske Undersogsler (UKU) side effect scale to assess side effects.[87] Among the specific side effects assessed, they found no significant difference in headaches between groups. At 12-week followup, significantly more patients in the sham procedure group had difficulties concentrating than did rTMS patients (1.22 vs. 0.71 on 0 to 3 scale, $P < 0.05$).

**Tier 3.** There were no eligible studies.

## VNS vs. sham

There were no eligible studies in an MDD only population and one study in an MDD/bipolar population in Tier 1. There were no eligible studies in Tiers 2 or 3.

**Tier 1.** There were no eligible studies in an MDD only population and one study in an MDD/bipolar population.

*MDD only.* There were no eligible studies.

*MDD/Bipolar mix.* Table 65 shows a tier one RCT (N = 235) that compared VNS versus sham.[80] The study used the COSTART dictionary to assess adverse events. Many adverse events were listed but no statistical analysis was conducted in the article. Numerous adverse events were more commonly reported in the VNS group than the sham group. (*P* = NR) These included voice alteration (68% vs. 38%), cough increased (29% vs. 9%), dyspnea (23% vs. 14%), dysphasia (21% vs. 11%), and neck pain (21% vs. 10%) (for all *P* = NR). One participant underwent device explantation due to infection. Eleven patients (4 in VNS group and 7 in sham group) had worsening depression requiring hospitalization.

Table 65. Adverse events assessed systematically comparing VNS to sham in Tier 1 patients diagnosed with major depressive disorder and bipolar disorder

| Author, Year<br>Design<br>Quality | Intervention and Sample Size<br>Study Details | Adverse Events (Pain, Concentration, Sleep) |
|---|---|---|
| Rush et al., 2005[80]<br>RCT<br>COSTART dictionary.<br>Fair | **VNS (n = 115)**<br>10 weeks of stimulation 20Hz, 500 us pulse width, on/off cycle 30sec/5min<br>Output current began at .25mA and was increased in .25mA increments until a comfortable level was reached. Maximum current 3.5mA<br>**Sham (n = 110)**<br>Device implanted but not turned on | No significant differences in specific adverse events reported<br>Overall serious adverse events, n:<br>VNS: 16 vs. Sham:14<br>(12 events in 11 patients [VNS:4, Sham: 7] were cases of worsening depression requiring hospitalization) |

COSTART, Coding Symbols for Thesaurus of Adverse Reaction Terms; Hz, hertz; min, minute; mA, milliamps; n, number; RCT, randomized controlled trials; sec, second; VNS, vagus nerve stimulation.

**Tier 2.** There were no eligible studies.

**Tier 3.** There were no eligible studies.

# Key Question 4c: Tolerability as Measured by Withdrawals due to Adverse Events—Overview

Withdrawals due to an adverse event (AE) illustrate the general tolerability of treatments for treatment resistant depression. People that cannot tolerate the adverse effects of the treatments fall into this category. Overall, reporting of withdrawals due to AE was limited by the fact that no statistical significance was reported by the authors when withdrawals occurred.

Overall, 17 studies reported withdrawals due to adverse events (Table 66). When considering only studies conducted in TRD (Tier 1) MDD only patients, we identified one head-to-head trial of ECT vs. rTMS[106] and three rTMS versus sham studies (4 articles).[64,65,67,83] In a Tier 1

MDD/bipolar population, we identified four studies, conducted hat compared rTMS to sham[17,68,74,75] and one study, in the same population that compared VNS to sham.[80]

Table 66. Number of good and fair quality studies by TRD tier and diagnostic mix that assess withdrawals due to adverse events

| Comparison and Population Included | Diagnosis | |
|---|---|---|
| | MDD only | MDD and Bipolar Disorder |
| **Tier 1: ≥ 2 failures** | | |
| ECT vs. rTMS | 1 | 0 |
| rTMS vs. Sham | 3 | 4 |
| VNS vs. Sham | 0 | 1 |
| Tier 2: ≥ 1 failure | | |
| rTMS vs. sham | 2 additional | 0 |
| CBT vs. Usual Care | 0 | 2 additional |
| Tier 3: Probable | | |
| ECT vs. rTMS | 0 | 2 additional |
| rTMS vs. Sham | 0 | 2 additional |

ECT, electroconvulsive therapy; MDD, major depressive disorder; rTMS, repetitive transcranial magnetic stimulation; vs., versus.

Additional eligible studies were found in Tiers 2 and 3. Two studies (3 articles) were conducted in Tier 2 patients with MDD comparing rTMS to sham, [71,72,100] and two Tier 2 studies in patients with an MDD/Bipolar mix (3 articles) compared CBT versus usual care.[59,60,81] Two head-to-head studies (4 articles) in Tier 3 compare ECT with rTMS in a population diagnosed with MDD and bipolar disorder.[54-56,107]

## Strength of Evidence: Tier 1 Only

Few studies provide relevant data (Table 67). One small study showed greater withdrawals in ECT vs rTMS (statistical significance not reported), leading to a grade of low strength of evidence that withdrawals due to adverse events were greater with ECT than rTMS. In the rTMS vs. sham group the results are mixed, with the data not providing a clear direction of effect of the treatment on withdrawals due to adverse events, resulting in strength grade of insufficient. There was low strength of evidence that there were greater withdrawals due to adverse events in the vagus nerve stimulation group compared to sham.

## Strength of Evidence: Considering all Tiers

Including all tiers did not change the strength of evidence for the three comparisons available in Tier 1, but consideration did add a new comparison: CBT vs. Usual Care (Table 68). For ECT vs. rTMS considering all tiers added two studies, both MDD/Bipolar mix, indicating no difference in withdrawals between ECT and rTMS results, which changes the strength of evidence from low suggesting relative benefit for ECT to insufficient evidence addressing a difference between the two interventions. For rTMS vs. sham group, the addition of four studies still produced mixed results without a clear direction of effect, and strength of evidence remained insufficient. For VNS vs. sham, no studies were added. An important addition to the strength of evidence table is the comparison of CBT versus usual care, which adds two studies. The comparison shows high strength of evidence that there were no withdrawals due to adverse events in the CBT group.

Table 67. Strength of evidence for withdrawals due to adverse events limited to Tier 1

| Comparison | Number Of Studies; Subjects | Risk of Bias Design Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| ECT vs. rTMS | 1; 30 | Medium<br>1 fair prospective cohort studies | Unknown | Direct | Imprecise | No difference in withdrawals between ECT versus rTMS<br><br>*P* = NR<br><br>Low |
| rTMS vs. sham | 7; 277 | Medium<br>6 RCTs<br>1 good, 5 fair | Inconsistent | Direct | Imprecise | Mixed results<br>Insufficient |
| VNS vs. sham | 1; 222 | Low<br>RCT/good | Unknown | Direct | Precise | Greater withdrawals due to AE in the VNS group<br>*P* = NR<br>Low |

AE, adverse events; ECT, electroconvulsive therapy; NR, not reported; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; VNS, vagus nerve stimulation; vs., versus.

Table 68. Strength of evidence for withdrawals due to adverse events considering all tiers

| Comparison | Number Of Studies; Subjects | Risk of Bias Design Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| ECT vs. rTMS | 3; 104 | Medium<br>2 RCT<br>1 fair, 1 good and one fair prospective cohort studies | Inconsistent | Direct | Imprecise | Mixed results<br><br>Insufficient |
| rTMS vs. sham | 11; 690 | Medium<br>10 RCTs<br>1 good, 9 fair | Inconsistent | Direct | Imprecise | Mixed results<br>Insufficient |
| VNS vs. sham | 1; 222 | Low<br>RCT<br>Good | Unknown | Direct | Precise | Greater withdrawals due to AE in the VNS group<br>*P* = NR<br>Low |
| CBT vs. Usual Care | 2; 182 | Low<br>2 RCT/1 good, 1 fair | Consistent | Direct | Precise | No withdrawals due to AE<br>High |

AE, adverse events; CBT, cognitive behavioral therapy; ECT, electroconvulsive therapy; NR, not reported; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; VNS, vagus nerve stimulation; vs., versus.

# Key Question 4c:Tolerability as Measured by Withdrawals due to Adverse Events—Key Points

Withdrawals due to adverse events illustrate the general tolerability of treatments for treatment resistant depression. Overall, reporting of withdrawals due to AE was limited by the fact that no statistical significance was performed by the authors when withdrawals occurred.

# Key Question 4c:Tolerability as Measured by Withdrawals due to Adverse Events—Detailed Analysis

## ECT vs. rTMS

Tier 1 consists of one prospective cohort study in an MDD population and none in an MDD/bipolar population (Table 69). Tier 2 has no eligible studies. Tier 3 has no studies in MDD only patients and two RCTs in MDD/bipolar mix patients.

Table 69. Withdrawals due to adverse events comparing ECT to rTMS in Tier 1 patients diagnosed with major depressive disorder

| Author, year<br>Design<br>Duration<br>Quality | Intervention and sample size<br>Study Details | Discontinuations during treatment |
|---|---|---|
| Schulze-Rauschenbach et al. 2005[106]<br>Prospective cohort<br>1 week (post-test measurement 8.8 days after txt)<br>Fair | **ECT (n = 14)**<br>Right unilateral treatment, Mean charge of 2-2.5 times the threshold for two weeks, mean # sessions 9.9 (SD 2.7)<br><br>**rTMS (n = 16)**<br>High frequency, 100% MT, 400-600 pps, mean # sessions 10.8 (SD 1.4) | Due to AEs, n (%):<br>ECT: 1 (7.1)<br>rTMS: 0 (0)<br>*P* = NR |

ECT, electroconvulsive therapy; MT, motor threshold; NR, not reported; pps, pulses per second; SD, standard deviation; rTMS, repetitive transcranial magnetic stimulation; txt, treatment.

**Tier 1.** There was one prospective cohort study in an MDD population and none in an MDD/bipolar population.

*MDD only.* One fair quality prospective cohort study[106] adequately reported withdrawals due to AE. This observational study reported greater withdrawals in the ECT versus rTMS group (7.1% versus 0% respectively).[106] Sample sizes were small, all with less than 25 patients per study arm.

*MDD/Bipolar mix*. There were no eligible studies.

**Tier 2.** There were no eligible studies.

**Tier 3.** There were no studies in MDD only patients and two RCTs in MDD/bipolar mix patients (Table 70).

*MDD only.* There were no eligible studies.

*MDD/Bipolar mix.* Two RCTs (3 articles, one good quality and one fair quality).[54,55,107] They reported no withdrawals due to AE.

Table 70. Withdrawals due to adverse events comparing ECT to rTMS in Tier 3 patients diagnosed with major depressive disorder and bipolar disorder

| Author, year<br>Design<br>Duration<br>Quality | Intervention and sample size<br>Study Details | Discontinuations during treatment |
|---|---|---|
| Eranti et al., 2007,[55]<br>McLoughlin et al., 2007,[54]<br>and Knapp et al., 2008[56]<br>RCT<br>3 weeks<br>Good | **ECT (n = 22)**<br>Twice weekly, stimulus dosing method, 1.5 times the seizure threshold, mean # sessions 6.3 (SD 2.5)<br>**rTMS (n = 24)**<br>High frequency, 110% MT, 1000 pps, 15 sessions, mean # sessions 13.7 (SD 2.7) | Due to AEs: 0 |
| O'Connor, 2003[107]<br>Prospective cohort<br>Up to 4 weeks<br>Fair | **ECT (n = 14)**<br>Unilateral, three times per week for two to four weeks.<br>rTMS (n = 14)<br>High frequency, 90% MT, 1600 pps, 10 sessions | Due to AEs: 0 |

AE, adverse events; ECT, electroconvulsive therapy; MT, motor threshold; n, number; pps, pulses per second; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation

## rTMS vs. sham

Tier 1 contains three RCTs in patients with MDD only (Table 71) and four RCTs in MDD/bipolar patients (Table 72). There are two RCTs in an MDD only population and no eligible studies in MDD/bipolar diagnosis patients in tier two. In tier three there were no studies in MDD only patients and two RCTs in patients with an MDD/bipolar mix diagnosis.

**Tier 1.** There are three RCTs in patients with MDD only and four RCTs in MDD/bipolar patients.

*MDD only.* Three fair RCTs that reported no withdrawals in either patients treated with rTMS or sham.[64,65,67,83]

*MDD/Bipolar mix.* Four fair RCTs reported withdrawals due to adverse events in patients previously treated two or more times previously for depression. Three of the studies showed no withdrawals due to adverse events.[17,68,74] There was one study that showed a difference in withdrawals due to adverse events rTMS 9.1 percent versus none for sham.[75] There are important differences between this study and the others in this group primarily in the strength of the intervention. As can be seen in table xx, the RCT that showed differences in withdrawals due to adverse events used more pulses per session, 1600 versus 750 to 1000 and 20 hertz versus 10 hertz, which could explain the differences in withdrawals due to adverse events within this group.

**Tier 2.** There are two RCTs in an MDD only population and no eligible studies in MDD/bipolar diagnosis patients (Table 73).

*MDD only.* One relatively large (n = 325)study compared tier two patients in an MDD only population.[72,100,101] The withdrawals due to adverse events were similar in the rTMS group (4.2 percent) versus sham (3.4 percent) over the four week time period. Additionally a small study (n = 45) compared withdrawals due to adverse events in four arms, high rTMS (n = 10), low-left rTMS (n = 10), low-right rTMS (n = 10) and sham (n = 15).[71]Two arms had no withdrawals but the low-left rTMS had a 50 percent withdrawals due to adverse event rate and 30 percent in the sham group.

Table 71. Withdrawals due to adverse events comparing rTMS to sham in Tier 1 patients diagnosed with major depressive disorder

| Author, year<br>Design<br>Duration<br>Quality | Intervention and sample size<br>Study Details | Discontinuations during treatment |
|---|---|---|
| Avery et al. 2006[67] and Avery et al. 2007[83]<br>RCT<br>4 weeks<br>Good | **rTMS (n = 35)**<br>High Frequency rTMS<br>110% MT, 1600 pps, 15 sessions<br>Sham (n = 33) | Due to AEs: 0 |
| Garcia-Toro et al. 2006[65]<br>RCT<br>2 weeks<br>Fair | **rTMS-1 (n = 10)**<br>High frequency, 110% MT, 1200 pps, 10 sessions plus low frequency 110%, 18000 pps, 10 sessions<br>rTMS-2 (n = 10)<br>Same as above but with individualize location<br>Sham: (n = 10) | Due to AEs: 0 |
| Holtzheimer et al. 2004[64]<br>RCT<br>2 weeks<br>Fair | **rTMS (n = 7 )**<br>High frequency, 110%, 1600 pps, 10 sessions<br>Sham (n = 8) | Due to AE:,0 |

AE, adverse events; MT, motor threshold; n, number; pps, pulses per second; RCT, randomized controlled trials, rTMS, repetitive transcranial magnetic stimulation.

Table 72. Withdrawals due to adverse events comparing rTMS to sham in Tier 1 patients diagnosed with major depressive disorder and bipolar disorder

| Author, year<br>Design<br>Duration<br>Quality | Intervention and Sample Size<br>Study Details | Discontinuations during treatment |
|---|---|---|
| Boutros, et al., 2002[68]<br>RCT<br>2 weeks<br>Fair | **rTMS (n = 12)**<br>High frequency rTMS, 80% MT, 800 pps, 10 sessions<br>Sham (n = 9) | Overall<br>rTMS: 9.1%<br>Sham: 8.6%<br>*P* = NR<br>Due to AEs: 0 |
| Fitzgerald et al. 2006[17]<br>RCT<br>6 week<br>Fair | **High and low rTMS (n = 25)**<br>High frequency, 100% MT, 750 pps, 10 sessions plus low requency 110%, 420 pps, 10 sessions<br>**Sham (n = 25)** | Due to AEs: 0 |
| Fitzgerald et al. 2003[74]<br>RCT,<br>Phase I: 2 weeks<br>Phase II: NA<br>Fair | **rTMS (n = 40)**<br>Low frequency, 100%, 300 pps, 10 sessions or high frequency, 100%, 1000 pps, 10 sessions<br>**Sham (n = 20)** | Due to AEs: 0 |
| Su et al., 2005[75]<br>RCT<br>2 weeks<br>Fair | **rTMS (n = 22)**<br>High frequency (5Hz), 100% MT, 1600 pps, 10 sessions or high frequency (20Hz) 100% MT, 1600 pps, 10 sessions (left)<br>Sham (n = 11) | Due to AEs, %:<br>All rTMS: 9.1<br>High rTMS: 0<br>Low rTMS: 17<br>Sham: 0 |

AE, adverse events; Hz, hertz; MT, motor threshold; n, number; pps, pulses per second; RCT, randomized controlled trials, rTMS, repetitive transcranial magnetic stimulation.

Table 73. Withdrawals due to adverse events comparing rTMS to sham in Tier 2 patients diagnosed with major depressive disorder

| Author, year<br>Design<br>Duration<br>Quality | Intervention and sample size<br>Study Details | Discontinuations during treatment |
|---|---|---|
| O'Reardon, 2007,[72] Janicak, 2007[100] and Solvason, 2007[101]<br>RCT<br>4 weeks primary end pt<br>Fair | **rTMS (n = 165)**<br>High frequency, 120% MT, 3000 pps, 20-30 sessions<br>Sham (n = 160) | Due to AEs:, %:<br>rTMS: 4.2<br>Sham 3.4<br>*P* = NR |
| Stern et al., 2007[71]<br>RCT<br>2 weeks of txt<br>Fair | **1. High rTMS (n = 10)**<br>High frequency, 110%, 1600 pps, 10 sessions<br>2. Low-left rTMS (n = 10)<br>Low frequency, (Left-DLPFC), 110% MT, 1600 pps, 10 sessions<br>3. Low-right rTMS (n = 10)<br>Low frequency, 110%, 1600 pps, 10 sessions<br>4. Sham (n = 15) | Due to AEs, %:<br>High rTMS: 1. 0<br>Low-left rTMS: 2. 50 (5/10)<br>Low-right rTMS: 3. 0<br>Sham: 4. 20 (3/15) |

AE, adverse events; DLPFC, dorsolateral prefrontal cortex; MT, motor threshold; n, number; NR, not reported; pps, pulses per second; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; txt, treatment.

*MDD/Bipolar mix.* There were no eligible studies.

**Tier 3.** There were no studies in MDD only patients and two RCTs in patients with an MDD/bipolar mix diagnosis (Table 74).

Table 74. Withdrawals due to adverse events comparing rTMS to sham in Tier 3 patients diagnosed with major depressive disorder and bipolar disorder

| Author, year<br>Design<br>Duration<br>Quality | Intervention and sample size<br>Study Details | Discontinuations during treatment |
|---|---|---|
| Bortolomasi et al. 2006[77]<br>RCT<br>1 week<br>Fair | **rTMS (n = 12)**<br>High frequency, 90% MT, 800 pps, 5 sessions<br>**Sham (n = 7)** | Due to AEs: 0 |
| George et al. 1997[79]<br>RCT, crossover<br>Primary endpoint after 2 weeks of txt<br>Fair | **rTMS (n = 12)**<br>High frequency, 80% MT, 800 pps, 10 sessions<br>**Sham (n = 12)** | Due to AEs: 0 |

AE, adverse events; MT, motor threshold; n, number; pps, pulses per second; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; txt, treatment.

*MDD only.* There were no eligible studies.

*MDD/Bipolar mix.* There were two small studies (n = 19 and 24) in tier three MDD/Bipolar mix population comparing rTMS to sham.[77,79] Neither had any withdrawals due to adverse events.

## VNS vs. sham

There were no eligible studies in patients with MDD only and one RCT in patients with an MDD/bipolar diagnosis in tier one. In tiers two and three there were no eligible studies.

**Tier 1.** There were no eligible studies in patients with MDD only and one RCT in patients with an MDD/bipolar mix diagnosis (Table 75).

Table 75. Withdrawals due to adverse events comparing VNS to sham in Tier 1 patients diagnosed with major depressive disorder and bipolar disorder

| Author, year<br>Design<br>Duration<br>Quality | Intervention and sample size<br>Study Details | Discontinuations during treatment |
|---|---|---|
| Rush et al., 2005[80],<br>RCT<br>10 weeks<br>Good | **VNS (n = 112)**<br>10 weeks of VNS therapy with continued medications. Medications could change after acute treatment phase<br>Sham (n = 110) | Due to AEs, %:<br>VNS: 2.7<br>Sham: 0 |

AE, adverse events; n, number; RCT, randomized controlled trial; VNS, vagus nerve stimulation

*MDD only.* There were no eligible studies.

*MDD/bipolar mix.* One good quality RCT (N = 222) comparing VNS to sham-control in a tier one population reported 2.7 percent withdrawals due to AE in the VNS group compared with none in the sham-control group over a 10 week treatment period.[80]

**Tier 2.** There were no eligible studies.

**Tier 3.** There were no eligible studies.

## CBT vs. Usual Care

There were no eligible studies in Tier 1. In an MDD population there were no eligible studies and 2 studies in an MDD/bipolar mix population in Tier 2. There were no eligible studies in Tier 3.

**Tier 1.** There were no eligible studies.

**Tier 2.** There were no eligible studies in MDD only and 2 studies in patients with an MDD/bipolar mix.

*MDD only***.** There were no eligible studies.

*MDD/Bipolar mix.*Two RCTs (4 articles, 1 good quality, 1 fair quality) comparing CBT to some form of usual care reported no withdrawals due to AE, as shown in Table 76.[59,60,81,104]These studies ranged in duration from 16 weeks treatment to 12-month followup periods.

**Tier 3.**There were no eligible studies.

# Key Question 4d: Adherence as Measured by Overall Withdrawals—Overview

Of 49 included studies, the three studies reporting compliance indicated 100 percent rate.[58,65,103] Overall withdrawals were used as a proxy to capture compliance as it was recorded more frequently. Out of the 49 included studies 20 studies (22 articles) reported total withdrawals (for any reason) during treatment (Table 77).

Table 76. Withdrawals due to adverse events comparing rTMS to sham in Tier 2 patients diagnosed with major depressive disorder and bipolar disorder

| Author, year<br>Design<br>Duration<br>Quality | Intervention and sample size<br>Study Details | Discontinuations during treatment |
|---|---|---|
| Harley, 2008[81]<br>RCT<br>16 weeks<br>Fair | **Dialectical Behavioral Therapy Skills Training (n = 13)** 16-session, once weekly group covered the 4 dialectical behavior therapy skill sets<br>Usual care (n = 11) | Due to AEs: 0 |
| Paykel, 1999,[59] Scott, 2000,[60] and Scott, 2003[104]<br>RCT<br>20 weeks<br>Good | **Cognitive Therapy plus Clinical Management (n = 80)**<br>16 CT sessions during a 20 week period<br>Clinical Management (n = 78)<br>Pts visited psychiatrist every 4 weeks and continued on current medication; seen every 8 wks during followup phase | Due to AEs: 0 |

AE, adverse events; CT, cognitive therapy; n, number; n, number; Pts, patients; RCT, randomized controlled trial; wks, weeks.

Table 77. Number of good and fair quality studies by TRD tier and diagnostic mix that assess overall withdrawals

| Comparison and Population Included | Diagnosis | |
|---|---|---|
| | MDD only | MDD and Bipolar Disorder |
| **Tier 1: ≥ 2 failures** | | |
| ECT vs. rTMS | 2 | 0 |
| rTMS vs. Sham | 3 | 4 |
| Tier 2: ≥ 1 failure | | |
| rTMS vs. Sham | 2 additional | 2 additional |
| CBT vs. Usual Care | 0 | 3 additional |
| Tier 3: Probable | | |
| ECT vs. rTMS | 0 | 2 additional |
| rTMS vs. Sham | 0 | 2 additional |

CBT, cognitive behavioral therapy; ECT, electroconvulsive therapy; rTMS, repetitive transcranial magnetic stimulation; vs., versus.

When considering only studies conducted in Tier 1 patients with MDD there were two head-to-head trials of ECT versus rTMS[51,106] and three rTMS versus sham studies (4 articles).[64,65,67,83] There were four tier one studies, conducted in an MDD/bipolar population that compared rTMS to sham.[17,68,74,75]

Additional eligible studies were found in Tiers 2 and 3. In Tier 2 MDD only populations, we identified two studies (three articles) comparing rTMS to sham[71,72,100] and one study comparing CBT to usual care. [86] In Tier 2 MDD/Bipolar mix populations, we identified two studies (3 articles) comparing CBT versus usual care.[59,60,81] Two head-to-head studies (four articles) in Tier 3 compare ECT with rTMS in a population diagnosed with MDD and bipolar disorder.[54-56,107] There are also two Tier 3 studies that compare rTMS to sham in a MMD and bipolar population.[77,79]

## Strength of Evidence: Tier 1 Only

The data addressing overall withdrawals is inconclusive, as can be seen in Table 78. The results are mixed for both ECT vs. rTMS and rTMS vs. sham, and the strength of evidence is insufficient make a conclusion regarding any difference between intervention arms.

Table 78. Strength of evidence for overall withdrawals during treatment limited to Tier 1

| Comparison | Number Of Studies; Subjects | Risk of Bias Design Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| ECT vs. rTMS | 2; 72 | Medium<br>1 RCT<br>Fair and one fair prospective cohort studies | Inconsistent | Direct | Imprecise | Mixed results<br>Insufficient |
| rTMS vs. sham | 7; 534 | Medium<br>7 RCTs/ fair | Inconsistent | Direct | Imprecise | Mixed results<br>Insufficient |

ECT, electroconvulsive therapy; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; vs., versus.

## Strength of Evidence: Considering all Tiers

As shown in Table 79, the addition of Tier 2 and Tier 3 studies does not change the overall results of insufficient strength of evidence in the analysis of overall withdrawals during treatment. The addition of two studies to the ECT versus rTMS continues to show mixed results. The five additional studies in the rTMS versus sham also continue to show mixed results. The expanded strength of evidence table does allow for the inclusion of CBT versus usual care, three studies that show mixed results in overall withdrawals during treatment, leading to an insufficient grade for strength of evidence.

Table 79. Strength of evidence for overall withdrawals during treatment considering all tiers

| Comparison | Number Of Studies; Subjects | Risk of Bias Design Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| ECT vs. rTMS | 4; 146 | Medium<br>3 RCT<br>2 fair, 1 good<br>1 fair prospective cohort study<br>1 fair | Inconsistent | Direct | Imprecise | Mixed results<br>Insufficient |
| rTMS vs. sham | 13; 646 | Medium<br>13 RCTs<br>1 good and 12 fair | Inconsistent | Direct | Imprecise | Mixed results<br>Insufficient |
| CBT vs. usual care | 3; 214 | Medium<br>3 RCT<br>1 good, 2 fair | Inconsistent | Direct | Imprecise | Mixed results<br>Insufficient |

CBT, cognitive behavioral therapy; ECT, electroconvulsive therapy; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; vs., versus.

# Key Question 4d: Adherence as Measured by Overall Withdrawals—Key Points

There are only three studies that assessed adherence or compliance during treatment. All three reported that all patients completed all required treatments as specified in the protocol.

Overall, reporting of withdrawals during treatment was limited by the fact that statistical significance was not reported. Studies were generally small, unlikely to have had power to show statistical or clinical significance, methods varied and there was significant heterogeneity across the populations studied.

# Key Question 4d: Adherence as Measured by Overall Withdrawals—Detailed Analysis

Adherence/compliance, as shown in Table 80, there were only three studies that reported adherence or compliance.[58,65,102,103] All of them reported 100 percent compliance.

Table 80. Adherence/compliance

| Author, year<br>Design<br>Quality | Intervention and Sample Size<br>Study Details | Results |
|---|---|---|
| Garcia-Toro et al. 2006[65]<br>RCT: outpatient<br>2 weeks<br>Fair (Tier 1) | **rTMS-1 (n = 10)**<br>High frequency, 110% MT, 1200 pps, 10 sessions plus low frequency 110%, 18000 pps, 10 sessions<br>**rTMS-2 (n = 10)**<br>Same as above but with individualized location<br>**Sham: (n = 10)** | All completed 10 rTMS sessions |
| Folkerts et al.[58]<br>RCT: patient status NR<br>4 weeks<br>Fair (Tier 1) | **ECT (n = 21)**<br>Unilateral<br>**Paroxetine (n = 18)**<br>Mean daily dosage 44 mg/day | All patients continued their respective therapies through scheduled end of treatment phase |
| Grunhaus et al., 2000[103] and Dannon et al., 2002[102]<br>RCT: Inpatient (n = 32) and outpatient (n = 8)<br>Fair (Tier 3) | **ECT (10 psychotic patients, 10 nonpyschotic patients)**<br>Unilateral, could be switched if not responding after 6th txt<br>rTMS (11 psychotic patients, 9 nonpsychotic patients)<br>High frequency rTMS<br>90% MT, 400-1200 pps, 20 sessions | All patients completed all treatments |

ECT, electroconvulsive therapy; mg/day, milligram per day; MT, motor threshold; n, number; NR, not reported; pps, pulses per second; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; txt, treatment.

## Overall Withdrawals: ECT vs. rTMS

There were two studies, one RCT and a prospective cohort study in tier one. There were no eligible studies in tier two and in tier three there were no MDD studies and in an MDD/bipolar mix there was one RCT and one prospective cohort study.

**Tier 1.** There were two studies, one RCT and a prospective cohort study.

*MDD only.* There are two Tier 1 studies that compare ECT to rTMS and reported overall withdrawals, as seen in Table 81. The first is a small RCT (n = 42) that resulted in more withdrawals in the ECT group of 15.1 percent than the rTMS group at 9.1percent (*P* = NR).[51] The second study was a small prospective cohort study (N = 30).[106] Similar to the RCT it showed that the ECT group experienced higher overall withdrawals of 7.1 percent versus 0 percent in the rTMS group but significance is not reported.

Table 81. Overall withdrawals comparing ECT to rTMS in Tier 1 patients diagnosed with major depressive disorder

| Author, year<br>Design<br>Quality | Intervention and Sample Size<br>Study Details | Results |
|---|---|---|
| Rosa et al., 2006[51]<br>RCT<br>Up-to 4 weeks<br>Fair | **ECT (n = 20)**<br>Right unilateral txt for two weeks, if pt not responding bilateral txt commenced<br><br>rTMS (n = 22)<br>High frequency, 100% MT, 2500 pps, 10-20 sessions (2-4 weeks) | Overall, %:<br>ECT 15.0<br>rTMS: 9.1 |
| Schulze-Rauschenbach et al. 2005[106]<br>Prospective cohort<br>1 week (post-test measurement 8.8 days after txt)<br>Fair | **ECT (n = 14)**<br>Right unilateral treatment, mean charge of 2-2.5 times the threshold for two weeks, mean # sessions 9.9 (SD 2.7)<br><br>**rTMS (n = 16)**<br>High frequency, 100% MT, 400-600 pps, mean # sessions 10.8 (SD 1.4) | Overall, %:<br>ECT: 7.1<br>rTMS: 0 |

ECT, electroconvulsive therapy; Mt, motor threshold; pps, pulses per second; pt, patient; RCT, randomized controlled trials; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation; txt, treatment.

*MDD/Bipolar mix.* There were no eligible studies.

**Tier 2.** There were no eligible studies.

**Tier 3.** There were no MDD studies and in an MDD/bipolar mix there was one RCT and one prospective cohort study.

*MDD only.* There were no eligible studies.

*MDD/Bipolar mix.* As shown in Table 82, two studies report overall withdrawals in a Tier 3 population comparing ECT to rTMS. A good-rated RCT reported overall withdrawals in the ECT group of zero percent compared to 25 percent in the rTMS arm (*P* = NR).[54-56] A small prospective cohort reported no overall withdrawals in either arm.[107]

## Overall Withdrawals: rTMS vs. sham

Tier 1 contains three RCTs in patients with MDD only and four RCTs in MDD/bipolar patients. There are two RCTs in an MDD only population and two eligible studies in an MDD/bipolar population in tier two. In Tier 3 there were no studies in MDD only patients and two RCTs in patients with an MDD/bipolar mix population.

**Tier 1.** *MDD only.* There are three RCTs that compare overall withdrawals in rTMS versus sham in a Tier 1 population (see Table 83). Two report that there are no withdrawals in either the rTMS or sham arms.[64,65] An RCT conducted in 68 patients showed an overall withdrawal rate of 9.1 percent in the rTMS arm and 8.6 percent in the sham arm (*P* = NR).[67,83]

Table 82. Overall withdrawals comparing ECT to rTMS in Tier 3 patients diagnosed with major depressive disorder and bipolar disorder

| Author, year<br>Design<br>Duration<br>Quality | Intervention and sample size<br>Study Details | Withdrawals during treatment |
|---|---|---|
| Eranti et al., 2007,[55]<br>McLoughlin et al., 2007,[54]<br>and Knapp et al., 2008[56]<br>RCT<br>3 weeks<br>Good | **ECT (n = 22)**<br>Twice weekly, stimulus dosing method, 1.5 times the seizure threshold, mean # sessions 6.3 (SD 2.5)<br>**rTMS (n = 24)**<br>High frequency, 110% MT, 1000 pps, 15 sessions, mean # sessions 13.7 (SD 2.7) | Overall, %<br>ECT: 0<br>rTMS: 25<br>*P* = NR |
| O'Connor, 2003[107]<br>Prospective cohort<br>Up to 4 weeks | **ECT (n = 14)**<br>Unilateral, 3 times per week for 2 to 4 weeks.<br>rTMS (n = 14)<br>High frequency, 90% MT,1600 pps, 10 sessions<br>Bipolar (%): NR | Overall: 0 |

ECT, electroconvulsive therapy; MT, motor threshold; n, number; NR, not reported; pps, pulses per second; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation.

Table 83. Overall withdrawals comparing rTMS to sham in Tier 1 patients diagnosed with major depressive disorder

| Author, year<br>Design<br>Quality | Intervention and Sample Size<br>Study Details | Results |
|---|---|---|
| Avery et al. 2006[67,83]<br>RCT<br>4 weeks<br>Good | **rTMS (n = 35)**<br>High frequency<br>110% MT, 1600 pps, 15 sessions<br>Sham (n = 33) | Overall, %:<br>rTMS: 9.1<br>Sham: 8.6<br>*P* = NR |
| Garcia-Toro et al. 2006[65]<br>RCT<br>2 weeks<br>Fair | **rTMS-1 (n = 10)**<br>High frequency, 110% MT, 1200 pps, 10 sessions plus low frequency 110%, 18000 pps, 10 sessions<br>**rTMS-2 (n = 10)**<br>Same as above but with individualized location<br>**Sham: (n = 10)** | Overall: 0 |
| Holtzheimer et al. 2004[64]<br>RCT<br>2 weeks<br>Fair | **rTMS (n = 7 )**<br>High frequency, 110%, 1600 pps, 10 sessions<br>Sham (n = 8) | Overall: 0 |

MT, motor threshold; n, number; NR, not reported; pps, pulses per second; RCT, randomized controlled trials; rTMS, repetitive transcranial magnetic stimulation

*MDD/Bipolar mix.* Four RCTs comprise the MDD/Bipolar mix in a Tier 1 population, as shown in Table 84. One RCT conducted in 40 patients had zero withdrawals in any arm.[74] Another small study (N = 33) had 9.1 percent overall withdrawals in the rTMS and sham groups.[75] Two studies had numerically greater withdrawals in the sham groups, though both were relatively small. The smallest, 21 patients, had overall withdrawals of 8.3 percent in the rTMS group and 30.0 in the sham group.[68] A larger study, 50 patients, had zero percent overall withdrawals in the rTMS group and 12 percent in the sham group.[17]

Table 84. Overall withdrawals comparing rTMS to sham in Tier 1 patients diagnosed with major depressive disorder and bipolar disorder

| Author, year<br>Design<br>Quality | Intervention and Sample Size<br>Study Details | Results |
|---|---|---|
| Boutros, et al., 2002[68]<br>RCT<br>2 weeks<br>Fair | **rTMS (n = 12)**<br>High frequency rTMS, 80% MT, 800 pps, 10 sessions<br>Sham (n = 9) | Overall, %:<br>rTMS: 8.3<br>Sham: 30.0 |
| Fitzgerald et al. 2006[17]<br>RCT<br>6 week<br>Fair | **High and low rTMS (n = 25)**<br>High frequency, 100% MT, 750 pps, 10 sessions plus low requency 110%, 420 pps, 10 sessions<br>Sham (n = 25) | Overall, %:<br>rTMS: 0<br>Sham: 12 |
| Fitzgerald et al. 2003[74]<br>RCT<br>Phase I: 2 weeks<br>Phase II: NA<br>Fair | **rTMS (n = 40)**<br>Low frequency, 100%, 300 pps, 10 sessions or high frequency, 100%, 1000 pps, 10 sessions<br>Sham (n = 20) | Overall: 0<br>Due to AEs: 0 |
| Su et al., 2005[75]<br>RCT<br>2 weeks<br>Fair | **rTMS (n = 22)**<br>High frequency (5Hz), 100% MT, 1600 pps, 10 sessions or high frequency (20Hz) 100% MT, 1600 pps, 10 sessions<br>Sham (n = 11) | Overall, %<br>rTMS 9.1<br>Sham: 9.1 |

AE, adverse events; MT, motor threshold; n, number; NA, not applicable; pps, pulses per second; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation.

**Tier 2.** *MDD only.* There were two RCTs in Tier 2 in MDD only patients, as seen in Table 85. A relatively large, 325 patients, had overall withdrawals of 13.3 percent in the rTMS arm and 16.3 percent in the sham arm.[72,100,101] A small study (n = 45) compared overall withdrawals in four arms, high rTMS (n = 10), low-left rTMS (n = 10), low-right rTMS (n = 10) and sham (n = 15). Two arms had no withdrawals but the low-left rTMS had a 20 percent overall withdrawals rate and 6.7 percent in the sham group.[71]

*MDD/Bipolar mix.* Table 86 provides the two studies that were found in a Tier 2 MDD/bipolar population.[76,87] Overall withdrawals were 0 percent in the rTMS arm and 30 percent in the sham arm. However, no significance was reported.[76] The final study in this group had overall withdrawals of 12.0 percent in the rTMS arm versus 4.2 percent but significance is not reported.[87]

**Tier 3.** There were no studies in MDD only patients and two RCTs in patients with an MDD/bipolar mix diagnosis.

*MDD only.* There were no eligible studies.

*MDD/Bipolar mix.* Two small studies, 19 and 24 patients, comparing rTMS and sham in Tier 3 subjects as seen in Table 87.[77,79] Neither of these studies had any overall withdrawals.

Table 85. Overall withdrawals comparing rTMS to sham in Tier 2 patients diagnosed with major depressive disorder

| Author, year<br>Design<br>Quality | Intervention and Sample Size<br>Study Details | Results |
|---|---|---|
| O'Reardon, 2007[72,100,101]<br>RCT<br>4 weeks primary end pt<br>Fair | **rTMS (n = 165)**<br>High frequency, 120% MT, 3000 pps, 20-30 sessions<br>Sham (n = 160) | Overall, %<br>rTMS 13.3<br>Sham: 16.3 |
| Stern et al., 2007[71]<br>RCT<br>2 weeks of txt<br>Fair | **High rTMS (n = 10)**<br>High frequency, 110%, 1600 pps, 10 sessions<br>Low-left rTMS (n = 10)<br>Low frequency, (Left-DLPFC), 110% MT, 1600 pps, 10 sessions<br>Low-right rTMS (n = 10)<br>Low frequency, 110%, 1600 pps, 10 sessions<br>Sham (n = 15) | Overall:<br>High rTMS: 0<br>Low-left rTMS: 20<br>Low-right rTMS: 0<br>Sham: 6.7 |

DLPFC, dorsolateral prefrontal cortex; MT, motor threshold; n, number; NA, not applicable; pps, pulses per second; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; txt, treatment.

Table 86. Overall withdrawals comparing rTMS to sham in Tier 2 patients diagnosed with major depressive disorder and bipolar disorder

| Author, year<br>Design<br>Quality | Intervention and Sample Size<br>Study Details | Results |
|---|---|---|
| Berman et al, 2000[76]<br>RCT<br>2 weeks<br>Fair | **rTMS (n = 10)**<br>High frequency, 80% MT, 800 pps, 10 sessions<br>Sham (n = 10) | Overall:<br>rTMS: 0<br>Sham: 30 |
| Bretlau et al., 2008[87]<br>RCT<br>3 weeks<br>Fair | **rTMS (n = 25)**<br>High frequency rTMS 100% MT, 1280 pps, 15 sessions over 3 weeks<br>Sham (n = 24)<br>Both groups received 20 mg escitalopram | Overall, %:<br>rTMS: 12.0<br>Sham: 4.2 |

mg, milligram; MT, motor threshold; n, number; pps, pulses per second; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation

Table 87. Overall withdrawals comparing rTMS to sham in Tier 3 patients diagnosed with major depressive disorder

| Author, year<br>Design<br>Quality | Intervention and Sample Size<br>Study Details | Results |
|---|---|---|
| Bortolomasi et al. 2006[77]<br>RCT<br>1 week<br>Fair | **rTMS (n = 12)**<br>High frequency, 90% MT, 800 pps, 5 sessions<br>**Sham (n = 7)** | Overall: 0 |
| George et al. 1997[79]<br>RCT, crossover<br>Primary endpoint after 2 weeks of txt<br>Fair | **rTMS (n = 12)**<br>High frequency, 80% MT, 800 pps, 10 sessions<br>**Sham (n = 12)** | Overall: 0 |

MT, motor threshold; n, number; pps, pulses per second; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; txt, treatment.

## Overall Withdrawals: CBT vs. Usual Care

There were no eligible studies in Tier 1. Tier 2 has one study in patients with MDD only and two studies diagnosed with MDD/bipolar mix and there were no eligible studies in Tier 3.

**Tier 1.** There were no eligible studies.

**Tier 2.** There was one study in patients with MDD only and two studies diagnosed with MDD/bipolar mix.

*MDD only.* Table 88 provides one small study, 32 patients, that was conducted in MDD only tier two patients.[86] Overall withdrawals was 16.7 percent in the CBT arm and 42.9 percent in the usual care arm. Statistical significance was not reported and the CBT arm had 26 subjects compared to six in the usual care arm.

Table 88. Overall withdrawals comparing CBT to sham in Tier 2 patients diagnosed with major depressive disorder

| Author, year<br>Design<br>Quality | Intervention and Sample Size<br>Study Details | Results |
|---|---|---|
| Moore et al. 1997[86]<br>RCT<br>Active phase ocurred during 12 month follow up phase<br>Fair | **CBT (n = 26)**<br>Minimum of 4 treatments in 1st month, 2 treatments in 2nd month and one per month following<br>**Continued or New Medication (n = 6)**<br>Medication dose within recognized therapeutic theshold | Overall, %<br>CBT: 16.7<br>Usual care: 42.9 |

CBT, cognitive behavioral therapy; n, number; RCT, randomized controlled trials.

*MDD/Bipolar mix.* Two studies (4 articles) compared CBT to usual care with mixed results in patients with MDD/bipolar mix in Tier 2 (see Table 89). The smaller one, 24 patients, had an overall withdrawal rate of 23.1 percent in the CBT arm and 18.2 percent in usual care.[81] A larger study, 158 patients, had overall withdrawals of 15.4 percent in the CBT arm versus 23.8 percent in the usual care arm.[59,60,104] For either study, statistical significance was not reported.

Table 89. Overall withdrawals comparing CBT to usual care in Tier 2 patients diagnosed with major depressive disorder and bipolar disorder

| Author, year<br>Design<br>Quality | Intervention and Sample Size<br>Study Details | Results |
|---|---|---|
| Harley, 2008[81]<br>RCT<br>16 weeks<br>Fair | **Dialectical Behavioral Therapy Skills Training (n = 13)**<br>16-session, once weekly group covered the 4 dialectical behavior therapy skill sets<br>Usual care (n = 11) | Overall, %<br>CBT :23.1<br>Usual care: 18.2 |
| Paykel, 1999[59] and Scott, 2000[60] and Scott, 2003[104]<br>RCT<br>20 weeks<br>Good | **Cognitive Therapy plus Clinical Management (n = 80)**<br>16 CT sessions during a 20 week period.<br>Clinical Management (n = 78)<br>Pts visited psychiatrist every 4 weeks and continued on current medication. Seen every 8 wks during followup phase | Overall, %<br>CBT: 15.4<br>Usual Care: 23.8 |

CBT, cognitive behavioral therapy; CT, cognitive therapy; n, number; pts, patients; RCT, randomized controlled trials; wks, weeks.

**Tier 3.**There were no eligible studies.

# Key Question 5: Efficacy and Harms for Selected Populations

## Overview

Studies that focused on subgroups, or included a subanalysis for a special population were eligible for consideration for this key question. Most studies were excluded because the subgroup analysis was not comparative between groups, but rather descriptive within an intervention group. No studies for this key question met the Tier 1 criteria. We identified three small Tier 2 (one or more treatment failures) randomized controlled trials, all on poststroke depression, comparing rTMS to a sham intervention.[108,109]

**Strength of evidence: tier 1 only.** No eligible studies for this tier.

**Strength of evidence: considering all tiers.** We rated the strength of evidence based on the same three outcomes focused on in KQ 1: change in depression severity, response, and remission (Table 90). Although these studies generally suggest that rTMS improves depression severity better than a sham procedure, given the consistency of results and small number of participants, the strength of evidence is low. For response and remission, the results are consistent, but the sample sizes are small, therefore the strength of the evidence is low.

## Key Points

We identified no Tier 1 TRD studies, nor did we identify any head-to-head comparisons. All three Tier 2 trials showed a greater decrease in depressive severity in those receiving rTMS treatment vs. sham. Two of the three trials found statistically significant improvements, but the third trial was underpowered to detect a difference. Response and remission rates were significantly greater in the active group only for the one trial that provided 15 sessions of rTMS over three weeks, in comparison to 10 sessions over two weeks in the other trials.

Table 90. Comparative clinical outcomes in those with two or more failures rTMS vs. sham for Tiers 1-3 combined; number of studies and subjects, and strength of evidence

| Comparison | Number of Studies; Subjects | Risk of bias Design Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| **rTMS vs. sham** | | | | | | |
| Change in depressive severity | 3; 112 | Medium<br>3 RCTs<br>All fair | Consistent | Direct | Imprecise | rTMS significantly better than sham<br>Low |
| Response | 3; 112 | Medium<br>3 RCTs<br>All fair | Consistent | Direct | Imprecise | rTMS better than sham, but not significantly in all studies<br>Low |
| Remission | 3; 112 | Medium<br>3 RCTs<br>All fair | Consistent | Direct | Imprecise | rTMS better than sham, but not significantly in all studies<br>Low |

RCT, randomized controlled trials; rTMS, repetitive transcranial magnetic stimulation; vs., versus.

## Detailed Analysis

Three Tier 2 RCTs focus on patients over the age of 50 with MDD and determined to have vascular depression secondary to a vascular accident.[108,109] As shown in Table 91 below, all three compare high frequency rTMS to a sham intervention and are of fair quality. One small trial was conducted in patients who failed two or more antidepressant treatments, the other two trials were in patients with one or more failures. All three studies were in moderately to severely depressed study populations (mean HAM-$D_{17}$ scores between 17 and 20 in each group) and all were discontinued off of any antidepressants they were receiving. No significant differences were reported for headache, local pain, or anxiety. No seizures occurred in either group.

**Poststroke depression**. Two experiments are presented in one article.[108] The active intervention in the first study applied 10 sessions of rTMS to 15 patients (15 in the sham group). In a modified–ITT treatment after three weeks of treatment, the rTMS group had a greater percent decrease in HAM-$D_{17}$ (33.1% vs. 13.6%, $P = 0.04$) tended to have a greater response rate, but the difference was not significant. Remission rates in each group were low, but also not significant. The second study increased the number of sessions to 15 and showed a greater decrease in depressive severity in the rTMS group with significantly improved response and remission rates after three weeks of treatment. In this experiment, 33 patients received 15 sessions (29 patients in sham group) and resulted in a greater percent decrease in HAM-$D_{17}$ (42.4% vs. 17.5%, $P = 0.01$), response rate (39.4% vs. 6.9%, $P = 0.003$) and remission (27.3% vs. 3.4%, $P = 0.01$) in comparison to the sham intervention group.

Table 91. Studies of efficacy and harms for patient subgroups

| Author, year<br>Study Design<br>Primary Endpoint(s)<br>Quality | Intervention and Sample Size<br>Study Details | Population Characteristics | Response<br>Remission<br>Change in Depressive Symptoms | Adverse Events<br>Quality of Life<br>Attrition |
|---|---|---|---|---|
| **Poststroke Depression** | | | | |
| Jorge et al. 2008[108]<br>Experiment 1<br>RCT, primary endpoint at 3 weeks, mITT<br>Failure required in current episode.<br>Fair | High frequency rTMS (n = 15)<br>10 sessions<br>Sham (n = 15)<br>**Concurrent medications**<br>All antidepressants discontinued<br>Strategy<br>Switch<br>Definitions<br>Remission: HAM-$D_{17}$ < 8 and did not meet criteria for major or minor depression | Subgroup<br>Patients with stroke/cerebral vascular disease<br>Diagnosis,%<br>MDD: 100<br>Baseline Depression: HAM-$D_{17}$<br>rTMS: 19.5 (5.8)<br>Sham: 19.9 (5.4) | HAM-$D_{17}$<br>Response, n (%)<br>rTMS: 5 (33.3)<br>Sham: 1 (6.7)<br>*P* = 0.08<br>Remission, n (%)<br>rTMS: 2 (13.3)<br>Sham: 1 (6.7)<br>*P* = 0.5<br>Change,%<br>rTMS: 33.1<br>Sham: 13.6<br>*P* = 0.04 | Adverse Events<br>Headache,%<br>rTMS: 5 (33)<br>Sham: 4 (27)<br>*P* = NR<br>No differences in frequency of headaches; all headaches were mild and responded to low dose analgesics<br>Local Pain, n (%)<br>rTMS: 1 (7)<br>Sham: 1 (7)<br>*P* = NR<br>Local discomfort, n (%)<br>rTMS: 4 (27)<br>Sham: 5 (33)<br>No difference in frequency of local discomfort<br>*P* = NR<br>Anxiety, n (%)<br>rTMS: 2 (13)<br>Sham: 0 (0)<br>*P* = NR<br>Seizures, n<br>rTMS: 0<br>Sham: 0<br>*P* = NR |
| Jorge et al. 2008[108]<br>Experiment 2<br>RCT, primary endpoint at 3 weeks, mITT<br>Failure not required in current episode.<br>Fair | **High frequency rTMS** (n = 33)<br>15 sessions<br>**Sham** (n = 29)<br>**Concurrent medications**<br>All antidepressants discontinued<br>**Strategy**<br>Switch<br>**Definitions**<br>Remission: HAM-$D_{17}$ <8 and did not meet criteria for major or minor depression | Subgroup<br>Patients with stroke/cerebral vascular disease<br>Diagnosis,%<br>MDD: 100<br>Baseline Depression, n (%):<br>HAM-$D_{17}$<br>rTMS: 18.4 (3.4)<br>Sham: 17.6 (5.6) | HAM-$D_{17}$<br>Response, n (%)<br>rTMS: 13 (39.4)<br>Sham: 2 (6.9)<br>*P* = 0.003<br>Remission<br>rTMS: 9 (27.3)<br>Sham: 1 (3.4)<br>*P* = 0.01<br>Change,%<br>rTMS: 42.4<br>Sham: 17.5<br>*P* < 0.01 | Adverse Events<br>Headache,%<br>rTMS: 7 (21)<br>Sham: 3 (10)<br>No differences between groups in frequency of headaches; all headaches were mild and responded to low dose analgesics<br>*P* = NR<br>Local Pain, n (%)<br>rTMS: 1 (3)<br>Sham: 0 (0)<br>*P* = NR<br>Local discomfort, n (%)<br>rTMS: 3 (9)<br>Sham: 1 (3)<br>No difference in frequency of local discomfort<br>*P* = NR |

HAM-$D_{17}$, 17-item Hamilton MDD, major depressive disorder; ITT, intent-to-treat analysis; mITT, modified intent-to-treat analysis; NR, not reported; NS, not significant; *P,* p-value; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; txt, treatment.

Table 91. Studies of efficacy and harms for patient subgroups (continued)

| Author, year<br>Study Design<br>Primary Endpoint(s)<br>Quality | Intervention and Sample Size<br>Study Details | Population Characteristics | Response<br>Remission<br>Change in Depressive Symptoms | Adverse Events<br>Quality of Life<br>Attrition |
|---|---|---|---|---|
| **Poststroke Depression** | | | | |
| Jorge et al. 2008[108]<br>Experiment 2<br>(continued) | | | | Anxiety, n (%)<br>rTMS: 1 (0)<br>Sham: 0 (0)<br>*P* = NR<br>Seizures, n<br>rTMS: 0<br>Sham: 0<br>*P* = NR |
| Jorge et al. 2004[109]<br>RCT, primary outcome at 3 weeks (2 weeks of txt, 1 week followup), ITT<br>Failure in current episode not required<br>Fair | **High Frequency rTMS** (n = 10)<br>10 sessions<br>**Sham** (n = 10)<br>**Concurrent medications**<br>All antidepressant medications discontinued<br>**Strategy**<br>Switch | Subgroup<br>Patients with stroke/cerebral vascular disease<br>Diagnosis,%<br>MDD: 100<br>Baseline Depression: HAM-$D_{17}$<br>rTMS: 20.1 (6.7)<br>Sham: 20.8 (6.0) | HAM-$D_{17}$<br>Response, n (%)<br>rTMS: 3 (30)<br>Sham: 0 (0)<br>*P* = NS<br>Remission, n (%)<br>rTMS: 1 (10)<br>Sham: 0 (0)<br>*P* = NS<br>Change Score<br>rTMS: 7.3<br>Sham: NR (can be calculated as 2.7)<br>*P* < 0.006<br>Change,%<br>rTMS: 38<br>Sham: 13 | Adverse Events<br>No significant differences in frequency of adverse events between active and sham rTMS groups<br>Neither group reported seizures or propagation of cortical excitability toipsilateral motor cortex |

In the third trial, 10 patients were treated with rTMS over 10 sessions (10 in sham group) and showed a greater decrease in depressive severity, but did not have the power to adequately compare response and remission rates.[109] Mean baseline depressive severity was moderate, with both groups averaging between 20 and 21 points on the HAM-$D_{17}$. Antidepressants were tapered to discontinuation prior to enrollment, so patients were switched to rTMS or control. An ITT analysis at three weeks, found outcomes favored the rTMS group. Compared to control, rTMS produced a greater decrease in depressive severity (- 7.3 vs.—2.7, $P < 0.006$), a greater likelihood of both response (3 out of 10 vs. 0 out of 10) and remission (1 of 10 vs. 0 of 10).

# Key Question 6: Health-related Outcomes—Overview

Understanding the burden of affective disorders on the quality of life of patients is an important component to establishing the overall effectiveness of treatment for these disorders. However, quality of life is rarely assessed in this body of literature. Previous ECT studies have associated ECT with a post-treatment quality-of-life improvement that can be maintained from 1 month to 1 year.[54] Very little quality-of-life data following rTMS, VNS, behavioral, or other nonpharmacologic treatments are available.

Numerous psychometric measures exist to assess an individual's level of functioning and execution of daily living activities, both health domains that are related to quality of life. The Global Assessment of Functioning (GAF) and the Range of Impaired Functioning Tool (LIFE-RIFT) are scales used to determine patients' ability to function in daily life.[110,111] The Medical Outcomes Study 36 Item Short Form (MOS SF-36 or SF-36) is an internationally recognized generic health survey instrument comprised of 36 items in eight independent health domains used to survey the health status of an individual.[112] Quality of life enjoyment and satisfaction questionnaire (Q-LES-Q) is a 16-item Quality of Life questionnaire (Q-LES-Q) that uses a self-report measure to obtain the degree of enjoyment and satisfaction of various areas of daily functioning.[113,114] Finally, the Social Adjustment Scale-Self-Report (SAS-SR) work subscale taps a subset of daily activities that may indirectly reflect patients' quality of life.[115]

The following KQ focuses on the comparative benefit of patient-reported health related outcomes using quality of life measures with TRD (MDD/Bipolar and MDD only). There were no head-to-head (direct) comparisons identified. Four indirect comparison studies were available and assessed general health status and mental and physical functioning. Two studies compared rTMS vs. sham, one study compared VNS vs. sham and one study compared CBT vs. control.

For TRD populations (Tier 1), we identified two studies, both in MDD/Bipolar mix samples (Table 92), one comparing rTMS vs. sham[74] and one comparing VNS vs. sham.[80] Both studies suggested greater benefit for rTMS over the control.

Considering additional tiers added, two Tier 2 studies of MDD only populations comparing rTMS vs. sham[72,100,101] and CBT vs. control group[81] showing no difference in outcomes (Table 92).

Table 92. Number of good and fair quality studies by TRD tier and diagnostic mix

| Comparison and Population Included | Diagnosis | |
|---|---|---|
| | MDD only | MDD and Bipolar Disorder |
| **Tier 1: ≥ 2 failures** | | |
| rTMS vs. Sham | 0 | 1 |
| VNS vs. Sham | 0 | 1 |
| Tier 2: ≥ 1 failure | | |
| rTMS vs. Sham | 1 | 0 |
| CBT vs. Control | 1 | 0 |

## Strength of Evidence: Tier 1 Only

No evidence directly compared the effect of nonpharmacologic treatment on patient-reported health-related outcomes. Two studies provided indirect evidence. Neither of these two Tier 1 studies assessed quality of life for a nonpharmacologic intervention vs. control, instead assessing general health status and mental and physical functioning, related health domains, for a nonpharmacologic treatment vs. a sham comparison. One study provided insufficient strength of evidence to assess whether there was a greater improvement in the ability to function following treatment with rTMS compared to sham, as results were mixed (Table 93).[74] Results were in the same direction favoring rTMS, but one of the active arms (low-right rTMS) produced statistically greater improvement than sham, while the second active arm (high-left rTMS) produced greater improvement that did not reach statistical significance. The other study provided low strength of evidence that health status did not differ significantly following treatment with VNS or sham.[80]

Table 93. Health related outcome measures by comparison for TRD (Tier 1); number of studies and subjects, and strength of the evidence domains

| | Number of studies; subjects | Risk of bias Design Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| High-left rTMS vs. Sham and Low-right rTMS vs. Sham | 1; 20 | Medium<br>1 RCT<br>1 Fair | NA | Direct | Imprecise | High-left rTMS, produces a greater improvement in health status and daily functioning than sham (*P* = 0.09), and low-right rTMS produces a greater improvement in health status and daily functioning than sham (*P* = 0.03)<br>Insufficient |
| VNS vs. sham | 1; 112 | Medium<br>1 RCT<br>1 Fair | NA | Direct | Imprecise | There is no difference between VNS vs. sham in daily functioning<br>Low |

NA, not applicable; RCT, randomized controlled trial; rTMS: transcranial magnetic stimulation; VNS, vagus nerve stimulation; vs., versus.

## Strength of Evidence: Considering all Tiers

Considering additional evidence from Tiers 2 and 3 added two studies (Table 94), one to the existing comparison of rTMS vs. sham[81] and another to a new comparison of CBT vs. control study.[72,100,101] The additional rTMS vs. sham comparison[72,100,101] provided more data showing significantly greater improvement with rTMS vs. sham and changed the strength of evidence from insufficient to a low strength of evidence that rTMS produces greater improvement in health status and functioning compared to sham. All results were in the same direction. Finally, the new CBT vs. control comparison provided low strength of evidence that functional impairment did not vastly differ following psychotherapy compared to a wait list control group.

Table 94. Health related outcome measures by comparison for Tiers 1-3 combined; number of studies and subjects, and strength of the evidence domains

| | Number of studies; subjects | Risk of bias Design Quality | Consistency | Directness | Precision | Results and Strength of Evidence |
|---|---|---|---|---|---|---|
| rTMS vs. Sham | 2; 175 | Medium<br>2 RCTs<br>1 Good, 1 Fair | Consistent | Direct | Imprecise | rTMS produces greater improvement in health status and daily functioning compared to sham<br>Low |
| VNS vs. sham | 1; 112 | Medium<br>1 RCT<br>Fair | N/A | Direct | Imprecise | No significant difference between VNS and sham in daily functioning<br>Low |
| CBT/DBT vs. control | 1; 10 | Medium<br>1 RCT<br>1 Fair | N/A | Direct | Imprecise | CBT produces nonsignificantly greater improvement in daily functioning compared to waitlist control<br>Low |

CBT, cognitive behavioral therapy; DBT, dialectical behavior therapy; RCT, randomized controlled trial; rTMS: transcranial magnetic stimulation ; VNS, vagus nerve stimulation; vs., versus.

# Key Question 6: Health-Related Outcomes-Key Points

No studies directly compared the effect of nonpharmacologic treatments on health-related outcomes.

We identified four relevant studies that compared outcomes related to quality of life for patients who underwent rTMS, VNS, vs. sham or CBT compared to a control group. Two studies[72,81,100,101] involved patients with MDD only and the other two studies[74,80] involved patients with MDD and/or bipolar disorder. The studies were funded by the United States federal government, hospitals, and universities. The active treatment duration across studies ranged from 2 weeks to 16 weeks.

Overall, the study samples were relatively small; two of the four studies had study samples of 50 or fewer patients. All studies were RCTs and we rated three as fair and one as good quality. One study found significant differences in GAF between one active arm and sham, but not between the other active arm and sham.[74] Additionally, two studies reported significant changes ($P < 0.05$) in the Social Adjustment Scale-Self-Report (SAS-SR) work subscale and the SF-36 Mental Component Score and the Q-LES-Q Total Score, respectively.[72,81,100,101]

# Key Question 6: Health-Related Outcomes—Detailed Analysis

## Repetitive Transcranial Magnetic Stimulation vs. Sham Control

**Tier 1: Patients with two or more failures.** *MDD only.* There were no eligible studies.

*MDD/Bipolar.* One study compared rTMS treatment (two versions—LFR-rTMS and HFL-rTMS) to a sham procedure and found no significant differences between the active rTMS groups compared with the sham group in the GAF mean score change (Table 95).[74] However,

they found a significant difference in the GAF mean score change score between the LFR-rTMS vs. sham groups ($P = 0.03$).

Table 95. Tier 1 studies including depressed populations with ≤20 percent of patients with bipolar disorder

| **Author, Year**<br>**Endpoint**<br>**Episode Failure**<br>**Quality**<br>**Tier** | **Intervention and Sample Size**<br>**Study Details** | **Results** |
|---|---|---|
| Fitzgerald et al., 2003[74]<br>2 weeks, all reported patients included<br>Did not require failure in the current episode<br>Fair<br>Tier1 | HFL-rTMS (n = 20)<br>High frequency, 10 sessions<br>LFR-rTMS (n = 20)<br>Low frequency, 10 sessions<br>**Sham (n = 20)**<br>Coil angled 45 degrees offhead, randomized left or right | Global Assessment of Functioning (GAF)<br>Baseline score, mean (SD)<br>G1: 43.0 (6.8)<br>G2: 43.5 (9.9)<br>G3: 42.7 (7.1)<br>**Endpoint score, mean** (SD)<br>At week 2<br>G1: 45.2 (7.1)<br>G2: 46.3 (8.5)<br>G3: 42.5 (6.8)<br>**Change, mean**<br>At week 2<br>G1: 2.2<br>G2: 1.4<br>G3: 0.2<br>G1 vs. G3:$P = 0.09$<br>G2 vs. G3 $P = 0.03$ |
| Rush et al., 2005[80]<br>10 weeks, all reported patients included<br>Two-six failures in current episode.<br>Fair<br>Tier 1 | VNS (n = 112)<br>10 weeks of VNS therapy with continued medications<br>Sham (n = 110)<br>Sham: device implanted but not turned on | Medical Outcomes Study Short Form-36 (MOS SF-36)<br>Baseline n<br>G1: n = 107<br>G2: n = 107<br>Change, mean (SD)<br>Physical component<br>G1: -0.9 (8.3)<br>G2: -1.6 (8.4)<br>$P = 0.480$<br>Mental component<br>G1: 5.0 (11.6)<br>G2: 4.0 (10.2)<br>$P = 0.406$. |

G, group; GAF, Global Assessment of Function; HFL-rTMS, high-frequency left-sided rTMS; Hz, hertz; LFR-rTMS, low-frequency right sided rTMS; MOS, SF-36; Medical Outcomes Study Short Form-36; n, number; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation; sec, second; VNS, vagus nerve stimulation.

**Tier 2: Patients with one or more failures.** *MDD only.* One study compared rTMS to sham procedure and found significant differences between the two groups in both the SF-36 mental component score ($P = 0.032$) and the Q-LES_Q total score ($P = 0.035$) (Table 96).[72,100,101]

*MDD/Bipolar.* There were no eligible studies.

**Tier 3:** No Tier 3 data was available for either the MDD only or MDD/Bipolar populations.

## Vagus Nerve Stimulation vs. Sham Control

**Tier 1: Patients with two or more failures.** *MDD only.* There were no eligible studies.

Table 96. Tier 2 studies on patients with major depressive disorder

| Author, Year<br>Endpoint<br>Episode Failure<br>Quality<br>Tier | Intervention and Sample Size<br>Study Details | Results |
|---|---|---|
| Harley et al., 2008[81]<br>16 weeks, all reported patients included<br><br>Did not require failure in the current episode<br>Fair<br>Tier 2 | DBT: Psychotherapy, a form of Cognitive Behavioral Therapy (CBT)<br>**(n = 10)**<br>16 weekly sessions lasting one hour and 30 minutes each,<br><br>Control (n = 9) | Lifework-The Range of Impaired Functioning Tool (Lifework-RIFT)<br>G1: CBT<br>G2: Control<br><br>Baseline score, mean (SD)<br>G1: 4.00 (0.94)<br>G2: 3.44 (1.24)<br>**Endpoint score, mean (SD)**<br>G1: 2.70 (1.34)<br>G2: 3.11 (1.69)<br>**Change, mean (SD)**<br>G1: -1.3<br>G2: -0.33<br>*P* = NS<br><br>Social Adjustment Scale-Self-Report (SAS-Self Report) work subscale<br>**Baseline score, mean (SD)**<br>G1: 82.50 (21.21)<br>G2: 69.22 (17.95)<br>**Endpoint score, mean (SD)**<br>G1: 65.70 (19.27)<br>G2: 69.56 (17.66)<br>**Change, mean (SD)**<br>G1: -16.80<br>G2: 0.34<br>*P* < 0.05 |
| O'Reardon et al., 2007,[72] Janicak et al,2007,[100] and Solvason et al, 2007[101]<br>6 weeks, all reported patients included.<br>Required to have failed at least one in this or most recent episode or four failed attempts in a lifetime.<br>Good<br>Tier 2 | **rTMS (n=155)**<br>High frequency, up to 30 sessions<br><br>G2: Sham rTMS (n=146)<br>Coil has embedded magnetic shield, limiting magnetic energy reaching cortex to 10% or less than active coil | Medical Outcomes Study Short Form-36 (MOS SF-36)<br>Baseline score, mean (SD)<br>Mental Component Score<br>G1: 20.4 (8.05)<br>G2: 20.4 (7.76)<br>Physical Component Score<br>G1: 50.5 (11.01)<br>G2: 48.8 (10.35)<br><br>Change, mean (SD)<br>Mental Component Score<br>At week 4<br>G1: 4.5 (10.16)<br>G2: 2.0 (9.42)<br>*P* = 0.019 |

Lifework-RIFT , Lifework-The Range of Impaired Functioning Tool; SAS-Self Report, Social Adjustment Scale-Self-Report work subscale; Q-LES Questionnaire, Quality of Life, Enjoyment and Satisfaction Questionnaire –Short Form; CBT, cognitive behavioral therapy; DBT, dialectical behavior therapy; G, group; HFL-rTMS, high-frequency left-sided rTMS; LFR-rTMS, low-frequency right sided rTMS; MDD, major depressive disorder; QOL, quality of life; rTMS, repetitive transcranial magnetic stimulation; SD, standard deviation; VNS, vagus nerve stimulation.

Table 96. Tier 2 studies on patients with major depressive disorder (continued)

| Author, Year<br>Endpoint<br>Episode Failure<br>Quality<br>Tier | Intervention and Sample Size<br>Study Details | Results |
|---|---|---|
| | | At week 6<br>G1: 5.7 (12.65)<br>G2: 2.9 (10.6)<br>*P* = 0.032<br>Physical Component<br>At week 4<br>G1: 0.3 (7.52)<br>G2: 0.2 (7.28)<br>*P* = 0.892<br>At week 6<br>G1: 0.1 (7.49)<br>G2: -0.2 (7.23)<br>*P* = 0.682<br>Quality of Life, Enjoyment and Satisfaction Questionnaire –Short Form (Q-LES Questionnaire)<br>Baseline score, mean (SD)<br>G1: 37.8 (8.23)<br>G2: 36.5 (7.87)<br>Endpoint score, mean (SD)<br>At week 4<br>G1: 41.4 (10.32)<br>G2: 39.0 (9.78)<br>At week 6<br>G1: 42.2 (12.28)<br>G2: 39.0 (10.15)<br>Change, mean (SD)<br>At week 4<br>G1: 3.50 (9.19)<br>G2: 3.80 (11.58)<br>At week 6<br>G1: 2.0 (9.24)<br>G2: 1.3 (9.85)<br>*P* = 0.035 |

*MDD/Bipolar.* One study compared VNS and a sham procedure using the MOS SF-36 to assess quality of life.[80]The intervention and control groups did not differ significantly on either the mental or physical components of the MOS SF-36 instrument.

**Tier 3:** No Tier 3 data was available for either the MDD only or MDD/Bipolar populations.

## Cognitive Behavioral Therapy vs. Control

**Tier 1:** No Tier 1 data was available for either the MDD only or MDD/Bipolar populations.

**Tier 2: Patients with one or more failures.** *MDD only.* The Harley et al. study, rated fair quality compared patients receiving psychotherapy such as CBT or IPT with a control group using the LIFE-RIFT instrument.[81] They found no significant differences between the intervention and control groups. They also used the SAS-SR work subscale as a measure of

QOL, reporting a significant difference ($P < 0.05$) between the psychotherapy group compared to the control group.

*MDD/Bipolar.* There were no eligible studies.

**Tier 3:** No Tier 3 data was available for either the MDD only or MDD/Bipolar populations.

# Chapter 4. Discussion

## Background

This review from the RTI-UNC Evidence-based Practice Center (EPC) provides a comprehensive summary of the available data addressing the comparative effectiveness of four nonpharmacologic treatments— electroconvulsive therapy (ECT), repetitive transcranial magnetic stimulation (rTMS), vagus nerve stimulation (VNS), and cognitive behavioral therapy or interpersonal psychotherapy (CBT or IPT)—as therapies for patients with treatment-resistant depression (TRD). For one issue (see key questions [KQ s] below), we also examined pharmacologic (antidepressant) interventions. The core patient population of interest was patients with major depressive disorder (MDD) who met our definition of TRD: failure to respond following two or more adequate antidepressant trials. We also included studies in which the patient population could include a "mix" of up to 20 percent of patients with bipolar disorder (i.e., 80 percent or more of patients had only MDD), assuming that this small mix would not substantially alter outcomes seen with MDD alone populations. In addition, we distinguished between patients for whom treatment was directed at the acute phase of disease and those for whom treatment was intended to maintain remission or to prevent relapse.

We structured our review to focus chiefly on our primary population of interest (MDD patients with TRD) but also considered data from studies that likely had a substantial proportion of TRD patients. We worked with our Technical Expert Panel to identify different tiers of definitions for TRD to use in our analytic strategy:

- **Tier 1** evidence (TRD as defined in this report): studies in which patients specifically had two or more failures of prior treatment with medications;
- **Tier 2** evidence: studies in which patients had one or more failures of prior treatment;
- **Tier 3** evidence: studies in which the number of prior failed treatments was not specified but the clinical situation suggested a high probability of patients having failed two or more prior antidepressant trials; this data has probable relevance to TRD. Studies which did not specify the number of failed treatments but noted that all subjects were referred for ECT were included in this tier.

The focus of each of the six KQ s or subquestions is listed below (key distinguished elements in italics).

- KQ 1a. Efficacy of nonpharmacologic interventions for *acute-phase* TRD (depressive severity, response, or remission).
- KQ 1b. Efficacy of nonpharmacologic vs. pharmacologic interventions for *acute-phase* TRD (depressive severity, response, or remission), for patients failing two or more prior treatments.
- KQ 2. Efficacy of nonpharmacologic interventions for *maintaining response or remission* with respect to TRD (e.g., preventing relapse or recurrence).
- KQ 3. Efficacy of nonpharmacologic interventions for *acute-phase* TRD as a function of particular *symptom subtypes* (e.g., catatonia or psychosis).

- KQ 4. Harms of nonpharmacologic interventions (i.e., safety, adverse events, or adherence issues).
- KQ 5. Efficacy or harms of nonpharmacologic treatments for selected patient subgroups defined by sociodemographic characteristics or coexisting conditions.
- KQ 6. Health-related outcomes of nonpharmacologic treatments (e.g., quality of life).

In the discussion below, we comment on findings from direct and indirect evidence; where differences were clinically meaningful, we provide the data also reported in Chapter 3. Respectively, these terms refer to head-to-head studies or studies involving a control group of some sort, such as a sham procedure or usual care (treatment as usual). As with Chapter 3, we include only studies for which we rated the quality as either good or fair; most studies were of only fair quality.

Finally, we graded the strength of evidence for major outcomes and comparisons. Detailed information was presented in Chapter 3; we comment in text below about the strength of evidence for main findings. Given our focus on TRD defined by two or more treatment failures, we first present findings from Tier 1 studies. Subsequently, we identify how consideration of Tiers 2 and 3 affects interpretation of the available evidence. To recap, the four levels of strength of evidence are as follows:

- **High: High confidence that the evidence reflects the true effect.** Further research is very unlikely to change our confidence in the estimate of effect.
- **Moderate: Moderate confidence that the evidence reflects the true effect.** Further research may change our confidence in the estimate of effect and may change the estimate.
- **Low: Low confidence that the evidence reflects the true effect.** Further research is likely to change our confidence in the estimate of effect and is likely to change the estimate.
- **Insufficient:** Evidence either is unavailable or does not permit estimation of an effect.

## Overview of Main Findings

A summary of our main findings are found in Table 97. A more detailed description follows.

**Table 97. Summary of findings on nonpharmacologic treatment of adult major depressive disorder with strength of evidence**

| Key Question and Comparison | Outcome | Tier 1 only | | Tiers 1- 3 | |
|---|---|---|---|---|---|
| | | Strength of Evidence[1] | Findings[2] | Strength of Evidence[1] | Findings[2] |
| **Key Question 1a. Comparative efficacy of nonpharmacologic treatments** | | | | | |
| ECT vs. rTMS | Change in depressive severity | Low | 1 fair trial: both ECT and rTMS improved symptom severity but did not differ significantly. | Insufficient | 1 good trial: ECT produced a significantly greater decrease in symptom severity than rTMS<br>2 fair trials: both ECT and rTMS improved symptom severity but did not differ significantly. |
| | Response rate | Low | 1 fair trial: ECT and rTMS did not differ significantly | Insufficient | 1 good trial: ECT produced significantly greater response rates than rTMS<br>2 fair trials: ECT and rTMS did not differ significantly. |
| | Remission rate | Low | 1 fair trial: ECT and rTMS did not differ significantly | Insufficient | 1 good trial: ECT produced significantly greater remission rates than rTMS.<br>2 fair trials: ECT and rTMS did not differ significantly. |
| ECT plus rTMS vs. ECT | Change in depressive severity | Low | 1 fair trial: Both ECT and ECT plus rTMS improved symptom severity but did not differ significantly. | Low | 2 fair trials: Both ECT and ECT plus rTMS improved symptom severity but did not differ significantly. |
| | Response rate | NA | No eligible studies identified | Low | 1 fair trial: ECT and ECT plus rTMS did not differ significantly. |
| | Remission rate | Low | 1 fair trial: ECT and ECT plus rTMS did not differ significantly. | Low | 1 fair trial: ECT and ECT plus rTMS did not differ significantly. |
| rTMS vs. sham | Change in depressive severity | High | 7 trials (3 good, 4 fair) rTMS had a significantly greater decrease in depressive severity than sham.<br>4 fair trials: rTMS had nonsignificantly greater decrease in depressive severity than sham.<br>1 fair trial: rTMS had greater decrease than sham but significance NR. | High | 13 trials (4 good, 9 fair): rTMS had a significantly greater decrease in depressive severity than sham.<br>5 fair trials: rTMS had a nonsignificantly greater decrease in depressive severity than sham<br>1 fair trial: rTMS had greater decrease than sham but significant NR. |

AE, adverse event; CBT, cognitive behavioral therapy; DBT, dialectical behavioral therapy; ECT ,electroconvulsive therapy; KQ , key question; NA, not applicable; NR, not reported; rTMS, repetitive transcranial magnetic stimulation; VNS, vagus nerve stimulation.

Table 97. Summary of findings on nonpharmacologic treatment of adult major depressive disorder with strength of evidence (continued)

| Key Question and Comparison | Outcome | Tier 1 only | | Tiers 1- 3 | |
|---|---|---|---|---|---|
| | | Strength of Evidence[1] | Findings[2] | Strength of Evidence[1] | Findings[2] |
| **Key Question 1a. Comparative efficacy of nonpharmacologic treatments** | | | | | |
| | Response rate | High | 4 trials (3 good, 1 fair): rTMS had a significantly higher response rate than sham.<br>1 fair trial: rTMS had a nonsignificantly higher response rate than sham.<br>4 fair trials: rTMS had a higher response rate than sham, but significance NR. | High | 5 trials (4 good, 1 fair): rTMS had a significantly higher response rate than sham.<br>2 trials (1 good, 1 fair): rTMS had a nonsignificantly higher response rate than sham.<br>5 fair trials: rTMS had a higher response rate than sham, but significance NR.<br>1 fair trial: rTMS and sham had equal response rates. |
| | Remission rate | Moderate | 2 good trials: rTMS had significantly greater remission rate than sham.<br>1 fair trial: rTMS had a greater remission rate than sham but significance NR. | High | 3 good trials: rTMS had a significantly greater remission rate than sham.<br>2 fair trials: rTMS had a greater remission rate than sham but significance NR.<br>1 fair trial: rTMS and sham did not differ significantly.. |
| VNS vs. sham | Change in depressive severity | Low | 1 good trial: VNS and sham did not differ significantly. | Low | No additional studies identified. |
| | Response rate | Low | 1 good trial: VNS and sham did not differ significantly. | Low | No additional studies identified. |
| Psychotherapy vs. control | Change in depressive severity | NA | No eligible studies identified. | Insufficient | 2 fair trials: psychotherapy group had a significantly greater decrease in depressive severity than control group.<br>1 good trial: psychotherapy and control groups did not differ significantly. |
| | Response rate | NA | No eligible studies identified. | Low | 1 fair trial: psychotherapy group had greater response rate than control group (significance NR) |
| | Remission rate | NA | No eligible studies identified. | Low | 1 good trial showed a significantly higher remission rate for psychotherapy than control<br>1 fair trial favored psychotherapy but significance NR. |

Table 97. Summary of findings on nonpharmacologic treatment of adult major depressive disorder with strength of evidence (continued)

| Key Question and Comparison | Outcome | Tier 1 only | | Tiers 1- 3 | |
|---|---|---|---|---|---|
| | | Strength of Evidence[1] | Findings[2] | Strength of Evidence[1] | Findings[2] |
| **KQ 1b. Comparative efficacy of nonpharmacologic and pharmacologic treatments** | | | | | |
| ECT vs. pharmaco-therapy | Change in depressive severity | Low | 1 fair trial: ECT had significantly greater improvement in symptom severity than pharmacotherapy | Low | No additional studies identified. |
| | Response rate | Low | 1 fair trial: ECT had significantly greater response rates than pharmacotherapy | Low | No additional studies identified. |
| Psycho-therapy vs. pharmaco-therapy | Change in depressive severity | NA | No eligible studies identified. | Low | 1 fair trial using both switching and augmenting strategies found no significant differences in change in depressive severity between CBT and pharmacotherapy.<br>1 small fair trial using a mixed strategy and conducting a treatment completer analysis did not report statistical significance |
| | Response rate | NA | No eligible studies identified. | Low | 1 good trial: CBT and pharmacotherapy did not differ significantly in response rate. |
| | Remission rate | NA | No eligible studies identified. | Low | 1 good trial: CBT and pharmacotherapy did not differ significantly in remission rate. |
| rTMS + pharmaco-therapy vs. pharmaco-therapy | Change in depressive severity | NA | No eligible studies identified. | Low | 1 fair trial: rTMS patients starting a new antidepressant had a significantly greater decrease in depressive severity than sham procedure patients and pharmacotherapy (as measured by effect size). |
| **KQ 2. Comparative efficacy for maintaining remission** | | | | | |
| ECT vs. rTMS | Maintenance of remission | NA | No eligible studies identified | Low | 2 fair trials: ECT and rTMS did not differ significantly |
| rTMS vs. sham | Maintenance of remission | Insufficient | 2 fair trials: relapse rates did not significantly differ between rTMS and sham, but too few patients were followed during the continuation phase to allow for meaningful conclusions. | Low | 6 fair trials: rTMS may maintain remission better than sham (depending on length of followup); significance NR for 4 trials. |
| CBT vs. usual care | Maintenance of remission | NA | No eligible studies identified. | Low | 1 fair trial: CBT had lower risk of relapse than usual care at all followup points but significance of results varied by how relapse was defined and was not reported for all follow-up measures. |

Table 97. Summary of findings on nonpharmacologic treatment of adult major depressive disorder with strength of evidence (continued)

| Key Question and Comparison | Outcome | Tier 1 only | | Tiers 1- 3 | |
|---|---|---|---|---|---|
| | | Strength of Evidence[1] | Findings[2] | Strength of Evidence[1] | Findings[2] |
| **KQ 3. Comparative efficacy for particular symptom subtypes** | | | | | |
| ECT vs. rTMS | Change in depressive severity | NA | No eligible studies identified. | Insufficient | 1 fair trial: Mixed results. ECT was significantly more effective than rTMS in the overall population and in psychotic patients, but differential use of antipsychotic with ECT (and not rTMS) substantially biased the results; ECT and rTMS were equally effective in nonpsychotic patients. |
| **KQ 4a. Impact of nonpharmacologic interventions on cognitive functioning** | | | | | |
| ECT vs. rTMS | Cognitive functioning | Low | 1 fair trial and 1 fair cohort study: ECT may have deleterious impact on cognitive functioning compared with rTMS (1 study: significant effect on 1-week recall; both studies: nonsignificant effect on all other measures) | Low | 2 trials (1 good, 1 fair) and 2 fair cohort studies: ECT may have deleterious impact on cognitive functioning compared with rTMS (1 study: significant differences in attention and orientation; 1 study: significant difference in acquisition; all other findings: differences nonsignificant or significance NR) |
| rTMS vs. sham | Cognitive functioning | Low | 4 trials (1 good, 3 fair): rTMS may have positive impact on cognitive functioning (2 trials: significant differences in memory, verbal fluency; all other findings nonsignificant or significance not reported). | Low | 5 trials (1 good, 4 fair): No change in findings for Tier 1 alone. |
| **KQ 4b. Specific adverse events** | | | | | |
| ECT vs. rTMS | Adverse events | NA | No eligible studies identified. | Low | 1 good trial: ECT and rTMS did not differ significantly in specific adverse events. |
| rTMS vs. sham | Adverse events | Low | 1 good trial: rTMS resulted in significantly more scalp pain at the stimulation site than sham. | Low | 3 trials (1 good, 2 fair): rTMS had significantly more scalp pain and difficulty starting urination than sham. |
| VNS vs. sham | Adverse events | Low | 1 fair trial: VNS and sham did not differ significantly in specific adverse events | Low | No additional studies identified. |

Table 97. Summary of findings on nonpharmacologic treatment of adult major depressive disorder with strength of evidence (continued)

| Key Question and Comparison | Outcome | Tier 1 only: Strength of Evidence[1] | Tier 1 only: Findings[2] | Tiers 1- 3: Strength of Evidence[1] | Tiers 1- 3: Findings[2] |
|---|---|---|---|---|---|
| **KQ 4c. Withdrawals due to adverse events** | | | | | |
| ECT vs. rTMS | Withdrawals | Low | 1 fair cohort study: no difference in withdrawals between ECT and rTMS groups. | Insufficient | 2 trials (1 good, 1 fair): ECT and rTMS did not differ significantly in withdrawals attributed to adverse events.<br>1 fair cohort study: ECT group had more withdrawals than rTMS group (significance NR) |
| rTMS vs. sham | Withdrawals | Insufficient | 7 trials (1 good, 6 fair): Trials showed mixed results about withdrawals attributed to adverse events. | Insufficient | 11 RCTs (1 good, 10 fair): Trials showed mixed results about withdrawals attributed to adverse events. |
| VNS vs. sham | Withdrawals | Low | 1 good trial: VNS had greater withdrawals attributed to adverse events than sham (significance NR). | Low | No additional studies identified. |
| CBT vs. usual care | Withdrawals | NA | No eligible studies identified | High | 2 RCTs (1 good, 1 fair): CBT had no withdrawals attributed to adverse events. |
| **KQ 4d. Adherence as measured by overall withdrawals** | | | | | |
| ECT vs. rTMS | Overall withdrawals | Insufficient | 1 fair trial and 1 fair cohort study: Studies showed mixed results about withdrawals. | Insufficient | 3 trials (1 good, 2 fair) and 1 cohort study: Studies showed mixed results about withdrawals. |
| rTMS vs. sham | Overall withdrawals | Insufficient | 7 fair trials: Trials showed mixed results about withdrawals. | Insufficient | 13 trials (1 good, 1 fair): Trials showed mixed results about withdrawals. |
| CBT vs. usual care | Overall withdrawals | NA | No eligible studies identified | Insufficient | 3 trials (1 good, 2 fair): Trials showed mixed results about withdrawals. |
| **KQ 5. Efficacy and harms for selected populations** | | | | | |
| rTMS vs. sham | Changes in depressive severity | NA | No eligible studies identified | Low | 3 fair trials: In post-stroke depression, rTMS groups had significantly better depressive severity than sham groups |
| | Response | NA | No eligible studies identified | Low | 3 fair trials: in post-stroke depression, rTMS groups had better response rates than sham groups, but significance varied. |
| | Remission | NA | No eligible studies identified | Low | 3 fair trials: in post-stroke depression, rTMS groups had better remission rates than sham groups, but significance varied. |

Table 97. Summary of findings on nonpharmacologic treatment of adult major depressive disorder with strength of evidence (continued)

| Key Question and Comparison | Outcome | Tier 1 only | | Tiers 1- 3 | |
|---|---|---|---|---|---|
| | | Strength of Evidence[1] | Findings[2] | Strength of Evidence[1] | Findings[2] |
| **KQ 6. Health-related outcomes** | | | | | |
| rTMS vs. sham | Health-related outcomes | Insufficient | 1 fair trial: Mixed results. rTMS (low-right) group had significantly greater improvement in health status and daily functioning than sham group; rTMS (high-left) group had nonsignificantly greater improvement in health status and daily functioning than sham group. | Low | 2 trials (1 good, 1 fair): rTMS groups had significantly greater improvement in health status and daily functioning than sham groups (except for high left rTMS group in one trial, which had nonsignificantly greater improvement in health status and daily functioning than the sham group). |
| VNS vs. sham | Health-related outcomes | Low | 1 fair trial: VNS and sham groups did not differ significantly in daily functioning | Low | No additional studies identified. |
| CBT/DBT vs. control | Health-related outcomes | NA | No eligible studies identified. | Low | 1 fair trial: CBT group had nonsignificantly greater improvement in daily functioning than waitlist control group. |

[1]Strength of evidence is based on the EPC program's modified version of the GRADE system; see text.

[2]Good and fair designations relate to quality ratings for each study.

## KQ 1a: Efficacy of Acute-Phase Interventions: Nonpharmacologic Interventions Against Each Other

**Direct evidence.** *Tier 1 trials.* The available head-to-head literature concerning the efficacy of the nonpharmacologic interventions for Tier 1 TRD is limited to two fair trials (both in MDD-only populations) (Table 98). One compared ECT and rTMS, and the other compared ECT and ECT plus rTMS. They showed, with low strength of evidence, no differences between treatment options for depressive severity, response rates, and remission rates. No trial involved a direct comparison of psychotherapy with another nonpharmacologic intervention.

Table 98. Number of studies of head-to-head comparisons of nonpharmacologic treatments, by comparison and level of treatment resistance

| Comparison | Definition of Treatment Resistance (Tier) | | |
|---|---|---|---|
| | Tier 1 (TRD): ≥ 2 failures | Tier 2: ≥ 1 failure | Tier 3 Probable |
| ECT plus rTMS vs. ECT | 1 | 0 | 1 additional |
| ECT vs. rTMS | 1 | 1 additional | 1 additional |

ECT, electroconvulsive therapy; rTMS, repetitive transcranial magnetic stimulation

*Tiers 1-3 trials.* Considering trials for Tiers 1-3 modified our strength of evidence grades (to insufficient) for comparisons of ECT with rTMS; two additional trials produced conflicting

results about these outcomes. Considering trials for Tiers 1-3 did not change findings for ECT vs. ECT plus rTMS; that is, the strength of evidence remained low that two interventions did not differ in these outcomes.

**Indirect evidence.** We identified trials that compared a nonpharmacologic intervention, generally rTMS, VNS, or psychotherapy, with a control or sham procedure; we identified no ECT vs. control studies (Table 99). The numbers of these trials with the same or similar control group were very small, so we could not pool them quantitatively. We could, however, assess the potential benefits of nonpharmacologic interventions vs. controls by calculating mean changes in depressive severity, relative risks of response, and relative risks of remission.

Table 99. Number of studies of nonpharmacologic interventions against controls or usual care, by comparison and level of treatment resistance (tier)

| Intervention and Control | Definition of Treatment Resistance | | |
|---|---|---|---|
| | Tier 1 (TRD): ≥ 2 failures | Tier 2: ≥ 1 failure | Tier 3 Probable |
| rTMS vs. sham procedure | 12 | 4 additional | 3 additional |
| rTMS plus pharmacotherapy vs. sham plus pharmacotherapy | 0 | 1 additional | 1 additional |
| VNS plus usual care vs. usual care | 1 | 0 | 0 |
| Psychotherapy plus usual care vs. usual care | 0 | 3 additional | 0 |

rTMS, repetitive transcranial magnetic stimulation; VNS, vagus nerve stimulation.

*Tier 1 trials.* rTMS was beneficial relative to controls receiving a sham procedure for all three outcomes (severity of depressive symptoms, response rate, remission rate). rTMS produced a greater decrease in depressive severity (high strength of evidence). Specifically, rTMS averaged a decrease in depressive severity measured by the Hamilton Rating Scale for Depression (HAM-D) of more than 5 points relative to sham control, and this change meets the minimum threshold of the 3-point HAM-D difference that is considered clinically meaningful. Response rates were greater with rTMS than sham (also high strength of evidence); those receiving rTMS were more than three times as likely to achieve a depressive response as patients receiving sham procedure. Finally, rTMS was also more likely to produce remission than the control procedure (moderate strength of evidence); patients receiving rTMS were more than 6 times as likely to achieve remission as those receiving the sham.

In the only other Tier 1 comparison, one good-quality VNS vs. sham control trial (a mixed MDD/bipolar population) reported no differences between the groups as measured by a change in depressive severity or response rates (low strength of evidence).

*Data from Tiers 1-3*. Considering evidence from trials in all tiers provided data consistent with just Tier 1 findings favoring rTMS over controls. These additional studies raised the strength of evidence grade for better remission rates for rTMS vs. control from moderate to high.

Compared with Tier 1 quantitative syntheses, meta-analyses involving data from studies in all three tiers provided more conservative point estimates and narrower confidence intervals for each outcome, but these analyses reflect what can be expected in Tier 1 populations. For example, rTMS produced a decrease in severity of approximately 5 HAM-D points relative to sham control. Also, patients receiving rTMS were more than two times as likely to achieve response and more than two times as likely to achieve remission than patients receiving sham control.

Finally, considering evidence from all tiers added studies of psychotherapy against controls. Three studies from MDD-only populations provided, overall, low strength of evidence that CBT

groups had better outcomes than control groups in response rates and remission rates. Evidence was insufficient concerning any difference between CBT and controls in changing depressive severity.

## KQ 1b: Efficacy of Acute-Phase Interventions: Nonpharmacologic Interventions Against Medications

**Direct evidence.** *Tier 1 trials.* The available head-to-head literature concerning the efficacy of the nonpharmacologic interventions compared with pharmacologic treatment (in this case, paroxetine) for Tier 1 trials is limited to one fair trial (a mixed MDD/bipolar population). ECT produced a significantly greater decrease in depressive severity (9 points by HAM-D) and significantly better response rates (71 percent vs. 28 percent) than medications (low strength of evidence) (Table 100).

Table 100. Number of studies involving pharmacotherapy comparisons, by level of treatment resistance (tier)

| Intervention | Definition of Treatment Resistance | | |
|---|---|---|---|
| | Tier 1 (TRD): ≥ 2 failures | Tier 2: ≥ 1 failure | Tier 3 Probable |
| ECT vs. pharmacotherapy | 1 | 0 | 0 |
| CBT vs. pharmacotherapy | 0 | 2 additional | 0 |
| rTMS + pharmacotherapy vs. pharmacotherapy | 0 | 1 additional | 0 |

ECT, electroconvulsive therapy; CBT, cognitive behavioral therapy

*Tiers 1-3 trials.* No additional trials were identified for ECT.

For psychotherapy versus pharmacotherapy, data from two trials (MDD-only patients) suggested that outcomes for the CBT groups did not differ significantly from those for pharmacotherapy control groups for depressive severity or response or remission rates (low strength of evidence). In mildly to moderately depressed patients, outcomes did not differ regardless of whether investigators augmented treatment with CBT or medications (response rates ranged from 28 percent to 35 percent, and remission rates from 31 percent to 33 percent), whether they switched patients to either CBT or a new medication (response rates ranged from 22 percent to 27 percent, and remission rates ranged from 27 percent to 31 percent), or whether they switched to CBT or continued medication (response and remission rates not provided).

Finally, a fair trial reported that rTMS plus a new pharmacotherapy produced a nonsignificantly greater decrease in depressive severity than a sham control plus that new pharmacotherapy (low strength of evidence).

**Indirect evidence**. *Tier 1 trials.* Indirect evidence about procedures or psychotherapy (vs. sham or nonpharmaceutical controls) were presented above as part of KQ 1.

We attempted to determine mean changes in depressive severity, relative risks of response, and relative risks of remission for pharmacologic vs. control studies to allow a comparison with similar outcomes in the nonpharmacologic vs. control trials (KQ 1a, indirect). However, there were no comparable, common control groups not receiving a mood-related medication to allow such comparisons.

Instead, we determined mean average outcomes for pharmacologic treatments. For switching strategies, mean pharmacologic response rates averaged 35.3 percent (95% CI, 28.4% to 42.3%) and mean remission rates averaged 19.8 percent (95% CI, 14.3% to 25.2%); for augmentation, mean response rates averaged 37.5 percent (27.1% to 47.4%) and mean response rates average 25.3 percent (15.6% to 34.9%); and for maintenance strategies, mean response rates averaged 29.9 percent (16.4% to 43.4%) and mean remission rates averaged 16.8 percent (7.8% to 25.8%).

While these results provide an idea of the general degree of response seen with next-step pharmacologic treatment in TRD, they serve as an uncontrolled case series and should only be compared to nonpharmacologic outcomes with caution.

## KQ 2. Efficacy of Nonpharmacologic Interventions for Maintenance of Remission or Prevention of Relapse

**Direct evidence.** With respect to maintaining remission (or preventing relapse), very little comparative evidence was available about the procedure-based interventions or psychotherapy.

*Tier 1 trials.* We found no direct comparisons for ECT, rTMS, VNS, or CBT.

*Tiers 1-3 trials.* For just Tiers 2-3 (in this case), two fair trials compared ECT with rTMS (one with MDD-only patients and one with a mixed MDD/bipolar patient population). The two interventions did not differ significantly in the likelihood of maintaining remission (low strength of evidence).

**Indirect evidence.** Information about interventions vs. control comparisons testing efficacy for remission or relapse was also very limited.

*Tier 1 trials.* Two fair trials compared rTMS with a sham procedure and found no significant differences except at short-term followup. Because very few patients were followed during the relapse prevention phases of these studies, we concluded that this evidence was insufficient to draw meaningful conclusions about rTMS. We had no eligible studies for ECT, VNS, or psychotherapy.

*Tier 1-3 trials*. For rTMS, six fair sham-controlled rTMS trials suggested that rTMS may maintain remission better than a sham procedure (low strength of evidence with little information about statistical significance of differences).

For CBT, one fair trial reported that CBT maintained remission better than usual care, with one measure showing significant benefit and a second measure nonsignificant benefit for CBT vs. sham (low strength of evidence).

## KQ 3. Efficacy of Nonpharmacologic Interventions for Patients with Different Symptomatology

**Direct evidence**. Very little comparative evidence addressed whether procedure-based treatments differed as a function of symptom subtypes, and we identified no Tier 1 trials.

For Tier 3 (in this case), a fair trial of a mixed MDD/bipolar population reported that patients with psychotic symptoms who received ECT had a significantly greater decrease in depressive severity than those receiving rTMS; however, the finding had a substantial cointervention bias concerning differences in continuation of psychotropic medications (ECT patients were allowed to continue antipsychotics, while rTMS patients were not). Accordingly, we judged strength of evidence to be insufficient.

No comparative evidence was available about psychotherapy in subgroups defined by symptom clusters.

**Indirect evidence.** We identified no studies testing either procedure-based or psychotherapeutic interventions against sham procedures or other controls.

## KQ 4. Harms of Nonpharmacologic Interventions

**Direct evidence.** In examining safety, adverse events, and adherence, we found some differences across the interventions in the harms and negative side effects to patients. For just this set of analyses, we examined both trials and cohort studies, and we focus on cognitive functioning, occurrence of specific adverse events, and withdrawals.

*Cognitive functioning*. For Tier 1 studies on cognitive functioning, ECT may have a greater negative (but transient) effect than rTMS within the six month period following treatment. Across Tier 1-3 studies, findings do not change materially. Strength of evidence in both cases is low.

*Specific adverse events.* We identified no Tier 1 studies comparing two interventions on this outcome. Across Tier 1-3 studies, one good trial reported that ECT and rTMS did not differ significantly (low strength of evidence).

*Withdrawals.* We looked at both withdrawals that investigators attributed to adverse events and overall numbers or rates of withdrawals. Evidence for ECT compared with rTMS (Tier 1 but especially Tiers 1-3) generally produced mixed or inconclusive information as to whether ECT had higher rates of withdrawals than rTMS groups (low or insufficient strength of evidence).

**Indirect evidence.** We attempted to include data from the same types of studies and for the same outcomes as for direct evidence. We identified no studies comparing ECTs vs. control.

*Cognitive functioning.* rTMS may have a positive impact on some aspects of cognitive functioning relative to a sham procedure (for four trials in Tier 1); most comparisons are not statistically significant and the strength of evidence is low in all cases. Findings and strength of evidence do not change materially when considering Tier 1-3 studies.

*Specific adverse events.* rTMS groups had significantly more scalp pain at the stimulation site and, when adding in Tier 2-3 trials, more difficulty starting urination than groups receiving sham procedures (low strength of evidence).

*Withdrawals.* Findings were mixed in both Tier 1 and Tier 1-3 studies as to whether rTMS groups had greater rates of withdrawals than groups receiving sham procedures (insufficient evidence).

There was low strength of evidence that there were greater withdrawals due to adverse events in the vagus nerve stimulation group compared to sham.

No Tier 1 studies reported on withdrawals for CBT groups versus those receiving some form of usual care. Some Tier 1-3 studies suggested that CBT patients did not withdraw (because of adverse events) (high strength of evidence); others reported mixed results about whether CBT and usual care patients differed in adherence, as measured by overall withdrawals (insufficient evidence).

## KQ 5. Efficacy or Harms of Nonpharmacologic Treatments for Selected Patient Subgroups

**Direct evidence.** We found no studies (in any tier) directly comparing nonpharmacologic interventions in selected populations, such as the elderly, those with stroke, or those with other medical comorbidities.

**Indirect evidence.** No Tier 1 trials provided indirect evidence about patient subgroups. However, three fair trials (all Tier 2), assessed whether the effects of nonpharmacologic treatments differed for a subpopulation comprising patients with post-stroke depression.

Specifically, compared with groups receiving sham procedure, patients receiving rTMS had significantly better levels of depression severity (low strength of evidence) and were more likely to have better response and remission (but significance varied) (low strength of evidence).

## KQ 6. Health-Related Outcomes of Nonpharmacologic Treatments

**Direct evidence.** With respect to patient-reported health-related outcomes, we focused on quality of life (various measures) and ability to function in daily life. No trials (either Tier 1 or Tier 1-3) directly compared the effects of nonpharmacologic treatments for these outcomes.

**Indirect evidence.** *Tier 1 trials.* Two trials (both in mixed MDD/bipolar populations) assessed general health status and mental and physical functioning (all health domains related to quality of life). One fair trial on rTMS vs. sham yielded mixed results (insufficient evidence) as to whether the procedures produced greater improvement in health status and daily functioning (as measured by the Global Assessment of Functioning scale). In the other fair trial, VNS and sham groups did not differ significantly in daily functioning (as measured by the 36-item Medical Outcomes Study Short Form [MOS SF-36]). No studies of psychotherapy were identified.

*Tier 1-3 trials.* Two rTMS vs. sham trials (one good, one fair) provided evidence that rTMS groups had significantly greater improvement in the MOS SF-36 and the Quality of Life, Enjoyment and Satisfaction Questionnaire (QLESQ) than controls (raising the strength of evidence grade to low). No additional studies were available for VNS.

The new CBT vs. control trial reported that the psychotherapy group had a nonsignificantly better level of daily functioning as measured by the Lifework-Range of Impaired Functioning Tool than a wait list control group (low strength of evidence).

# Applicability

For the limited amount and low strength of evidence available, the data for Tier 1 (TRD) is generally applicable to TRD populations. Populations enrolled in these trials appeared representative of our target population. With the exception of rTMS, which had varying doses and length of treatment, the studied interventions appeared comparable with those in routine use. Measured outcomes on the whole reflected the most important clinical outcomes for depression measures, although reporting was inconsistent; outcomes for the other key questions were much more restricted. Follow-up periods were generally short than desirable, but most were sufficient to measure an initial acute phase treatment response. Study settings were a mixture of inpatient and outpatient, as ECT is generally an inpatient procedure while the others are generally outpatient. The use of varying definitions of TRD in the trials and the absence of analyses considering the effect of the number of current treatment failures on outcomes hindering interpretation of data, leading to the use of a tiered system. The evidence base combining data for Tiers 1-3 on the whole produced findings that were consistent with Tier 1 TRD data and also appear applicable to TRD populations.

## Limitations of the Evidence Base

**Lack of a use of standard definition of TRD.** Comparison of any of the potential interventions in the field, nonpharmacologic or otherwise, is hampered by variable definitions of TRD. While these definitions appear to be consolidating towards a single meaning—2 or more

treatment failures in the current episode—very few studies of TRD have applied it. Use of multiple definitions makes synthesis of the available information difficult, as the effect of combining patients with 1 treatment failure with those of 2 or more (or 4 or more) remains unclear.

Similarly, the failure of studies to describe the number of treatment failures prevented us from being able to stratify our outcomes by the number of failed trials within Tier 1 studies and assess the role of number of failures in TRD on outcomes.

**Failure to consistently assess number of failures in current episode.** Given the difficulty in accurately assessing adequacy of prior treatment trials over a lifetime, a history of failed treatment attempts in a current episode is likely a more accurate measure of treatment resistance. Likely, many of those who reported lifetime histories of two or more failures did have them in the current episode, but few studies required such a failure in their selection criteria so that many studies may be mixing current failure with more chronic failures.

**Few head-to-head studies of nonpharmacologic intervention.** The small number of existing head-to-head studies limits the strength of all our findings to either low or insufficient evidence, making firm conclusion about comparative effectiveness impossible. Only two studies occurred in our main population of interest: Patients with MDD who had failed two or more antidepressants.[51,52]

**Heterogeneity of the populations (MDD and MDD/bipolar mix).** This mixture of diagnostic disorders in samples made interpretation of the data difficult. Populations studied included MDD and MDD/Bipolar mix patients. We selected a 20 percent cut-off to decrease the likelihood of the mix affecting outcomes (e.g., in a study of 40 patients, if 8 had bipolar disorder and were roughly evenly distributed between treatment arms, their outcomes would need to be extreme to substantially affect outcome). This need to clarify a specific cut-off, however, excluded studies that may have had relevant populations. Further, because results were not stratified by MDD and bipolar disease, the precision of the effect on the nonpharmacologic outcomes may have been distorted.

**Failure to consider a spectrum of depressive severity.** Most patients involved in studies were severely depressed and analyses did not assess how the degree of depression along the severe spectrum may affect outcomes in comparative studies. For example, the most severely depressed may have different outcomes with one vs. another intervention than those who are severely depressed but to a lesser degree.

**Heterogeneity of interventions and intervention strategies.** The literature is characterized by a large variety of treatment strategies used (augmentation, switch, a combination of the two), a wide variety of treatment parameters used (length and dose of ECT, number of rTMS sessions), and variable and uncontrolled use of psychotropic medications, all of which make interpretation and synthesis of the studies difficult.

**Outcome elements assessed.** While reporting one or two of the pertinent outcomes, the majority of the relevant studies did not assess both response and remission rates. These measures are especially important to allow a clinically meaningful interpretation of findings.

**Few comparisons of nonpharmacologic to pharmacologic treatments in TRD patients.** For many clinicians, following failure of two antidepressant treatments, the next step is not consideration of a nonpharmacologic treatment but is usually considering of a different pharmacologic strategy. The role of nonpharmacologic interventions in the sequence of treatment choices remains unclear.

**Inadequate study design to assess longer term outcomes.** Studies need to have more long term monitoring over time so that the outcomes can be further studied. For example, the available studies for ECT did not follow patients long enough to assess potential cumulative effects on cognitive functioning that may distinguish it from other interventions. Additionally, longer monitoring periods are necessary to compare the maintenance of remission.

**Studies were not designed to answer many of the outcomes relevant to key question**. Outcomes such as relapse, cognitive functioning, adverse events, withdrawal due to adverse events, and health-related outcomes are not often primary outcomes, limiting the power to adequately test hypotheses about such differences between nonpharmacologic interventions.

**These treatments are quite different.** Differences in these interventions—how long it takes to reach an adequate dose, how effectively patients can be blinded, how long it takes to obtain a response, how long the results last—make it challenging to directly compare these varying treatments. For example, with rTMS, if there is no effect in two weeks, one might consider switching treatments, whereas with CBT, such a latency would not be a cause for concern.

## Limitations of this Review

This area of comparative clinical research is early in its infancy and few relevant trials were available. The paucity of data limited our ability to pool findings statistically. Specifically, we were not able to quantitatively synthesize data from head-to-head comparisons, nor were we able to indirectly compare the nonpharmacologic literature by pooling data from studies sharing equivalent control groups. Our synthesis, then, is primarily qualitative.

The dearth of relevant trials also prevented us from assessing whether key elements might suggest one nonpharmacologic treatment over another. In particular, we were unable to assess what the effect on outcome was of key, clinically relevant elements of interest: population variables (MDD and MDD/bipolar mix; varying depressive severity; and requiring treatment failures to be in the current episode) and intervention variables (using an augment vs. switch treatment strategy; varying by nonpharmacologic treatment characteristics).

## Future Research

This area of comparative clinical research is early in its infancy. Key areas for future research need primarily to lay more robust foundations for an evidence base which can better inform decisions for clinicians and patients..

**The field needs a standard definition of TRD that investigators should use in their clinical trials research.** Comparison of any of the potential interventions in the field, nonpharmacologic or otherwise, is hampered by the variability in TRD definitions. Although these definitions appear to be converging on a single meaning—two or more treatment failures in the current episode—very few studies of TRD have applied it. Progress in this area of research requires better standardization of this concept, so that future reviews of the evidence do not need to resort to differentiating, as we did, between “tier 1” studies (i.e., TRD by this definition based on two or more treatment failures) and “tier 2 or 3” types of studies. The latter do provide information that helps illuminate likely impacts of these interventions on patients with TRD, but that is not the same thing as having robust studies focused clearly on the patient population of greatest interest.

**More clinical trials comparing nonpharmacologic interventions with other nonpharmacologic options and with pharmacologic treatmentsare necessary to inform decisionmaking in TRD.** Clinicians, patients, and policymakers need additional relevant data to guide difficult treatment decisions about what to do next: try another medication trial (and should it be an augmentation, switch, or combination strategy?); add (or switch to) rTMS, ECT, VNS, or psychotherapy?

Also, given that treatment options for many TRD patients include medications, trials should directly compare nonpharmacologic interventions with each other *and* with pharmacologic treatments.

**The numbers of failures in current episode should be delineated carefully.** This information, more likely to be accurate than life-time histories of failures, can help investigators determine whether the particular number of failures, or reaching a particular number of failures in a current episode, can help differentiate between nonpharmacologic treatment choices. For example, for patients with two failures in a current episode, the outcomes may not differ between cognitive therapy and rTMS; however, for patients with a different (higher or lower) number of failures in the current episode, one nonpharmacologic treatment may indeed be better than the other. Currently, we do not know what the proper threshold is for selection of treatment. Clarification of the scientific basis for such a decision would substantially improve decisionmaking.

**Determining whether responses differ for patients with MDD and those with bipolar disorder will help to guide future clinical trial design.** Our decision to include trials with patient populations including up to 20 percent with bipolar disorder (i.e, the "mixed" populations noted earlier) was guided by clinical experience and common sense but not by data. Testing to see whether outcomes differ between the two groups can yield information about inclusion criteria (should the mix be 0%, 10%, 20%, etc.?) that may be useful to investigators in designing TRD trials and may be important to consider as a potential covariate in analyses involving such mixes.

**Greater consideration should be given to the role that the spectrum of depressive severity plays.** Using a finer gradation of depressive severity than investigators now typically employ might identify whether particularly severe degrees of depression, most commonly understood currently as a HAM-D $\geq$ 20, may respond differently to the available nonpharmacologic interventions than do less severe levels of depression. These gradations may lead clinicians to a better understanding of severe depression and its role in guiding treatment selection in TRD.

**Direct comparisons of treatment strategies, holding consistent any coexisting or concomitant therapies, are imperative.** Decisionmakers need to know whether outcomes with nonpharmacologic treatments are better when such a treatment augments the current treatment, replaces the current treatment, or replaces the current treatment in combination with another treatment. When ongoing treatment is uncontrolled and reflects a variety of treatments—e.g., some patients continue with atypical antipsychotics, some with mood stabilizers, some with no psychotropic medications—results of study studies are difficult, if not impossible, to interpret.

**Consistent reporting of changes in depressive severity, response rates, and remission rates is crucial** To allow for better comparisons of clinical outcomes in this difficult-to-treat population, all three measures offer useful information for clinicians. Thus, for either trials or observational studies, investigators should attempt to collect data on all three routinely.

**Application of consistent, accepted protocols in trials.** Making sure that patients receive equivalent doses of different nonpharmacologic interventions is more difficult than making sure of this for pharmacologic interventions. Nevertheless, investigators designing trials of nonpharmacologic therapies can attempt to do so by implementing standard accepted protocols for their trials. Such "dosing" is, admittedly, difficult to control when that protocol is in the process of being developed, as with rTMS, but this standardization is a goal well worth trying to reach.

**More careful and consistent assessment of adverse events is required.** Adverse event reporting is quite limited and over a short time span, and what exists is variable and inconsistent. Systematic collection and more consistent reporting of data on harms—i.e., adverse events and negative side effects—and information about attrition and withdrawal would provide useful information to help balance information now focused on clinical benefits. Use of the CONSORT statement (http://www.consort-statement.org/home/), which guides proper reporting of study information (including the presentation of adverse events), would strengthen reporting both harms and other clinical trial findings; it would also aid in the critical appraisal and interpretation of all study results. Further, a more informative assessment of adverse events would require studies to be able to assess long term and cumulative outcomes.

**Including key relevant measures and subgroups in subsequent research is desirable.** As indicated by the review, nearly no evidence exists on how the effectiveness of nonpharmacologic treatments differs (or not) as a function of symptom subtypes or for subgroups defined by sociodemographic characteristic (such as age) or coexisting medical conditions (e.g., post-stroke or post-myocardial infarction depression; perinatal depression). Also essentially missing is information about health-related outcomes, especially those reported by patients, that concern their quality of life or levels of functional impairment Subsequent studies should focus on employing known, reliable, and valid measures of patient-reported outcomes, such as the MOS SF-36,[116] the QLESQ,[117] and the EQ-5D.[118]

**Including comparisons of newer nonpharmacologic interventions will be important in future research.** As new nonpharmacologic treatments are developed and tested, investigators should try to include them as potential comparators. At the time we started this comparative effectiveness review, clinical trial data on some of the developing nonpharmacologic interventions, such as magnetic seizure therapy.[119-121] or deep brain stimulation,[122-124] were insufficient (from the published literature) for us to try to include them. As the evidence bases grow to support the efficacy of such additional nonpharmacologic interventions, the newer strategies should be included in comparative effectiveness study designs.

# Conclusion

Our review suggests that comparative clinical research on nonpharmacologic interventions in a TRD population is early in its infancy, and many clinical questions about efficacy and effectiveness remain unanswered. Interpretation of the data is substantially hindered by varying definitions of TRD and the paucity of relevant studies. The greatest volume of evidence is for ECT and rTMS; however, even for the few comparisons of treatments that are supported by some evidence, the strength of evidence is low for benefits. There was low strength of evidence that ECT and rTMS produced a similar likelihood of maintaining remission. ECT produced a greater negative impact on cognitive functioning compared to rTMS. The few trials addressing subpopulations, subtypes, and health-related outcomes provided low or insufficient evidence of

differences between nonpharmacologic interventions. The most urgent next steps for research are to apply a consistent definition of TRD, to conduct more head-to-head clinical trials comparing nonpharmacologic interventions to themselves and to pharmacologic treatments, and careful delineation of the number of adequate treatment failures in the current episode.

# References


1. 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.

2. 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.

3. American Psychiatric Association. Practice guideline for the treatment of patients with major depressive disorder (revision). American Journal of Psychiatry. 2000 Apr;157(4 Suppl):1-45.

4. Qaseem A, Snow V, Denberg TD, et al. Using second-generation antidepressants to treat depressive disorders: a clinical practice guideline from the American College of Physicians. Ann Intern Med. 2008 Nov 18;149(10):725-33.

5. Kupfer DJ. Long-term treatment of depression. J Clin Psychiatry. 1991 May;52 Suppl:28-34.

6. Trivedi MH, Rush AJ, Wisniewski SR, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. Am J Psychiatry. 2006 Jan;163(1):28-40.

7. Thase ME, Rush AJ. When at first you don't succeed: sequential strategies for antidepressant nonresponders. J Clin Psychiatry. 1997;58 Suppl 13:23-9.

8. 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. American Journal of Psychiatry. 2006;163(11):1905-17.

9. Russell JM, Hawkins K, Ozminkowski RJ, et al. The cost consequences of treatment-resistant depression. J Clin Psychiatry. 2004 Mar;65(3):341-7.

10. Crown WH, Finkelstein S, Berndt ER, et al. The impact of treatment-resistant depression on health care utilization and costs. J Clin Psychiatry. 2002 Nov;63(11):963-71.

11. Greenberg P, Corey-Lisle PK, Birnbaum H, et al. Economic implications of treatment-resistant depression among employees. Pharmacoeconomics. 2004;22(6):363-73.

12. Gartlehner G, Gaynes BN, Hansen RA, et al. Comparative benefits and harms of second-generation antidepressants: background paper for the American College of Physicians. Ann Intern Med. 2008 Nov 18;149(10):734-50.

13. Tharyan P. The Cochrane Schizophrenia Group: preparing, maintaining and disseminating the evidence for interventions used for people with schizophrenia. Int Rev Psychiatry. 2005 Apr;17(2):115-21.

14. Greenhalgh J, Knight C, Hind D, et al. Clinical and cost-effectiveness of electroconvulsive therapy for depressive illness, schizophrenia, catatonia and mania: systematic reviews and economic modelling studies. Health Technol Assess. 2005 Mar;9(9):1-156, iii-iv.

15. Lisanby SH. Electroconvulsive therapy for depression. N Engl J Med. 2007 Nov 8;357(19):1939-45.

16. Centers for Medicare & Medicaid Services. Update-Inpatient Psychiatric Facilities Prospective Payment System (IPF PPS). Rate Year 2009, Publication # 100-04 2008 June 27.

17. Fitzgerald PB, Benitez J, de Castella A, et al. A randomized, controlled trial of sequential bilateral repetitive transcranial magnetic stimulation for treatment-resistant depression. Am J Psychiatry. 2006 Jan;163(1):88-94.

18. Feldman G. Cognitive and behavioral therapies for depression: overview, new directions, and practical recommendations for dissemination. Psychiatr Clin North Am. 2007 Mar;30(1):39-50.

19. Schramm E, van Calker D, Dykierek P, et al. An intensive treatment program of interpersonal psychotherapy plus pharmacotherapy for depressed inpatients: acute and long-term results. Am J Psychiatry. 2007 May;164(5):768-77.

20. George MS, Belmaker RH, eds. Transcranial magnetic stimulation in clinical psychiatry. Arlington, VA: American Psychiatric Publishing 2007.

21. Shapira B, Tubi N, Lerer B. Balancing speed of response to ECT in major depression and adverse cognitive effects: role of treatment schedule. J Ect. 2000 Jun;16(2):97-109.

22. Rossi S, Hallett M, Rossini PM, et al. Safety, ethical considerations, and application guidelines

for the use of transcranial magnetic stimulation in clinical practice and research. Clin Neurophysiol. 2009 Dec;120(12):2008-39.

23. Daban C, Martinez-Aran A, Cruz N, et al. Safety and efficacy of Vagus Nerve Stimulation in treatment-resistant depression. A systematic review. J Affect Disord. 2008 Sep;110(1-2):1-15.

24. Patel M, Patel S, Hardy DW, et al. Should electroconvulsive therapy be an early consideration for suicidal patients? J ECT. 2006 Jun;22(2):113-5.

25. Tew JD, Jr., Mulsant BH, Haskett RF, et al. Acute efficacy of ECT in the treatment of major depression in the old-old. Am J Psychiatry. 1999 Dec;156(12):1865-70.

26. Van Der Wurff FB, Stek ML, Hoogendijk WJ, et al. The efficacy and safety of ECT in depressed older adults: a literature review (Brief record). International Journal of Geriatric Psychiatry. 2003(10):894-904.

27. Simpson KN, Welch MJ, Kozel FA, et al. Cost-effectiveness of transcranial magnetic stimulation in the treatment of major depression: a health economics analysis. Adv Ther. 2009 Mar;26(3):346-68.

28. Walter G, Martin J, Kirkby K, et al. Transcranial magnetic stimulation: experience, knowledge and attitudes of recipients. Aust N Z J Psychiatry. 2001 Feb;35(1):58-61.

29. Marangell LB, Martinez M, Jurdi RA, et al. Neurostimulation therapies in depression: a review of new modalities. Acta Psychiatr Scand. 2007 Sep;116(3):174-81.

30. Melkerson MN. Special Premarket 510(k) Notification for NeuroStar TMS Therapy System for Major Depressive Disorder. 2008 December 16 [cited March 13, 2010]; Available from: http://www.accessdata.fda.gov/cdrh_docs/pdf8/K083538.pdf

31. Shuchman M. Approving the vagus-nerve stimulator for depression. New England Journal of Medicine. 2007;356(16):1604-7.

32. Lulic D, Ahmadian A, Baaj AA, et al. Vagus nerve stimulation. Neurosurg Focus. 2009 Sep;27(3):E5.

33. Cohen LJ, Allen JC, Jr. Estimating the potential savings with vagus nerve stimulation for treatment-resistant depression: a payer perspective. Curr Med Res Opin. 2008 Aug;24(8):2203-17.

34. Rush AJ, Siefert SE. Clinical issues in considering vagus nerve stimulation for treatment-resistant depression. Exp Neurol. 2009 Sep;219(1):36-43.

35. Rush AJ, Sackeim HA, Marangell LB, et al. Effects of 12 months of vagus nerve stimulation in treatment-resistant depression: a naturalistic study. Biol Psychiatry. 2005 Sep 1;58(5):355-63.

36. George MS, Rush AJ, Marangell LB, et al. A one-year comparison of vagus nerve stimulation with treatment as usual for treatment-resistant depression. Biol Psychiatry. 2005 Sep 1;58(5):364-73.

37. Luty SE, Carter JD, McKenzie JM, et al. Randomised controlled trial of interpersonal psychotherapy and cognitive-behavioural therapy for depression. Br J Psychiatry. 2007 Jun;190:496-502.

38. Bortolotti B, Menchetti M, Bellini F, et al. Psychological interventions for major depression in primary care: a meta-analytic review of randomized controlled trials. Gen Hosp Psychiatry. 2008 Jul-Aug;30(4):293-302.

39. Butler AC, Chapman JE, Forman EM, et al. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. Clin Psychol Rev. 2006 Jan;26(1):17-31.

40. 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.

41. Shelton RC, Osuntokun O, Heinloth AN, et al. Therapeutic options for treatment-resistant depression. CNS Drugs. 2010 Feb 1;24(2):131-61.

42. Berlim MT, Turecki G. What is the meaning of treatment resistant/refractory major depression (TRD)? A systematic review of current randomized trials. Eur Neuropsychopharmacol. 2007 Nov;17(11):696-707.

43. Berlim MT, Turecki G. Definition, assessment, and staging of treatment-resistant refractory major depression: a review of current concepts and methods. Can J Psychiatry. 2007 Jan;52(1):46-54.

44. Fornaro M, Giosue P. Current nosology of treatment resistant depression: a controversy

resistant to revision. Clin Pract Epidemiol Ment Health. 2010;6:20-4.

45. Malhi GS, Parker GB, Crawford J, et al. Treatment-resistant depression: resistant to definition? Acta Psychiatr Scand. 2005 Oct;112(4):302-9.

46. Goldberg DP, Andrews G, Hobbs MJ. Where should bipolar disorder appear in the meta-structure? Psychological Medicine. 2009;39(12):2071-81.

47. Lisanby SH, Gutman D, Luber B, et al. Sham TMS: intracerebral measurement of the induced electrical field and the induction of motor-evoked potentials. Biol Psychiatry. 2001 Mar 1;49(5):460-3.

48. Loo CK, Taylor JL, Gandevia SC, et al. Transcranial magnetic stimulation (TMS) in controlled treatment studies: are some "sham" forms active? . Biol Psychiatry. 2004;47(4):325-31.

49. Inventory of Depressive Symptomatology (IDS) and Quick Inventory of Depressive Symptomatology (QIDS). 2010 [cited 2010 May 28 ]; Available from: http://www.ids-qids.org/index2.html

50. Owens DK, Lohr KN, Atkins D, et al. Grading the strength of a body of evidence when comparing medical interventions--Agency for Healthcare Research and Quality and the Effective Health Care Program. Journal of Clinical Epidemiology. 2010;63(5):513-23.

51. Rosa MA, Gattaz WF, Pascual-Leone A, et al. Comparison of repetitive transcranial magnetic stimulation and electroconvulsive therapy in unipolar non-psychotic refractory depression: a randomized, single-blind study. Int J Neuropsychopharmacol. 2006 Dec;9(6):667-76.

52. Pridmore S. Substitution of rapid transcranial magnetic stimulation treatments for electroconvulsive therapy treatments in a course of electroconvulsive therapy. Depress Anxiety. 2000;12(3):118-23.

53. Grunhaus L, Schreiber S, Dolberg OT, et al. A randomized controlled comparison of electroconvulsive therapy and repetitive transcranial magnetic stimulation in severe and resistant nonpsychotic major depression. *Biol Psychiatry* 2003:324-31.

54. McLoughlin DM, Mogg A, Eranti S, et al. The clinical effectiveness and cost of repetitive transcranial magnetic stimulation versus electroconvulsive therapy in severe depression: a multicentre pragmatic randomised controlled trial and economic analysis. Health Technol Assess. 2007 Jul;11(24):1-54.

55. Eranti S, Mogg A, Pluck G, et al. A randomized, controlled trial with 6-month follow-up of repetitive transcranial magnetic stimulation and electroconvulsive therapy for severe depression. Am J Psychiatry. 2007 Jan;164(1):73-81.

56. Knapp M, Romeo R, Mogg A, et al. Cost-effectiveness of transcranial magnetic stimulation vs electroconvulsive therapy for severe depression: A multi-centre randomised controlled trial. *Journal of Affective Disorders* 2008:273-85.

57. Chistyakov AV, Kaplan B, Rubichek O, et al. Effect of electroconvulsive therapy on cortical excitability in patients with major depression: a transcranial magnetic stimulation study. Clin Neurophysiol. 2005 Feb;116(2):386-92.

58. Folkerts HW, Michael N, Tölle R, et al. Electroconvulsive therapy vs. paroxetine in treatment-resistant depression -- a randomized study. Acta Psychiatr Scand. 1997 Nov;96(5):334-42.

59. Paykel ES, Scott J, Teasdale JD, et al. Prevention of relapse in residual depression by cognitive therapy. A controlled trial. Archives of General Psychiatry. 1999;56(9):829-35.

60. Scott J, Teasdale JD, Paykel ES, et al. Effects of cognitive therapy on psychological symptoms and social functioning in residual. British Journal of Psychiatry. 2000;177(NOV.):440-6.

61. Padberg F, Zwanzger P, Thoma H, et al. Repetitive transcranial magnetic stimulation (rTMS) in pharmacotherapy-refractory major depression: comparative study of fast, slow and sham rTMS. *Psychiatry Res* 1999:163-71.

62. Kauffmann CD, Cheema MA, Miller BE. Slow right prefrontal transcranial magnetic stimulation as a treatment for medication-resistant depression: a double-blind, placebo-controlled study. Depression and Anxiety. 2004;19(1):59-62.

63. Garcia-Toro M, Mayol A, Arnillas H, et al. Modest adjunctive benefit with transcranial magnetic stimulation in medication-resistant depression. Journal of Affective Disorders. 2001 May 2001;64(2):271-5.

64. Holtzheimer PE, 3rd, Russo J, Claypoole KH, et al. Shorter duration of depressive episode may

predict response to repetitive transcranial magnetic stimulation. Depress Anxiety. 2004;19(1):24-30.

65. Garcia-Toro M, Salva J, Daumal J, et al. High (20-Hz) and low (1-Hz) frequency transcranial magnetic stimulation as adjuvant treatment in medication-resistant depression. Psychiatry Res. 2006 Jan 30;146(1):53-7.

66. Pascual-Leone A, Rubio B, Pallardo F, et al. Rapid-rate transcranial magnetic stimulation of left dorsolateral prefrontal cortex in drug-resistant depression. Lancet. 1996 Jul 27;348(9022):233-7.

67. Avery DH, Holtzheimer PE, 3rd, Fawaz W, et al. A controlled study of repetitive transcranial magnetic stimulation in medication-resistant major depression. Biol Psychiatry. 2006 Jan 15;59(2):187-94.

68. Boutros NN, Gueorguieva R, Hoffman RE, et al. Lack of a therapeutic effect of a 2-week sub-threshold transcranial magnetic stimulation course for treatment-resistant depression. Psychiatry Research. 2002 Dec 2002;113(3):245-54.

69. Manes F, Jorge R, Morcuende M, et al. A controlled study of repetitive transcranial magnetic stimulation as a treatment of depression in the elderly. International Psychogeriatrics. 2001 Jun 2001;13(2):225-31.

70. Moser DJ, Jorge RE, Manes F, et al. Improved executive functioning following repetitive transcranial magnetic stimulation. *Neurology* 2002:1288-90.

71. Stern WM, Tormos JM, Press DZ, et al. Antidepressant effects of high and low frequency repetitive transcranial magnetic stimulation to the dorsolateral prefrontal cortex: a double-blind, randomized, placebo-controlled trial. J Neuropsychiatry Clin Neurosci. 2007 Spring;19(2):179-86.

72. O'Reardon JP, Solvason HB, Janicak PG, et al. Efficacy and safety of transcranial magnetic stimulation in the acute treatment of major depression: a multisite randomized controlled trial. *Biol Psychiatry* 2007:1208-16.

73. Bocchio-Chiavetto L, Miniussi C, Zanardini R, et al. 5-HTTLPR and BDNF Val66Met polymorphisms and response to rTMS treatment in drug resistant depression. Neurosci Lett. 2008 May 30;437(2):130-4.

74. Fitzgerald PB, Brown TL, Marston NA, et al. Transcranial magnetic stimulation in the treatment of depression: a double-blind, placebo-controlled trial. Arch Gen Psychiatry. 2003 Oct;60(10):1002-8.

75. Su T-P, Huang C-C, Wei IH. Add-On rTMS for medication-resistant depression: a randomized, double-blind, sham-controlled trial in Chinese patients. Journal of Clinical Psychiatry. 2005 Jul 2005;66(7):930-7.

76. Berman RM, Narasimhan M, Sanacora G, et al. A randomized clinical trial of repetitive transcranial magnetic stimulation in the treatment of major depression. Biological Psychiatry. 2000 Feb 2000;47(4):332-7.

77. Bortolomasi M, Minelli A, Fuggetta G, et al. Long-lasting effects of high frequency repetitive transcranial magnetic stimulation in major depressed patients. Psychiatry Res. 2007 Mar 30;150(2):181-6.

78. Moller AL, Hjaltason O, Ivarsson O, et al. The effects of repetitive transcranial magnetic stimulation on depressive symptoms and the P(300) event-related potential. Nord J Psychiatry. 2006;60(4):282-5.

79. George MS, Wassermann EM, Kimbrell TA, et al. Mood improvement following daily left prefrontal repetitive transcranial magnetic stimulation in patients with depression: a placebo-controlled crossover trial. Am J Psychiatry. 1997 Dec;154(12):1752-6.

80. Rush AJ, Marangell LB, Sackeim HA, et al. Vagus nerve stimulation for treatment-resistant depression: a randomized, controlled acute phase trial. Biological Psychiatry. 2005 Sep 2005;58(5):347-54.

81. Harley R, Sprich S, Safren S, et al. Adaptation of dialectical behavior therapy skills training group for treatment-resistant depression. J Nerv Ment Dis. 2008 Feb;196(2):136-43.

82. Wiles NJ, Hollinghurst S, Mason V, et al. A randomized controlled trial of cognitive behavioural therapy as an adjunct to pharmacotherapy in primary care based patients with treatment resistant depression: A pilot study. Behavioural and Cognitive Psychotherapy. 2008 Jan 2008;36(1):21-33.

83. Avery DH, Holtzheimer PE, 3rd, Fawaz W, et al. Transcranial magnetic stimulation reduces pain in patients with major depression: a sham-controlled study. J Nerv Ment Dis. 2007 May;195(5):378-81.

84. Möller AL, Hjaltason Ó, Ívarsson Ó, et al. The effects of repetitive transcranial magnetic stimulation on depressive symptoms and the P-sub-3-sub-0-sub-0 event-related potential. Nordic Journal of Psychiatry. 2006 2006;60(4):282-5.

85. 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* 2007:739-52.

86. Moore RG, Blackburn I-M. Cognitive therapy in the treatment of non-responders to antidepressant medication: A controlled pilot study. Behavioural and Cognitive Psychotherapy. 1997 1997;25(3):251-9.

87. Bretlau LG, Lunde M, Lindberg L, et al. Repetitive transcranial magnetic stimulation (rTMS) in combination with escitalopram in patients with treatment-resistant major depression: a double-blind, randomised, sham-controlled trial. *Pharmacopsychiatry* 2008:41-7.

88. Thase ME, Kupfer DJ, Jarrett DB. Treatment of imipramine-resistant recurrent depression: I. An open clinical trial of adjunctive L-triiodothyronine. J Clin Psychiatry. 1989 Oct;50(10):385-8.

89. Fava M, Rush AJ, Wisniewski SR, et al. A comparison of mirtazapine and nortriptyline following two consecutive failed medication treatments for depressed outpatients: a STAR*D report. *Am J Psychiatry* 2006:1161-72.

90. Mazeh D, Shahal B, Aviv A, et al. A randomized, single-blind, comparison of venlafaxine with paroxetine in elderly patients suffering from resistant depression. *Int Clin Psychopharmacol* 2007:371-5.

91. Fredman B, d'Etienne J, Smith I, et al. Anesthesia for electroconvulsive therapy: effects of propofol and methohexital on seizure activity and recovery. Anesth Analg. 1994 Jul;79(1):75-9.

92. Little JT, Kimbrell TA, Wassermann EM, et al. Cognitive effects of 1- and 20-hertz repetitive transcranial magnetic stimulation in depression: preliminary report. Neuropsychiatry Neuropsychol Behav Neurol. 2000 Apr;13(2):119-24.

93. Corya SA, Williamson D, Sanger TM, et al. A randomized, double-blind comparison of olanzapine/fluoxetine combination, olanzapine, fluoxetine, and venlafaxine in treatment-resistant depression. *Depress Anxiety* 2006:364-72.

94. Shelton RC, Williamson DJ, Corya SA, et al. Olanzapine/fluoxetine combination for treatment-resistant depression: a controlled study of SSRI and nortriptyline resistance. *J Clin Psychiatry* 2005:1289-97.

95. Poirier MF, Boyer P. Venlafaxine and paroxetine in treatment-resistant depression. Double-blind, randomised comparison. *Br J Psychiatry* 1999:12-6.

96. McGrath PJ, Stewart JW, Fava M, et al. Tranylcypromine versus venlafaxine plus mirtazapine following three failed antidepressant medication trials for depression: a STAR*D report. *Am J Psychiatry* 2006:1531-41; quiz 666.

97. Nierenberg AA, Papakostas GI, Petersen T, et al. Lithium augmentation of nortriptyline for subjects resistant to multiple antidepressants. *Journal of Clinical Psychopharmacology*. US: Lippincott Williams & Wilkins 2003:92-5.

98. Thase ME, Corya SA, Osuntokun O, et al. A randomized, double-blind comparison of olanzapine/fluoxetine combination, olanzapine, and fluoxetine in treatment-resistant major depressive disorder. *Journal of Clinical Psychiatry*. US: Physicians Postgraduate Press 2007:224-36.

99. Shelton RC, Tollefson GD, Tohen M, et al. A novel augmentation strategy for treating resistant major depression. *Am J Psychiatry* 2001:131-4.

100. Janicak PG, Nahas Z, Lisanby SH, et al. Durability of acute tesponse to TMS in the treatment of major depression: relapse during a continuation pharmacotherapy extension study. Society for Biological Psychiatry Annual Meeting; 2007 May; San Diego, CA; 2007.

101. Solvason HB, Husain M, Fitzgerald PB, et al. TMS in the acute treatment of major depression: Improvments in functional status and quality of life. *Society fo Bilogical Psychiatry Annual Meeting*. San Diego, CA 2007.

102. Dannon PN, Dolberg OT, Schreiber S, et al. Three and six-month outcome following courses of either ECT or rTMS in a population of severely depressed individuals--preliminary report. Biol Psychiatry. 2002 Apr 15;51(8):687-90.

103. Grunhaus L, Dannon PN, Schreiber S, et al. Repetitive transcranial magnetic stimulation is as effective as electroconvulsive therapy in the treatment of nondelusional major depressive disorder: an open study. Biol Psychiatry. 2000 Feb 15;47(4):314-24.

104. Scott J, Palmer S, Paykel E, et al. Use of cognitive therapy for relapse prevention in chronic depression. Cost-effectiveness study. Br J Psychiatry. 2003 Mar;182:221-7.

105. Dannon PN, Schreiber S, Dolberg OT, et al. Transcranial magnetic stimulation is effective in the treatment of relapse of depression. International Journal of Psychiatry in Clinical Practice. 2000;4(3):223-6.

106. Schulze-Rauschenbach SC, Harms U, Schlaepfer TE, et al. Distinctive neurocognitive effects of repetitive transcranial magnetic stimulation and electroconvulsive therapy in major depression. Br J Psychiatry. 2005 May;186:410-6.

107. O'Connor M, Brenninkmeyer C, Morgan A, et al. Relative effects of repetitive transcranial magnetic stimulation and electroconvulsive therapy on mood and memory: a neurocognitive risk-benefit analysis. Cogn Behav Neurol. 2003 Jun;16(2):118-27.

108. Jorge RE, Moser DJ, Acion L, et al. Treatment of vascular depression using repetitive transcranial magnetic stimulation. Arch Gen Psychiatry. 2008 Mar;65(3):268-76.

109. Jorge RE, Robinson RG, Tateno A, et al. Repetitive transcranial magnetic stimulation as treatment of poststroke depression: a preliminary study. Biol Psychiatry. 2004 Feb 15;55(4):398-405.

110. Aas IH. Global assessment of functioning (GAF): properties and frontier of current knowledge. Ann Gen Psychiatry. 2010;9(1):20.

111. Leon AC, Solomon DA, Mueller TI, et al. The range of impaired functioning tool (LIFE-RIFT): a brief measure of functional impairment. Psychol Med. 1999;29(4):869-78.

112. Ware JE, Snow KK, Kolinski M, et al. SF-36 Health Survey manual and interpretation guide. Boston, MA: The Health Institute, New England Medical Centre 1993.

113. Wisniewski SR, Rush AJ, Bryan C, et al. Comparison of quality of life measures in a depressed population. The Journal of nervous and mental disease. 2007;195(3):219-25.

114. Endicott J, Nee J, Harrison W, et al. Quality of Life Enjoyment and Satisfaction Questionnaire: A New Measure. 1993.

115. Achard S, Chignon JM, Poirier-Littre MF, et al. Social adjustment and depression: value of the SAS-SR (Social Adjustment Scale Self-Report). Encephale. 1995;21(2):107-16.

116. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Medical Care. 1992;30(6):473-83.

117. Endicott J, Nee J, Harrison W, et al. Quality of Life Enjoyment and Satisfaction Questionnaire: a new measure. Psychopharmacol Bull. 1993;29(2):321-6.

118. Kind P, ed. The EuroQol instrument: An index of health-related quality of life. Quality of life and pharmacoEconomics in clinical trials. 2nd ed: Philadelphia Lippincott-Raven Publishers 1996.

119. Lisanby SH, Luber B, Schlaepfer TE, et al. Safety and feasibility of magnetic seizure therapy (MST) in major depression: randomized within-subject comparison with electroconvulsive therapy. Neuropsychopharmacology. 2003 Oct;28(10):1852-65.

120. White PF, Amos Q, Zhang Y, et al. Anesthetic considerations for magnetic seizure therapy: a novel therapy for severe depression. Anesth Analg. 2006 Jul;103(1):76-80, table of contents.

121. Kirov G, Ebmeier KP, Scott AI, et al. Quick recovery of orientation after magnetic seizure therapy for major depressive disorder. Br J Psychiatry. 2008 Aug;193(2):152-5.

122. Malone DA, Jr., Dougherty DD, Rezai AR, et al. Deep brain stimulation of the ventral capsule/ventral striatum for treatment-resistant depression. Biol Psychiatry. 2009;65(4):267-75.

123. McNeely HE, Mayberg HS, Lozano AM, et al. Neuropsychological impact of Cg25 deep brain stimulation for treatment-resistant depression: preliminary results over 12 months. J Nerv Ment Dis. 2008 May;196(5):405-10.

124. Mayberg HS, Lozano AM, Voon V, et al. Deep brain stimulation for treatment-resistant depression. Neuron. 2005 Mar 3;45(5):651-60.