

# *Draft Comparative Effectiveness Review*

---

Number xx

## **Menopausal Symptoms: Comparative Effectiveness Review of Therapies**

**Prepared for:**

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

**Contract No. xxx-xx-xxxx**

This information is distributed solely for the purposes of predissemination peer review. 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 (AHRQ) determination or policy.

**Prepared by:**

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

**Investigators:**

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

**AHRQ Publication No. xx-EHCxxx**  
**<Month Year>**

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

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

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.

This document is in the public domain and may be used and reprinted without special permission. Citation of the source is appreciated.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact [EffectiveHealthCare@ahrq.hhs.gov](mailto:EffectiveHealthCare@ahrq.hhs.gov).

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

**Suggested citation:** <Authors>. <Topic>. Evidence Report/Technology Assessment. No. <#>. (Prepared by <EPC Name> under Contract No. <##>.) Rockville, MD: Agency for Healthcare Research and Quality. <Month, Year>. <http://www.ahrq.gov/clinic/epcix.htm>. Accessed: <Date>. <URL>.

## Preface

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

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

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

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

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

Stephanie Chang, M.D., M.P.H.  
Director, EPC Program  
Center for Outcomes and Evidence  
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

William Lawrence, M.D., M.S.  
Task Order Officer  
Center for Outcomes and Evidence  
Agency for Healthcare Research and Quality

## Acknowledgments

### Key Informants

In designing the study questions, the EPC consulted several Key Informants who represent the end-users of research. The EPC sought the Key Informant input on the priority areas for research and synthesis. Key Informants are not involved in the analysis of the evidence or the writing of the report. Therefore, in the end, study questions, design, methodological approaches, and/or conclusions do not necessarily represent the views of individual Key Informants.

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

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

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

### Technical Expert Panel

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

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

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

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

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

## Peer Reviewers

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

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

The list of Peer Reviewers follows:

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

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

# Menopausal Symptoms: Comparative Effectiveness Review of Therapies

## Structured Abstract

**Objectives:** To systematically review and synthesize evidence evaluating the comparative effectiveness of treatments for menopausal symptom relief, along with potential long-term benefits and harms of those treatments.

**Data Sources:** The following electronic databases were searched through March 2012: MEDLINE®, EMBASE®, Cochrane Controlled Trials Register, and AMED Allied and Complementary Medicine. Grey literature searches included clinicaltrials.gov, the FDA website, and relevant conference abstracts.

**Review Methods:** Menopausal symptoms of interest included: vasomotor symptoms, quality of life, psychological symptoms, sexual function, urogenital atrophy, and sleep dysfunction. Randomized clinical trials provided the evidence base for symptom relief. Because outcomes were reported using varied measures, standardized effect measures were calculated to allow pooling. Network meta-analyses were performed when possible, along with pairwise comparisons. Long-term benefits and harms of interest included: breast, colon, endometrial, and ovarian cancer; coronary heart disease and venous thromboembolic events; gall bladder disease; and osteoporotic fractures. Systematic reviews, cohort, and case-control studies provided evidence.

**Results:** Evidence from 254 trials provided results for vasomotor symptoms (187 trials), quality of life (108 trials), psychological symptoms (90 trials), sexual function (76 trials), urogenital atrophy (63 trials), and sleep dysfunction (48 trials). The most commonly studied agents were estrogens, isoflavones, and SSRI/SNRIs. Estrogens of any dose without apparent difference between doses or mode of administration appeared the most effective relieving vasomotor symptoms and were accompanied by better quality of life scores. Improvements in depression, anxiety, and global measures of mental health were found. Estrogens administered vaginally diminished pain during sex and testosterone increased reported sexual activity. Measures of urogenital atrophy were most convincingly improved with vaginal estrogens. Estrogens also improved sleep, but the effect appeared modest. Over the long term, estrogen combined with progestin has both beneficial (fewer osteoporotic fractures) and harmful (increased risk of breast cancer, gall bladder disease, venous thromboembolic events, and stroke) effects; estrogens given alone do not appear to increase breast cancer risk. There is limited evidence on the long-term effects of most nonhormone treatments.

**Conclusions:** Women experiencing menopausal symptoms can consider a number of treatments of varying efficacy. There is considerable certainty that estrogens are most effective relieving common symptoms. Estrogens are accompanied by other potential long-term benefits and harms that require considering. Compared with estrogen, other agents have lesser efficacy and limited evidence on long-term benefits and harms.

# Contents

Executive Summary .....	ES-1
Introduction.....	1
Background.....	1
Menopausal Treatment Strategies.....	2
Objectives .....	5
Population(s), Interventions, Comparators, Outcomes, Timing, and Setting .....	5
Key Questions.....	6
Analytic Framework .....	8
Methods.....	9
Topic Refinement and Review Protocol.....	9
Literature Search Strategy.....	9
Grey Literature Search Strategy.....	10
Inclusion and Exclusion Criteria.....	10
Population(s), Interventions, Comparators, Outcomes .....	10
Study Designs—Inclusion and Exclusion Criteria .....	11
Key Question 1—Symptom Relief .....	11
Key Question 2—Other Benefits/Harms Hormones.....	12
Key Question 3—Nonhormone Other Benefits/Harms .....	13
Key Question 4—Subgroups .....	14
Key Question 1 and Key Question 4 Duplicate Populations .....	14
Study Selection Process .....	14
Data Extraction and Management.....	15
Key Question 1 and Key Question 4.....	15
Key Question 2 .....	16
Key Question 3 .....	16
Evidence Tables.....	16
Quality Assessment of Individual Studies .....	17
Randomized, Controlled Trials .....	17
Cohort Studies.....	18
Case Control Studies.....	18
Data Synthesis.....	18
Overall Approaches and Meta-Analyses for Direct Comparisons.....	18
Use of Standardized Effect Measures .....	19
Pooling .....	20
Minimal Clinically Important Differences.....	20
Indirect Comparisons with Mixed Treatment Comparisons Techniques .....	21

Outcome Measures.....	21
Strength of the Body of Evidence.....	22
Applicability .....	24
Peer Review and Public Commentary .....	24
Results.....	26
Overview.....	26
Agents .....	26
Results of Literature Searches .....	27
Key Question 1. Effectiveness of Different Treatments for Postmenopausal Symptoms .....	29
Description of Included Studies .....	29
Navigating Key Question 1 Results .....	29
Vasomotor Symptoms.....	31
Quality of Life.....	46
Psychological Symptoms .....	61
Sexual Function .....	79
Urogenital Atrophy .....	90
Sleep.....	100
Compounded Hormone Therapies .....	108
Key Question 2. Long-Term Effects of Hormone Therapy Preparations.....	108
Breast Cancer .....	109
Gall Bladder Disease.....	111
Colorectal Cancer.....	112
Coronary Heart Disease, Stroke, and Venous Thromboembolic Events .....	114
Endometrial Cancer .....	116
Osteoporotic Fractures .....	116
Ovarian Cancer .....	118
Strength of Evidence – Long-Term Effects of Hormone Therapy Preparations .....	118
Key Question 3. Nonhormone Other Benefits/Harms .....	119
Breast Cancer .....	121
Gall Bladder Disease.....	123
Colorectal Cancer.....	123
Coronary Heart Disease, Stroke, or Thromboembolism.....	124
Endometrial Cancer .....	125
Osteoporotic Fractures .....	125
Ovarian Cancer .....	127
Strength of Evidence – Nonhormone Other Benefits/Harms .....	127
Adverse Events .....	128

Key Question 4. Effectiveness of Treatments for Menopausal Symptoms in Selected Subgroups .....	128
Vasomotor symptoms .....	129
Sexual Function .....	132
Psychological Symptoms .....	133
Quality of Life.....	134
Sleep Dysfunction .....	135
Urogenital Atrophy .....	136
Discussion .....	138
Introduction.....	138
Symptom Relief .....	138
Vasomotor Symptoms.....	138
Quality of Life.....	139
Psychological Symptoms .....	139
Sexual Function .....	140
Urogenital Atrophy .....	141
Sleep.....	142
Limitations of the Evidence Base on Symptom Relief.....	142
Compounded Hormone Therapies .....	143
Other Benefits and Harms.....	144
Hormone Therapy Preparations .....	144
Nonhormone Therapy Preparations .....	145
Symptom Relief in Subgroups.....	146
Research Gaps.....	146
Implications for Clinical and Policy Decision-Making .....	147
Limitations of the Comparative Effectiveness Review Process .....	147
Conclusions.....	148
References.....	149
List of Abbreviations .....	165

#### Tables

Table 1. All therapies inclusion/exclusion criteria for the relief of vasomotor symptoms, sleep disturbance, psychological symptoms, urogenital atrophy, sexual dysfunction, and quality of life .....	12
Table 2. Nonhormone therapies trial inclusion/exclusion for agent-specific adverse events .....	14
Table 3. Nonhormone therapies trial inclusion/exclusion—coronary heart disease, stroke, or thromboembolism; gall bladder disease; osteoporotic fractures; or endometrial, breast, colorectal, or ovarian cancer.....	14
Table 4. NNTs according to SMD and control group response rates (e.g., achieving some minimal important improvement).....	20
Table 5. Minimal clinically important difference for various scales .....	21

Table 6. Strength of evidence rating domains. ....	23
Table 7. Agents and categorizations for purposes of review .....	26
Table 8. Comparison matrix example. Estimate B represents E-High (high-dose estrogen) compared with placebo; estimate A represents E-Standard (standard-dose estrogen) compared with an SSRI/SNRI .....	30
Table 9. Characteristics of trials assessing efficacy of treatment on vasomotor symptoms .....	33
Table 10. Vasomotor symptoms estimates of comparative efficacy as standardized effect sizes and 95 percent credible intervals from network meta-analysis. Treatments are ordered left to right from most to least efficacious. Highlighted effects are those where the credible interval does not overlap zero. The negative effects reflect improvement (lower on the symptom scale) for the agent on the left versus comparators to its right from intersecting treatments listed on the diagonal .....	37
Table 11. Approximate conversion of standardized effect sizes (SMDs) to reduction in daily hot flashes in trials assessing moderate to severe hot flashes .....	37
Table 12. Vasomotor symptoms rankings of comparative efficacy, standard deviations, and 95% credible intervals (integer values because they arise from a distribution of integers) .....	37
Table 13. Vasomotor symptoms pairwise effect estimates (pooled random effect estimates or single-trial effects if only data available) .....	38
Table 14. Trials comparing placebo with progestins, reporting vasomotor outcomes .....	40
Table 15. Trials comparing placebo with other prescription agents, reporting vasomotor outcomes .....	41
Table 16. Trials comparing estrogens of similar dose but by different route, reporting vasomotor outcomes .....	41
Table 17. Trials comparing nonprescription agents with placebo, reporting vasomotor outcomes .....	42
Table 18. Trials comparing estrogen with a nonprescription agent, reporting vasomotor outcomes .....	43
Table 19. Trials comparing nonprescription agents, reporting vasomotor outcomes .....	43
Table 20. Strength of evidence ratings domains for vasomotor symptoms <sup>a</sup> .....	44
Table 21. Characteristics of trials assessing efficacy for quality of life outcomes .....	47
Table 22. Quality-of-life estimates of comparative efficacy as standardized effect sizes and 95 percent credible intervals from network meta-analysis. Treatments are ordered left to right from most to least efficacious. Highlighted effects are those where the credible interval does not overlap zero. The effects reflect improvement (lower on the scale) for the agent on the left versus comparators to its right from intersecting treatments listed on the diagonal .....	52
Table 23. Quality-of-life rankings of comparative efficacy, standard deviations, and 95% credible intervals (integer values because they arise from a distribution of integers) .....	52
Table 24. Quality of life pairwise effect estimates (pooled random effect estimates or single trial effects if only data available) .....	53
Table 25. Trials comparing different modes of estrogen administration, reporting quality-of-life outcomes .....	55
Table 26. Trials comparing different doses of the same nonprescription treatment, reporting quality-of-life outcomes .....	56
Table 27. Trials comparing antidepressants, reporting quality-of-life outcomes .....	56
Table 28. Trials comparing nonprescription treatments, reporting quality-of-life outcomes .....	57
Table 29. Strength of evidence ratings domains for quality-of-life .....	59

Table 30. Characteristics of trials assessing efficacy for psychological symptoms .....	62
Table 31. Psychological outcomes pairwise effect estimates (pooled random effect estimates or single trial effects if only data available) .....	68
Table 32. Trials comparing different modes of estrogen administration reporting psychological outcomes .....	72
Table 33. Trials comparing estrogen with estrogen plus testosterone reporting psychological outcomes .....	72
Table 34. Trials comparing progestin alone with placebo reporting psychological outcomes .....	73
Table 35. Trials comparing prescription treatments with placebo reporting psychological outcomes .....	73
Table 36. Trials comparing nonprescription agents with placebo reporting psychological outcomes .....	74
Table 37. Trials comparing nonprescription agents with nonprescription agents reporting psychological outcomes .....	75
Table 38. Trials comparing antidepressants with antidepressants, reporting psychological outcomes .....	75
Table 39. Trials comparing antidepressants with nonprescription agents, reporting psychological outcomes .....	76
Table 40. Strength of evidence ratings domains for psychological symptoms .....	77
Table 41. Characteristics of trials assessing efficacy for sexual function .....	79
Table 42. Pooled effect sizes from single trials for improvement in sexual function .....	81
Table 43. Trials comparing different estrogen doses, reporting a global assessment of sexual function .....	84
Table 44. Trials comparing different estrogen modes of administration, reporting pain, interest, and global sexual function outcomes .....	85
Table 45. Trials comparing estrogens with testosterone, reporting sexual function outcomes ....	86
Table 46. Trials comparing estrogens with nonprescription agents reporting sexual function outcomes .....	86
Table 47. Trials comparing nonprescription agents with placebo, reporting sexual function outcomes .....	87
Table 48. Strength of evidence ratings domains for sexual function .....	88
Table 49. Characteristics of trials assessing efficacy for urogenital atrophy .....	90
Table 50. Pooled standardized effect sizes from trials for improvement in urogenital atrophy among vaginal estrogen doses. The estimate in each cell represents comparison of the treatment intersecting the diagonal above it to that on the right (n is number of trials pooled). For example, the pooled standardized effect size for standard dose estrogen versus placebo is -0.41. ....	92
Table 51. Pooled standardized effect sizes from single trials for improvement in urogenital atrophy among nonvaginal agents. The estimate in each cell represents comparison of the treatment intersecting the diagonal above it to that on the right (n is number of trials pooled). For example, the pooled standardized effect size for standard-dose estrogen versus placebo is -0.31 .....	93
Table 52. Hormone therapies compared, reporting urogenital atrophy outcomes .....	96
Table 53. Trials comparing different modes of estrogen administration, reporting urogenital atrophy symptoms .....	96
Table 54. Trials comparing nonprescription agents with placebo, reporting urogenital atrophy symptoms .....	97

Table 55. Trials comparing nonprescription agents, reporting urogenital atrophy outcomes .....	97
Table 56. Strength of evidence ratings domains for urogenital atrophy .....	98
Table 57. Characteristics of trials assessing efficacy for sleep dysfunction .....	100
Table 58. Pooled standardized effect sizes from trials for improvement in sleep dysfunction. The estimate in each cell represents comparison of the treatment intersecting the diagonal above it to that on the right (n=number of trials pooled). For example, the pooled standardized effect size for standard dose estrogen versus placebo is 0.25. ....	103
Table 59. Trials comparing similar estrogen doses, and reporting sleep outcomes .....	105
Table 60. Trial comparing hormone with a nonprescription agent, reporting sleep outcomes ..	105
Table 61. Trials comparing prescription treatments with placebo reporting sleep outcomes ....	105
Table 62. Trials comparing nonprescription agents with placebo and reporting sleep outcomes .....	106
Table 63. Trials comparing nonprescription agents, and reporting sleep outcomes .....	106
Table 64. Strength of evidence ratings domains for sleep outcomes .....	107
Table 65. Evidence base for long-term effects of hormone therapies .....	109
Table 66. Overall breast cancer incidence among women treated with estrogen/progestin .....	110
Table 67. Overall breast cancer incidence among women treated with estrogen alone .....	110
Table 68. Gall bladder disease incidence among women treated with estrogen/progestin .....	112
Table 69. Gall bladder disease incidence among women treated with estrogen alone .....	112
Table 70. Overall colorectal cancer incidence among women treated with estrogen/progestin. ....	113
Table 71. Overall colorectal cancer incidence among women treated with estrogen alone .....	113
Table 72. Coronary heart disease, stroke, and venous thromboembolic events incidence among women treated with estrogen/progestin .....	114
Table 73. Coronary heart disease, stroke, and venous thromboembolic events incidence among women treated with estrogen alone .....	114
Table 74. Overall endometrial cancer incidence among women treated with estrogen/progestin .....	116
Table 75. Osteoporotic fracture incidence among women treated with estrogen/progestin .....	117
Table 76. Osteoporotic fracture incidence among women treated with estrogen alone .....	117
Table 77. Ovarian cancer incidence among women treated with estrogen/progestin .....	118
Table 78. Strength of evidence assessment for long-term effects of hormone therapies <sup>a</sup> .....	118
Table 79. Evidence base for long-term effects of nonhormone therapies .....	119
Table 80. Study quality assessment for Key Question 3 RCTs .....	120
Table 81. Study quality assessment for Key Question 3 cohort studies .....	120
Table 82. Study quality assessment for Key Question 3 case control studies .....	121
Table 83. Nonhormone therapies and breast cancer .....	122
Table 84. Nonhormone therapies and colorectal cancer .....	123
Table 85. Nonhormone therapies and CHD, stroke, or thromboembolism .....	124
Table 86. Nonhormone therapies and osteoporotic fractures .....	126
Table 87. Nonhormone therapies and ovarian cancer .....	127
Table 88. Strength of evidence assessment for long-term effects of nonhormone therapies <sup>a</sup> ....	127
Table 89. Evidence base for subgroup analyses reporting vasomotor outcomes .....	129
Table 90. Quality assessment for studies reporting vasomotor subgroup analyses .....	129
Table 91. Evidence base for subgroup analyses reporting sexual function outcomes .....	132
Table 92. Quality assessment for studies reporting sexual function subgroup analyses .....	132
Table 93. Evidence base for subgroup analyses reporting psychological outcomes .....	133

Table 94. Quality assessment for studies reporting psychological symptoms subgroup analyses .....	133
Table 95. Evidence base for subgroup analyses reporting quality of life outcomes.....	134
Table 96. Quality assessment for studies reporting quality of life subgroup analyses .....	135
Table 97. Evidence base for subgroup analyses reporting sleep dysfunction outcomes .....	135
Table 98. Quality assessment for studies reporting sleep dysfunction subgroup analyses.....	136
Table 99. Evidence base for subgroup analyses reporting urogenital symptom outcomes .....	136
Table 100. Quality assessment for studies reporting urogenital symptoms subgroup analyses .	136
Table 101. Magnitude and strength of evidence of treatments for vasomotor symptoms; standardized effect sizes from pairwise comparisons .....	138
Table 102. Magnitude and strength of evidence of treatments for quality of life; standardized effect sizes from pairwise comparisons .....	139
Table 103. Magnitude and strength of evidence of treatments for psychological symptoms; standardized effect sizes from pairwise comparisons .....	140
Table 104. Magnitude and strength of evidence of treatments for sexual function; standardized effect sizes from pairwise comparisons .....	141
Table 105. Magnitude and strength of evidence of treatments for urogenital atrophy; standardized effect sizes from pairwise comparisons .....	142
Table 106. Magnitude and strength of evidence of treatments for sleep; standardized effect sizes from pairwise comparisons .....	142
Table 107. Long-term effects of hormone therapy preparations summarized.....	144
Table 108. Long-term effects of nonhormone therapy preparations summarized.....	145

## Figures

Figure 1. Analytic framework.....	8
Figure 2. PRISMA diagram .....	28
Figure 3. Network of comparisons included in vasomotor analyses — line thickness (and circle area) proportional to the number of comparisons .....	34
Figure 4. Caterpillar plot displaying all vasomotor symptoms comparisons included in the network analysis and 95% credible intervals.....	39
Figure 5. Network of comparisons included in quality of life analyses—line thickness and circle area are proportional to the number of comparisons .....	49
Figure 6. Caterpillar plot displaying all quality-of-life comparisons included in the network analysis and 95% credible intervals.....	54
Figure 7. Caterpillar plot displaying all depression comparisons and 95 percent confidence intervals.....	70
Figure 8. Caterpillar plot displaying all anxiety comparisons and 95 percent confidence intervals .....	70
Figure 9. Caterpillar plot displaying all global mental health comparisons and 95 percent confidence intervals .....	71
Figure 10. Caterpillar plot for sexual function: pain, global, and interest—standardized effects and 95 percent confidence intervals.....	82
Figure 11. Caterpillar plot sexual function: satisfying sexual episodes with testosterone compared with placebo—mean difference and 95 percent confidence intervals .....	82

Figure 12. Urogenital atrophy life caterpillar plot displaying pooled comparisons and 95 percent confidence intervals .....	94
Figure 13. Caterpillar plot displaying all pooled sleep comparisons and 95% confidence intervals .....	104

**Appendixes**

Appendix A. Literature Search Strategy	
Appendix B. List of Excluded Studies	
Appendix C. Abstraction Forms	
Appendix D. Estrogen Dosing Categories	
Appendix E. Evidence Base for Key Question 1	
Appendix F. Vasomotor Symptom Plots	
Appendix G. Quality of Life Plots	
Appendix H. Psychological Symptoms Plots	
Appendix I. Sexual Function Plots	
Appendix J. Urogenital Atrophy Plots	
Appendix K. Sleep Plots	
Appendix L. Quality Assessments from Nelson Report	
Appendix M. Short-Term Adverse Effects for Nonhormone Therapies	
Appendix N. Effectiveness of Treatments for Menopausal Symptoms in Selected Subgroups	

# Executive Summary

## Background

Menopause is defined as the permanent cessation of menstruation and ovulation. “Spontaneous” menopause occurs after 12 months of amenorrhea, as ovarian hormone secretion diminishes gradually, on average around the age of 51. Menopause may be induced prematurely through medical interventions such as surgery, chemotherapy, or radiation. In the United States, the number of women entering menopause each year is estimated to be approximately 2 million.<sup>1</sup>

Current terminology describing the stages of menopause was detailed in 1991 at the Stages of Reproductive Aging Workshop (STRAW).<sup>2-4</sup> The STRAW stages define the time from beginning of irregular menses through the first 12 months of amenorrhea as perimenopause and the period from the last menses to death as postmenopause; the first 5 postmenopausal years are defined as early postmenopause, which is followed by late postmenopause.<sup>2-4</sup>

During menopause, approximately 85 percent of women report experiencing symptoms of varying type and severity.<sup>5</sup> Types of symptoms experienced may include<sup>1</sup>:

- Vasomotor symptoms: Hot flushes are recurrent, transient episodes of intense heat on the face and upper body, sometimes followed by chills. These symptoms can occur while sleeping, producing intense perspiration. Individual hot flushes may last from one to five minutes. After irregular menses, vasomotor symptoms are the second most frequently reported perimenopausal symptoms.
- Sleep disturbances: Lengthy times to fall asleep, inability to sleep through the night, or inability to resume sleeping when woken prematurely, are signs of insomnia. Sleep apnea symptoms range from slight airflow reductions causing snoring, to periodic cessation of breathing.
- Psychological symptoms: Depression, anxiety, and mood disturbances may occur. Depressive symptoms can range from a depressed mood to clinical depression. A depressed mood may not require treatment, but if clinical depression is suspected, assessment and treatment are recommended. Symptoms of anxiety may include tension, nervousness, panic, and worry.
- Urogenital issues: Urinary incontinence and vaginal atrophy may occur. Vaginal atrophy involves vaginal walls that are thin, pale, dry, and sometimes inflamed. These changes cause discomfort and potential trauma during intercourse and pelvic examinations.
- Sexual function effects: Dyspareunia (pain during intercourse) and decreased libido are also reported by perimenopausal and postmenopausal women.

Longitudinal studies have shown that during early postmenopause, the prevalence of vasomotor symptoms among women ranges from 30 to 80 percent, depressed mood occurs in approximately one-third, and sleep disturbance in more than 40 percent.<sup>6-8</sup> Vasomotor symptoms generally begin 2 years before menopause, peak 1 year after menopause, and then diminish.<sup>9</sup> Differences in symptoms have been found among subpopulations of women. In the Penn Ovarian Aging Trial<sup>10</sup> and the Study of Women’s Health Across the Nation,<sup>11</sup> researchers report

differences in prevalence and duration of vasomotor symptoms among women depending on ethnicity and body mass index (BMI).

## Objectives

To systematically review and synthesize evidence evaluating the comparative effectiveness of treatments for menopausal symptoms, including potential benefits and harms other than symptom relief.

## Key Questions

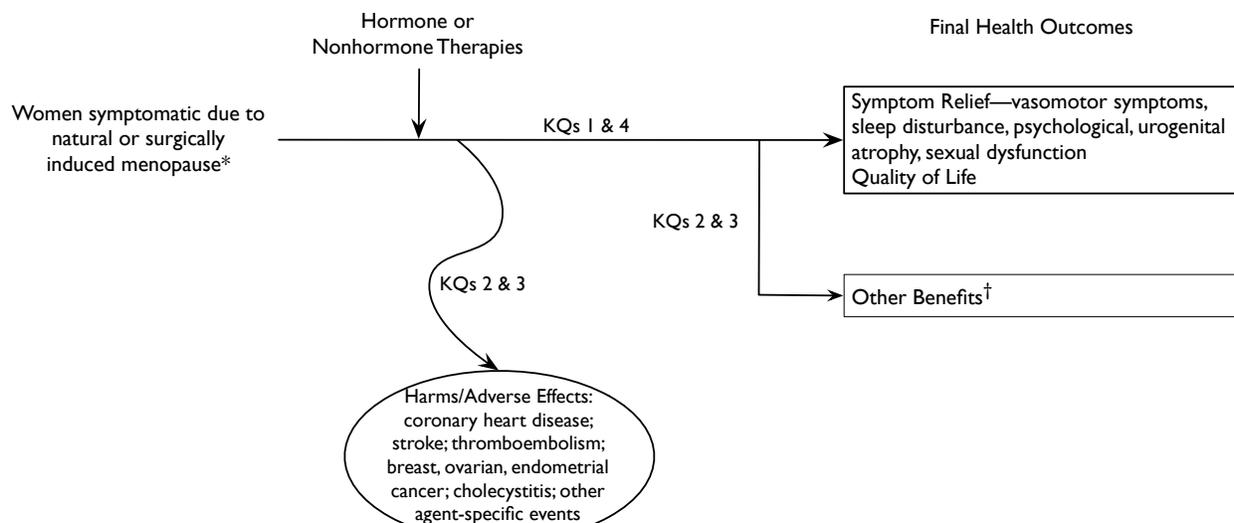
Key Question 1. What is the comparative effectiveness of different treatments for reducing symptoms of menopause (vasomotor symptoms, sleep disturbance, psychological symptoms, urogenital atrophy, and sexual dysfunction) and for improving quality of life? Individual agents will be compared to the extent permitted by the evidence.

Key Question 2. What are the effects of hormone therapy preparations on coronary heart disease, stroke, or thromboembolism; gall bladder disease; osteoporotic fractures; or endometrial, breast, colorectal, or ovarian cancers? Exposure will be examined according to duration of use and initiation relative to age and onset of menopause. (For women desiring contraception, combined estrogen-progestin and progestin-only contraceptives are included.)

Key Question 3. What are the effects of nonhormone therapy preparations on coronary heart disease, stroke, or thromboembolism; gall bladder disease; osteoporotic fractures; or endometrial, breast, colorectal, or ovarian cancer? Exposure will be examined according to duration of use and initiation relative to age and onset of menopause. What are the significant agent-specific harms/adverse effects of nonhormone therapies?

Key Question 4. Do effectiveness and adverse effects vary among subgroups of participants defined by demographics, symptom severity, other medications, and comorbidities or according to agent, preparation, or dose?

**Figure A. Analytic Framework**



KQ = key question

\*Excludes women with breast cancer or receiving tamoxifen

<sup>†</sup>Osteoporotic fractures and colorectal cancer

## **Methods**

### **Input from Stakeholders**

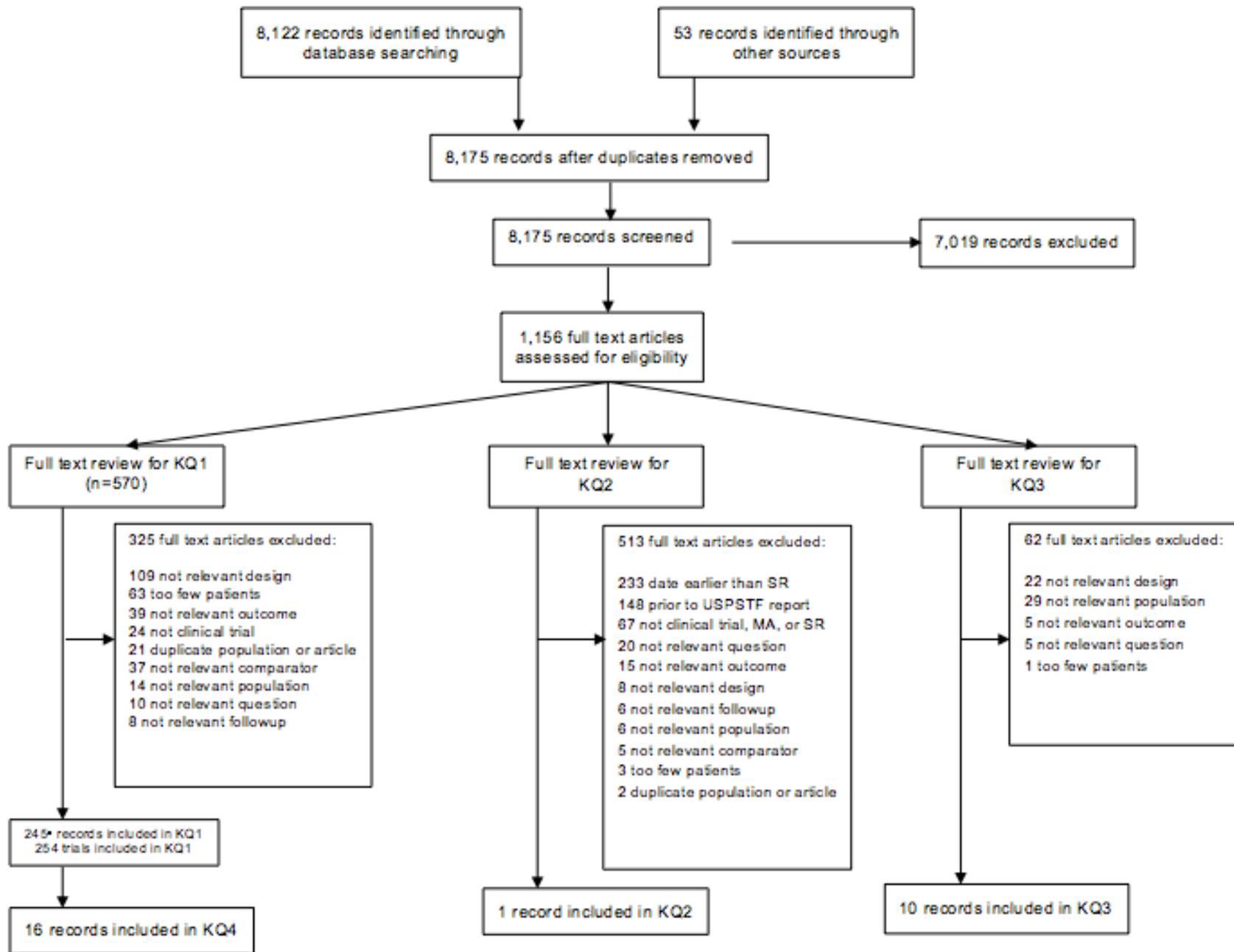
Input was sought from Key Informants representing clinicians (general medicine and gynecology), academicians, researchers, and patients during topic refinement. Key Questions were subsequently posted and public comment obtained. A technical expert panel was assembled including content and clinical experts. Comments were reviewed with appropriate changes to Key Questions made.

### **Data Sources and Selection**

The final literature search, including articles through March 2012, was run on MEDLINE®, EMBASE®, Cochrane Controlled Trials Register, and AMED Allied and Complementary Medicine. The reference lists for systematic reviews and meta-analyses were also screened to identify additional references. The grey literature search included extensive reviews of clinicaltrials.gov, the FDA Web site, and relevant conference abstracts. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram (Figure B) depicts the flow of search screening and study selection.

For KQ1, symptom relief from any therapy, randomized controlled trials (RCT) with 25 or more participants per arm, and a followup of 4 weeks or longer for centrally acting agent therapies and 12 weeks for all other therapies, were included. For KQ2, long-term effects of hormone therapies, systematic reviews and meta-analyses were included. For KQ3, a two part question looking at adverse events and long-term effects of nonhormone therapies, trials included in KQ1 which also reported adverse events were included, and randomized controlled

Figure B. PRISMA diagram



MA: meta-analysis; SR: systematic review

\*9 records presented results from two distinct patient populations and were divided into 2 trials each

trials and observational studies were included to assess long-term effects. For KQ4, subgroup analyses of symptom relief from any therapy, trials from KQ1 that reported subgroup analyses were included.

## **Data Abstraction and Quality Assessment**

### **Data Abstraction**

#### **Key Question 1 and Key Question 4**

Data were abstracted into tables created in DistillerSR. Two training sets of three articles each were abstracted by all team members. Results of each training set were reviewed to discuss any discrepancies in abstraction. Final data abstraction was performed by one team member, and verified by a different team member, with inconsistencies identified and resolved by consensus. The following data were abstracted:

- Trial Characteristics: author, year, country, number of trial sites, trial design, total number randomized, length of followup, intervention, uterine status, disclosures and conflicts of interest, funding, primary and secondary outcomes
- Trial Arm Characteristics: participant information such as number, age, ethnicity, BMI, time since menopause, tobacco use; treatment specifics such as type of treatment, dosage, dosage category, and mode of administration
- Outcomes: scale; results from baseline, 12-weeks, and final assessments; mean scores, mean changes, percent reductions, standard deviations, 95 percent confidence intervals, pre/post intervention comparisons, and between group comparisons

When only graphical outcomes were presented, figures were digitized. For KQ1, standardized effect sizes were calculated from available estimates of treatment effects, variances, and p-values.

#### **Key Question 2**

Data abstracted from the systematic reviews and meta analyses include the following: included trials, treatment type, treatment dose, length of followup, and results.

#### **Key Question 3**

Summary tables of long-term effects of nonhormone therapies contained the following information: condition, treatment, study design, study descriptions, and results.

Agent-specific adverse events for nonhormone therapies were categorized using a system recommended by the International Federation of Pharmaceutical Manufacturers and Associations.<sup>12</sup> The following data were abstracted for each category: author, year, country, treatment, dose, trial size, total adverse events, and percentage of events.

### **Quality Assessment**

In adherence with the EPC Program *Methods Guide for Effectiveness and Comparative Effectiveness Reviews* (hereafter *Methods Guide*),<sup>13</sup> the general approach to grading trials was performed by applying the criteria of the U.S. Preventive Services Task Force (USPSTF).<sup>14</sup> Discordant assessments were resolved with input from a third reviewer.

Study quality of RCTs was assessed by: assembly of comparable groups; blinding of researchers and subjects; concealment of group assignment; maintenance of comparable groups; differential loss to followup; equal and reliable measurements; clearly defined interventions; important outcomes considered and defined; and intention to treat analysis.

Study quality of cohort studies was assessed by: assembly of comparable groups; maintenance of comparable groups; differential loss to followup; equal and reliable measurements; important outcomes considered and defined; and statistical adjustment for potential confounders.

Study quality of case control studies was assessed by: accurate ascertainment of cases; nonbiased selection of cases and controls; response rate; equal application of diagnostic tests; accurate and equal measure of exposure; and attention to potential confounders.

## Data Synthesis and Analysis

For Key Question 1, trials employed a variety of outcome instruments. Standardized effect measures were calculated and pooled according to the EPC Program *Methods Guide*.<sup>15</sup> Calculating the standardized mean difference (SMD), which is  $(\text{effect1} - \text{effect2}) / \text{standard deviation}$ , allows for comparison of results across studies using different measures. Analyses were performed in R<sup>16</sup> using the meta,<sup>17</sup> compute.es,<sup>18</sup> and ggplot2<sup>19</sup> packages. Clinical heterogeneity, and appropriateness for pooling, was judged on the basis of study characteristics in concert with subject matter knowledge. Because the goal of any pooling is to estimate unconditional effects,<sup>20</sup> random-effects models were used. The magnitude of statistical heterogeneity was examined by using tau<sup>2</sup> owing to limitations of the I<sup>2</sup> metric and because between-trial variances are more intuitively interpreted on the effect estimate scale.<sup>21</sup> Evidence of possible publication bias were explored using funnel plots.

For vasomotor symptoms and QoL outcomes, network meta-analyses were performed including the most relevant comparisons with sufficient data. Network meta-analysis formally allows quantitative indirect comparisons. The random-effects network meta-analysis was performed pooling standardized effects in a Bayesian model described by Chaimani (<http://www.mtm.uoi.gr/3.continuousmodeldescription.pdf>). Models were fitted in OpenBUGS 3.2.2 using noninformative priors and convergence assessed using the Brooks-Gelman-Rubin plot and statistic (no value exceeded 1.002 in the model), autocorrelation and history plots. A burn-in of 10,000 samples was discarded and subsequent 50,000 analyzed. Rankings were estimated for the probability a treatment was most effective, next most effective, and so on. Effect estimates and accompanying 95 percent credible intervals were obtained from the samples. To evaluate consistency we compared available pairwise estimates to the network results.<sup>22</sup> We examined all pairwise comparisons individually in random effects models and graphically using forest funnel plots.

## Strength of the Body of Evidence

Strength of evidence (SOE) assessments were based on the Evidence-based Practice Center (EPC) approach,<sup>23</sup> which is conceptually similar to the GRADE system.<sup>24</sup> Two reviewers graded the strength of evidence, resolving disagreements by consensus. Details for the strength of evidence approach are also available at the AHRQ Effective Healthcare site, [http://effectivehealthcare.ahrq.gov/repFiles/2009\\_0805\\_grading.pdf](http://effectivehealthcare.ahrq.gov/repFiles/2009_0805_grading.pdf).

We adopted a point-based approach to SOE ratings. Each rating started at high (3 points) and was downgraded by one point each for: high risk of bias, inconsistent or unknown consistency,

imprecise or unknown precision, indirect body of evidence, and suspected reporting bias. Domain ratings were entered into a spreadsheet which provided a summary SOE. If the summary SOE remained 3 with no downgrades, strength of evidence was rated high; if the summary SOE equaled 2, strength of evidence was rated moderate; if the summary SOE equaled 1, strength of evidence was rated low; if the summary SOE was zero or lower, strength of evidence was rated insufficient. Following AHRQ guidance for assessing evidence on equivalence and non-inferiority, studies can be appropriately considered individually in the presence of clinical heterogeneity—“the lack of meta-analysis does not necessarily preclude a conclusion of EQ-NI [Equivalence-noninferiority], just as it does not preclude an evaluation of the strength of evidence in relation to a particular outcome.”<sup>25</sup>

## Results

Results are presented below for symptom relief (KQ1), other benefits and harms (KQ2 and KQ3), and symptom relief among subgroups (KQ4). Following these results, are discussions on research gaps, implications for clinical policy and decision making, limitations of the comparative effectiveness review process, and conclusions.

## Symptom Relief

Summary results are presented by outcome (vasomotor symptoms, quality of life, psychological symptoms, sexual function, urogenital atrophy, and sleep disturbances), followed by a brief discussion of compounded hormone therapies and limitations of the evidence base for symptom relief. Investigators utilized many different measurement rating scales to evaluate treatment effects. Pooling across scales can be accomplished only by standardized mean differences (SMD). Although enabling pooling, standardized effects pose challenges for clinical interpretation. To place their magnitudes into context, with control-group event rates of 20 to 60 percent SMDs can be expressed as approximate odds ratios. For example, SMDs and corresponding odds ratios (in parentheses) are as follows: SMD -0.2 (OR 0.7), -0.3 (0.6), -0.4 (0.5), 0.3 (2), 0.6 (3), and 0.75 (4). Although the odds ratios exceed relative risks when placebo group event rates exceed 10 percent, they provide a rough guide to the relative effect. For example, a typical placebo response rate of women with vasomotor symptoms is approximately 25 percent.

## Vasomotor Symptoms

A large body of evidence was identified comparing the efficacy of agents with placebo and other active treatments for the relief of vasomotor symptoms (Table A). Trials were most numerous for estrogens, isoflavones, SSRI/SNRIs, gabapentin or pregabalin, ginseng, and black cohosh. Estrogens of any dose appeared more effective than any other comparator without apparent difference between doses or mode of administration. Few differences were apparent in the network meta-analysis among SSRI/SNRIs, isoflavones, gabapentin/pregabalin, and black cohosh. Whether ginseng might have any effect is unclear. A host of other agents have been studied, but evidence is limited to single trials.

The efficacy of estrogens in treating vasomotor symptoms is well established. The comparative effectiveness of other agents relative to estrogens has been less clear. Albeit limited by the trial quality, the findings here show that other agents can ameliorate vasomotor symptoms,

but none have estrogen’s effectiveness. Conclusions concerning relative effectiveness can also be drawn.

**Table A. Magnitude and strength of evidence of treatments for vasomotor symptoms; standardized effect sizes from pairwise comparisons**

Comparisons	Comparators	Effect Size (SMD) (95% CI)	Strength of Evidence
9	Estrogen High vs. Placebo	-0.72 (-0.99 to -0.44)	High
36	Estrogen Standard vs. Placebo	-0.79 (-0.92 to -0.66)	
46	Estrogen Low/Ultralow vs. Placebo	-0.70 (-0.83 to -0.58)	
13	Estrogen High vs. Standard	-0.15 (-0.40 to 0.09)	High
7	Estrogen High vs. Low/Ultralow	-0.16 (-0.39 to 0.07)	
21	Estrogen Standard vs. Low/Ultralow	-0.10 (-0.22 to 0.02)	
10	SSRI/SNRI vs. Placebo	-0.40 (-0.54 to -0.26)	High
29	Isoflavones vs. Placebo	-0.41 (-0.58 to -0.25)	Moderate
4	Gabapentin/Pregabalin vs. Placebo	-0.33 (-0.45 to -0.22)	Moderate
3	Black Cohosh vs. Placebo	-0.26 (-0.43 to -0.09)	Low
3	Ginseng vs. Placebo	-0.41 (-0.83 to 0.02)	Low
8	Estrogen mode a vs. mode b	Not estimated	Moderate

## Quality of Life

Trials evaluating numerous agents reported some quality of life metric, but the evidence base included more than a single trial for estrogens, isoflavones, SSRI/SNRIs, ginseng, black cohosh, and DHEA (Table B). Compared with placebo, improved quality of life scores accompanied estrogens with standardized effect sizes exceeding 0.40 with moderate or high strength of evidence; effect sizes for all other agents were lesser in magnitude or low SOE. Similarly, estrogens ranked highest in the network comparison. For estrogens, there was no apparent difference in effect according to mode of administration. Quality of life scores were reported from trials of many nonprescription agents, but results from single trials do not allow conclusions concerning effects.

We found improved global quality of life scores in women taking estrogens. Two of the larger trials, “Women’s International Study of long Duration Oestrogen after The Menopause” (WISDOM)<sup>26</sup> and WHI,<sup>27, 28</sup> report no effect of estrogens on quality of life, but this is likely attributable to older age and less symptom severity of enrolled women in these trials. For the larger body of comparisons in women receiving estrogens, despite between-trial variability, results were more consistent. The general pattern of comparative efficacy seen with quality of life scores paralleled results for other vasomotor and other symptoms.

**Table B. Magnitude and strength of evidence of treatments for quality of life; standardized effect sizes from pairwise comparisons**

Comparisons	Comparators	Effect Size (SMD) (95 % CI)	Strength of Evidence
4	Estrogen High vs. Placebo	0.70 (0.40 to 1.01)	High
21	Estrogen Standard vs. Placebo	0.63 (0.47 to 0.78)	
12	Estrogen Low/Ultralow vs. Placebo	0.40 (0.24 to 0.56)	
6	Estrogen Standard vs. High	0.07 (-0.05 to 0.18)	High
8	Estrogen Standard vs. Low/Ultralow	0.13 (-0.04 to 0.29)	
2	Estrogen High vs. Low/Ultralow	-0.10 (-0.29 to 0.09)	
5	SSRI/SNRI vs. Placebo	0.27 (0.18 to 0.36)	

18	Isoflavones vs. Placebo	0.17 (0.06 to 0.29)	Moderate
3	Black Cohosh vs. Placebo	0.40 (0.18 to 0.63)	Moderate
3	Ginseng vs. Placebo	0.19 (0.01 to 0.36)	Low
3	DHEA vs. Placebo	Not estimated	Insufficient
7	Estrogen mode a vs. mode b	Not estimated	Moderate

## Psychological Symptoms

Just over one third of trials examining symptom treatment reported a psychological outcome—depression, anxiety, and global mental health—and often more than one. Only half specified some psychological symptom as a primary outcome. Overall, the samples were not selected to represent populations with clinical depression or anxiety. Compared with placebo, standardized effect sizes were in general not large (i.e., SMD between -0.5 and 0) for any of the agents studied for any psychological domain (Table C). Furthermore, the strength of evidence was at least moderate only for some effects of estrogens and SSRI/SNRIs.

An increased risk for depression, in the absence of prior depressive illness, during the menopausal transition has been described<sup>29</sup> and may be associated with vasomotor symptoms.<sup>30</sup> The effects assessed here may provide guidance when menopausal women are experiencing psychological symptoms.

**Table C. Magnitude and strength of evidence of treatments for psychological symptoms; standardized effect sizes from pairwise comparisons**

Comparisons	Comparators	Effect Size (SMD) (95% CI)	Strength of Evidence
<b>Depression</b>			
4	Estrogen High vs. Placebo	-0.64 (-0.94 to -0.33)	Moderate
11	Estrogen Standard vs. Placebo	-0.19 (-0.31 to -0.07)	Moderate
3	Estrogen Low/Ultralow vs. Placebo	-0.04 (-0.41 to 0.31)	Insufficient
8	Isoflavones vs. Placebo	-0.41 (-0.69 to -0.13)	Low
3	SSRI/SNRI vs. Placebo	-0.40 (-0.59 to -0.22)	Moderate
<b>Anxiety</b>			
2	Estrogen High vs. Placebo	-0.35 (-0.58 to 0.13)	Low
8	Estrogen Standard vs. Placebo	-0.16 (-0.34 to 0.03)	Insufficient
3	Estrogen Low/Ultralow vs. Placebo	-0.19 (-0.41 to 0.02)	Low
7	Isoflavones vs. Placebo	-0.53 (-0.87 to -0.23)	Low
2	SSRI/SNRI vs. Placebo	-0.31 (-0.53 to -0.08)	Low
<b>Global</b>			
9	Estrogen Standard vs. Placebo	-0.03 (-0.10 to 0.04)	Insufficient
7	Estrogen Low/Ultralow vs. Placebo	-0.24 (-0.45 to -0.02)	High
6	Isoflavones vs. Placebo	-0.12 (-0.26 to 0.01)	Low
4	SSRI/SNRI vs. Placebo	-0.39 (-0.63 to -0.15)	Moderate
2	Gabapentin/Pregabalin vs. Placebo	-0.22 (-0.46 to 0.03)	Insufficient

## Sexual Function

Some measure of sexual function was reported in less than a third of trials; half of those trials specified the outcome as primary (Table D). Outcomes were reported in four domains: pain (dyspareunia), a global metric, activity, and interest. Vaginal estrogens improved pain most convincingly (high strength of evidence), while lower pain scores with oral estrogens were less

certain (low strength of evidence). There was a modest improvement in global measures with estrogens. No agent appeared to enhance measures of interest. Sexually satisfying episodes were more frequent with testosterone compared with placebo—slightly more than one extra episode reported every 4 weeks (strength of evidence moderate). Overall, these results are generally consistent with evidence-informed expert clinical opinion.<sup>1</sup>

The Prevalence of Female Sexual Problems Associated with Distress and Determinants of Treatment Seeking (PRESIDE)<sup>31</sup> estimated approximately 15 percent of women aged 45 to 64 experienced some form of sexual distress. We identified one quantitative review on sexual outcomes during menopause, which included literature published between 1972 and 1992.<sup>32</sup> The effect of estrogen therapy on all four sexual function domains combined (108 studies total) yielded a standardized effect of -0.67 (SD 1.23), which was somewhat larger in magnitude than that obtained in this review.

**Table D. Magnitude and strength of evidence of treatments for sexual function; standardized effect sizes from pairwise comparisons**

Comparisons	Comparators	Effect Size (SMD) (95% CI)	Strength of Evidence
<b>Pain (lower is better)</b>			
10	Vaginally applied estrogens vs. placebo	-0.50 (-0.71 to -0.29)	High
3	Oral estrogens vs. Placebo	-0.44 (-1.05 to 0.17)	Low
13	All estrogens vs. Placebo	-0.49 (-0.69 to -0.29)	High
<b>Global (higher is better)</b>			
10	All estrogens vs. Placebo	0.28 (0.16 to 0.41)	High
2	SNRIs vs. Placebo	0.11 (0.02 to 0.19)	Insufficient
<b>Interest (higher is better)</b>			
3	All estrogens vs. Placebo	0.43 (-0.02 to 0.89)	Insufficient
3	Isoflavones vs. Placebo	0.31 (-0.24 to 0.86)	Insufficient
<b>Pain, Interest, Global</b>			
10	Estrogen Mode a vs. Mode b	Not Estimated	Moderate
<b>Activity</b>		<b>SSE/4 weeks</b>	
4	Testosterone, no women with intact uteri/ovaries	1.05 (0.64 to 1.45)	Moderate
4	Testosterone, women with/without uteri/ovaries	1.31 (0.89 to 1.72)	
8	Testosterone, all trials	1.17 (0.88 to 1.46)	

SSE Satisfying sexual episodes

## Urogenital Atrophy

One-quarter of trials reported urogenital atrophy outcomes—a primary outcome in 60 percent. Although multiple scales were employed, the strength of evidence was high that either oral or vaginal estrogens improve symptoms with standardized effect sizes for vaginal estrogens approximately twice that of nonvaginal estrogens (Table E). The strength of evidence was low for other agents (isoflavones and black cohosh).

The conclusions here are similar to those provided to clinicians<sup>1</sup> when considering treating symptoms that may be experienced by as many as 40 percent of postmenopausal women.<sup>33</sup> A

2006 Cochrane review including 19 trials concluded that vaginal or oral estrogens were equally effective for treating vaginal atrophy.<sup>34</sup> These results indicate, albeit indirectly based on placebo comparisons, that a greater magnitude of effect for vaginal compared with oral administration.

**Table E. Magnitude and strength of evidence of treatments for urogenital atrophy; standardized effect sizes from pairwise comparisons**

Comparisons	Comparators	Effect Size (SMD) (95% CI)	Strength of Evidence
11	Estrogen vaginal vs. placebo	-0.60 (-0.86 to -0.33)	High
16	Nonvaginal estrogen vs. placebo	-0.31 (-0.40 to -0.22)	High
5	Isoflavones vs. placebo	-0.57 (-0.90 to -0.24)	Low
2	Black Cohosh vs. placebo	-0.27 (-0.44 to -0.11)	Low
7	Estrogen mode a vs. mode b	not estimated	Low

## Sleep

Many trials ascertained self-reported sleep outcomes, but only a single trial examined a drug FDA-approved for use in insomnia (eszopiclone) that was highly effective. On a standardized effect scale, sleep improved with eszopiclone approximately three-fold greater than with estrogens or any other agent. This suggests that modestly improved sleep accompanies other agents, including estrogens, used to treat menopausal symptoms (Table F).

While sleep disturbances during menopause are common,<sup>35</sup> how often they are secondary to menopausal symptoms is not well defined. Sedative hypnotic agents are not generally used to treat menopausal symptoms and so were not represented in the trials identified. Reported improvements in sleep evident with other agents such as estrogens is possibly due to treatment of vasomotor symptoms, but requires evidence not considered here.

**Table F. Magnitude and strength of evidence of treatments for sleep; standardized effect sizes from pairwise comparisons**

Comparisons	Comparators	Effect Size (SMD) (95%CI)	Strength of Evidence
25	Estrogen vs. placebo	0.36 (0.26 to 0.46)	High
3	Estrogen vs. estrogen different dose	-0.25 (-0.67 to 0.18)	Insufficient
6	Isoflavones vs. placebo	0.35 (-0.43 to 1.13)	Insufficient
2	Ginseng vs. placebo	Not Estimated	Insufficient

## Compounded Hormone Therapies

Compounded hormone therapies are commonly prescribed, often in combination with some testing for hormone levels, with effectively no direct evidence base. We identified a single randomized, controlled trial examining pharmacokinetics in 40 women studied for 16 days.<sup>36</sup> No studies were identified examining the safety of the compounding practices for hormone therapies.

## Limitations of the Evidence on Symptom Relief

The body of evidence synthesized for Key Question 1 was large with many trials rated poor quality. However, the challenges of synthesizing this evidence extends far beyond trial quality to limitations incompletely incorporated in strength of evidence assessments. These include:

- Use of different outcome scales or metrics

- Necessity of calculating standardized effect sizes and inherent difficulties estimating from publications
- Potential differences in populations represented by trial samples
- Potential for selective outcome reporting

Calculating standardized effect sizes is not without challenges. There were a number of ways to obtain effect sizes from the continuous measures reported; trials typically did not report a between-group difference and variance (standard deviation) allowing the most straightforward calculating of standardized effects. To avoid excluding trial results, other calculations were required using p-values that were not reported exactly or simply nonsignificant. Where results were pooled, excluding nonsignificant results lacking a p-value would introduce bias. While imputation allowed including those results, it introduces uncertainty. Fortunately, the number of p-values requiring imputation was small.

A separate issue is that while trial populations included women experiencing menopause, they were differences in mean age, length of follow-up, and symptom severities. While the initial intent was to examine subgroups according to characteristics such as the presence of a uterus, lack of reporting did not allow doing so. Results then apply to average women across all trials.

It is also difficult to evaluate potential selective outcome reporting from the included trials. Vasomotor symptoms were reported in about three-quarters of trials, but all other outcomes in fewer than half. While some trials, such as those of sexual function or vaginal atrophy, were clearly not designed to primarily assess all outcomes, insignificant results may have gone unreported. For some of the outcomes reported, in only half was the outcome reported as primary. Results do not allow assessing whether effects on different outcomes are independent.

## Other Benefits and Harms

Summary results are presented first for hormone therapy preparations, then for nonhormone therapy preparations, followed by a discussion of limitations of the evidence base for other benefits and harms.

## Hormone Therapy Preparations

Evidence included in the recent Nelson et al report for the USPSTF<sup>37</sup> was reviewed here with concordant conclusions. In the Nelson et al report, a majority of evidence was derived from WHI trials, representing an older population, but one which overlaps with the population for this review. Findings from large observational studies with younger populations were included to inform the discussion on applicability. Still, the picture of long-term effects emerges with reasonable clarity as summarized in Table G.

The USPSTF review reported differences in event rates with estrogen/progestin or estrogen compared with placebo. However, extrapolating absolute rates from the WHI samples to the target population of this review is potentially problematic. In broad relative terms, gall bladder disease is the most frequent occurrence with thromboembolic events, stroke, and breast cancer less frequent. While less frequent they are not insignificant.

**Table G. Long-term effects of hormone therapy preparations summarized**

Outcome	Risk	Treatment vs. Placebo	Strength of Evidence	Comment
Breast Cancer	↑	Estrogen/Progestin	High	

	↓	Estrogen	Low	Inconsistent
Gall bladder disease	↑	Estrogen/Progestin	Moderate	Consistency unknown with 1 trial
	↑	Estrogen	Moderate	Consistency unknown with 1 trial
Venous Thromboembolic Events	↑	Estrogen/Progestin	Moderate	Consistency unknown with 1 trial
	↑	Estrogen	High	
Stroke	↑	Estrogen/Progestin	Moderate	Consistency unknown with 1 trial
	↑	Estrogen	High	
Ovarian Cancer	↑	Estrogen/Progestin	Low	Consistency unknown with 1 trial; imprecise with few cases
Colorectal Cancer	—	Estrogen/Progestin	Low	Consistency unknown with 1 trial; imprecise with wide CI
	—	Estrogen	Moderate	Consistency unknown with 1 trial
CHD	—	Estrogen/Progestin	Low	Consistency unknown with 1 trial; imprecise with wide CI
	—	Estrogen	Moderate	Consistency unknown with 1 trial
Endometrial Cancer	—	Estrogen/Progestin	Low	Inconsistent and imprecise
Osteoporotic Fractures	↓	Estrogen/Progestin	Moderate	Inconsistency between 2 trials
	↓	Estrogen	Moderate	Consistency unknown with 1 trial

Risk: ↑ increased, ↓ decrease, — no change;  
CI: confidence interval

## Nonhormone Therapy Preparations

The evidence base informing other potential benefits and harms of nonhormone therapies is limited, but does not suggest harmful long-term effects are likely for those agents studied (Table H). We identified large trials examining vitamin E, small trials of isoflavones, and observational studies evaluating antidepressants that did not always distinguish risks for the classes of agents used to treat symptoms (SSRI/SNRI). While no salient benefits were identified, neither were safety signals apparent. However, given the large numbers of women potentially taking these agents some caution is advised particularly for nonprescription agents. For example, the possibility of increased mortality with high-dose vitamin E has been raised.<sup>38</sup> Additionally, case reports of hepatotoxicity with black cohosh have been published.<sup>39</sup> This association has been debated,<sup>40</sup> but surveillance for adverse effects of nonprescription agents is generally inadequate. Safety data are also needed for the broad array of herbs and botanicals used to treat menopausal symptoms.

**Table H. Long-term effects of nonhormone therapy preparations summarized**

Outcome	Risk	Treatment vs. Placebo	Strength of Evidence
Breast cancer	—	Vitamin E	High
Breast cancer	—	SSRI/SNRI	Low
Colorectal Cancer	—	Vitamin E	High
Cardiovascular Events	—	Vitamin E	High
Cardiovascular Death	↓	Vitamin E	Low
Osteoporotic Fractures	↑	SSRI	Low
Osteoporotic Fractures	↓	Isoflavones	Insufficient
Ovarian Cancer	—	Vitamin E	Low
Breast, Endometrial, Ovarian Cancer	↑	Any Antidepressant	Insufficient

Risk: ↑ increased, ↓ decrease, — no change

## Limitations of the Evidence Base on Other Benefits and Harms

One limitation of the evidence base concerning long-term outcomes of hormone therapies derives from necessity to rely on results of randomized, controlled trials. There are well-described discrepant conclusions concerning these associations between observational studies and randomized, controlled trials.<sup>41</sup> The discrepancies have been attributed to two primary reasons—selection bias and time-varying confounding.<sup>42-44</sup> While the association with cardiovascular outcomes has been most scrutinized, difficulties assessing causal effects of hormone therapy from observational data appear to extend to other outcomes, including hip fractures<sup>42</sup> and colorectal cancer.<sup>44 4044</sup> As noted throughout, trials have been conducted from a target population overlapping with the one for this review, creating some challenges for assessing applicability.

There are several limitations to the evidence base of nonhormone therapies to consider. Many studies included women of all ages and, therefore, were excluded unless subgroup analyses on older women or menopausal women were specified. Much of the research available on the long-term effects of isoflavones and vitamin E consisted of population-based dietary studies and, therefore, did not meet inclusion criteria. Intermediate outcomes were reported in many of the studies, for example, bone density rather than osteoporotic fractures, and cholesterol rather than cardiovascular events. Finally, in studies that included all women rather than focusing on menopausal women, it was difficult to discern if exposure (to antidepressants, isoflavones) occurred during menopausal years.

## Symptom Relief in Subgroups

A small subset of trials identified for Key Question 1 reported subgroup analyses on symptom relief: 10 for hormone therapies, two nonhormone prescription therapies, and four nonprescription therapies. No subgroup analyses could be pooled, as no two trials had the same comparators, definitions of subgroups, and outcomes. The sparse evidence did not allow rating strength of evidence.

## Research Gaps

The principal gaps in the evidence on symptom relief include the following: safety data on nonprescription agents, lack of evidence on compounded hormone therapies, potential for predicting treatment response, and independence of some treatment outcomes:

- A large number of nonprescription agents were studied in individual trials. These agents are unregulated and safety data may be limited or absent. As women may elect to try these agents, those data need to be available.
- Millions of women use compounded hormone treatments. Yet there is a stark absence of evidence concerning the safety of compounded hormone therapies, and the diagnostic methods often accompanying their use.
- While the efficacy of estrogen treatment for symptom relief is so substantial that identifying some predictors of response would unlikely be fruitful. However, for the less-efficacious interventions, identifying predictors could be helpful for women having reasons to forgo hormone treatments.

The most important previous gaps in the evidence concerning long-term effects of hormone therapies have been filled. For some nonhormone therapies, with reasonable certainty (i.e.,

moderate or greater strength of evidence) significant safety issues have not been apparent; the same cannot be said for the entirety of the nonprescription agents.

Finally, estrogen therapy has the greatest efficacy relieving most symptoms and is accompanied by other potentially important benefits as well as some tradeoffs in the form of harms (varying according to whether combined with progestin). Given the number of outcomes to consider with different exposure effects (e.g., duration of use); the overall risk-benefit calculus is not simple. Juxtaposing evidence concerning symptom relief (as obtained here) with models for the long-term benefits and harms<sup>45</sup> according to patient characteristics (i.e., lower risk of hip fracture in blacks) could facilitate informed decisions by women and health care providers.

## **Implications for Clinical and Policy Decision-Making**

The implications of the conclusions from this review for clinical decision-making are straightforward. The results provide a guide to comparative efficacy alongside potential long-term benefits and harms; all are weighed in clinical decisions. For vasomotor symptoms and quality of life, the review provides clinicians with a simple ranked efficacy comparison for the most commonly used treatments.

## **Limitations of the Comparative Effectiveness Review Process**

This review was a large undertaking with many complexities. These included the variable manner in which trials reported results, multiple trial arms and multiple treatments, along with the goal of not excluding results for any a priori potentially arbitrary reason. Obtaining standardized effects can be challenging.<sup>46</sup> There are multiple ways to obtain an effect measure and standard deviation for each trial arm; different approaches may not yield identical results. Furthermore, given multiple trial arms and multiple outcomes, the number of calculations required was substantial. Confidence intervals and strength of evidence ratings do not incorporate this analytical uncertainty. Pooled estimates should be interpreted with this understanding.

Analyses of the multiple treatments required imposing some classification scheme that has limitations. For example, the estrogen dose categorization scheme did not consider progestin, or distinguish between combined and sequential progestin administration. Progestin use was problematic to distinguish because trials may have not given to women without a uterus, yet reported an effect for the entire sample.

Finally, interpreting network and pairwise meta-analyses deserves comment. In the pairwise meta-analysis, the randomized comparison is entirely preserved when pooling. Underlying the network of comparisons is an assumed exchangeability (similarity) of patient samples or the population from which they were drawn. All enrolled women were menopausal or perimenopausal, but there were some differences in the samples as noted in the review. Despite this, the closeness of almost all the network and pairwise estimates argues any discrepancies likely small.

## **Conclusions**

Women experiencing symptoms of menopause can consider a number of potential treatments of varying efficacy. From a large body of evidence, there is considerable certainty that estrogens are most effective relieving the common symptoms. Estrogens are accompanied by other

potential long-term benefits and harms that require considering. Compared with estrogen, other agents have lesser efficacy and limited evidence on long-term benefits and harms.

# Introduction

## Background

Menopause is defined as the permanent cessation of menstruation and ovulation. After 12 months of amenorrhea without pathological etiology, menopause is considered “natural” or “spontaneous.” Menopause can also be induced through medical interventions such as surgery, chemotherapy, or radiation. It occurs naturally between the ages of 42 and 58<sup>2-4</sup> and is a consequence of reproductive senescence. The average age at onset appears fixed, as it has been unchanged since ancient Greece.<sup>47</sup> In the United States, the number of women entering menopause (approximately 2 million per year<sup>1</sup>) will remain generally stable or even decline as baby boomers age. But given the continued improvement in life expectancy at age 50, the number of menopausal years will increase both for individual women and the population as a whole.

Current terminology describing the stages of menopause was detailed in 1991 at the Stages of Reproductive Aging Workshop (STRAW).<sup>2-4</sup> The STRAW stages define the time from beginning of irregular menses through the first 12 months of amenorrhea as perimenopause and the period from the last menses to death as postmenopause;<sup>2-4</sup> the first 5 postmenopausal years are defined as early postmenopause, which is followed by late postmenopause.

During menopause, approximately 85 percent of women report experiencing symptoms of varying type and severity.<sup>5</sup> Types of symptoms experienced include<sup>1</sup>:

- Vasomotor symptoms are recurrent, transient episodes of flushing, with intense heat on the face and upper body, sometimes followed by chills. These symptoms can occur while sleeping and can produce intense perspiration (night sweats). Individual hot flushes may last from one to five minutes. After irregular menses, vasomotor symptoms are the second most frequently reported perimenopausal symptom.
- Increases in sleep disturbances such as insomnia and sleep apnea/hypopnea may occur. Insomnia includes lengthy times to fall asleep, inability to sleep through the night, or inability to resume sleeping when woken prematurely. Sleep apnea symptoms range from slight airflow reductions which cause snoring, to periodic cessation of breathing (apnea).
- Psychological symptoms such as depression, anxiety, and mood disturbances may also occur in perimenopausal and postmenopausal women. The term “depression” may include a depressed mood or an intense adjustment reaction to a life event which may not require treatment. The term may also include clinical depression. If clinical depression is suspected, assessment and treatment are recommended. Symptoms of anxiety may include tension, nervousness, panic, and worry.
- Urogenital issues such as urinary incontinence and vaginal atrophy may occur. Vaginal atrophy describes vaginal walls that are thin, pale, dry, and sometimes inflamed. These changes cause discomfort and potential trauma during intercourse and during pelvic examinations.
- Sexual function effects such as dyspareunia (pain during intercourse) and decreased libido are also reported by perimenopausal and postmenopausal women.

Longitudinal studies have shown that during the early postmenopausal period the prevalence of vasomotor symptoms among women ranges from 30 to 80 percent, depressed mood occurs in approximately one-third, and sleep disturbance in more than 40 percent; diminished sexual function and vaginal dryness are also common.<sup>6-8</sup> A natural history of symptoms has been described, including the presence, severity, and time since menopause. For example, vasomotor symptoms generally begin 2 years before menopause, peak 1 year after menopause, and then diminish over the next 10 years.<sup>9</sup> However, differences in symptoms have been found among different subpopulations of women. In the Penn Ovarian Aging Trial, moderate to severe vasomotor symptoms lasted a median of 10.2 years; black women experienced a longer median duration of vasomotor symptoms, while women with a high body mass index tended to have shorter symptom duration.<sup>10</sup> In the Study of Women's Health Across the Nation, the prevalence of vasomotor symptoms was greater among black and Hispanic women and women with a higher body mass index.<sup>11</sup>

## Menopausal Treatment Strategies

### Overview

Estrogens have been a mainstay for treating menopausal symptoms, but are surrounded by controversy. Estrogens were approved by the U.S. Food and Drug Administration (FDA) in 1942 for treating menopausal symptoms, and by 1947, the *Physician's Desk Reference* listed more than 50 estrogen preparations approved for treating menopausal symptoms. In 1995, an estimated 37 percent of women aged 50 years or older in the United States reported using hormone therapy (estrogen with or without progestin),<sup>48</sup> owing in part to the results of observational studies interpreted to support a protective effect for cardiovascular disease. The clinical landscape shifted abruptly in 2002 with the first results from the Women's Health Initiative (WHI), a randomized comparison of estrogen/progestin versus placebo. Not only was cardiovascular risk increased, but overall harms from the treatment exceeded benefits.<sup>49</sup> Although subsequent evaluation of the body of evidence has indicated that interpretations of the results are more complex,<sup>50</sup> particularly for the target population included in this review, the consequences for hormone therapy use in the United States remain uncertain.<sup>51</sup>

In addition to decreasing estrogen production in menopausal women, the decrease of androgen production is of concern. Androgens affect sexual desire, muscle mass and strength, body mass index, and adipose tissue distribution. Androgens may also affect energy and psychological health. Two major androgens in women are testosterone and dehydroepiandrosterone (DHEA). In women with naturally occurring menopause, there is not a sudden decrease in androgen production, but in women with surgical menopause, testosterone levels decrease by about 50%.<sup>1</sup> A Cochrane review has reported sufficient evidence to suggest that supplementing estrogen therapy or estrogen/progestin therapy with testosterone has a beneficial effect on menopausal women experiencing sexual dysfunction.<sup>52</sup> DHEA is available without prescription as a dietary supplement, and is therefore under limited regulation. The efficacy of DHEA supplements for the treatment of menopausal symptoms has not been established.

Generally prepared for the individual patient, compounding of hormone therapy combines several hormones and employs nonstandard routes of administration.<sup>53</sup> Compounded hormones are claimed to be biochemically similar or identical to endogenous hormones. Compounded preparations typically contain estriol and can have variable potency.<sup>54</sup> Growing interest in

compounded hormones is undisputed; evidence from surveys of pharmacists, practitioners, and patients suggests a growing market for and belief in their effectiveness.<sup>55, 56</sup> In 2003, approximately 30 million prescriptions for compounded products were filled.<sup>57</sup> The products are heavily marketed, currently a \$1 billion industry and growing.<sup>58</sup>

While hormone therapy can relieve menopausal symptoms, concerns about potential risks (especially cardiovascular disease, uterine and breast cancer) provide reason to consider other agents. Both nonhormone prescription medications and nonprescription agents including complementary and alternative medicine (CAM) therapies have been studied in comparison with hormone therapy or placebo. These studies focus primarily on the relief of vasomotor symptoms.<sup>59</sup> Nonhormone prescription therapies include selective serotonin reuptake inhibitors (SSRI) and serotonin norepinephrine reuptake inhibitors (SNRI), eszopiclone, clonidine, methyl dopa, gabapentin, and pregabalin; biologic CAM therapies include isoflavones, red clover (*Trifolium pratense*), black cohosh (*Cimicifuga racemosa*), St. John's wort (*Hypericum perforatum*), ginseng, flax seed, vitamin E, dong quai (*Angelica sinensis*), and dehydroepiandrosterone (DHEA). Postulated mechanisms for SSRIs and SNRIs include central effects on serotonin, dopamine, or norepinephrine,<sup>60</sup> while the potential benefit of isoflavones is thought to be mediated through their affinity for estrogen receptors. In the Study of Women's Health Across the Nation, depending on ethnicity, 20 to 70 percent of participants reported using some form of CAM therapy during the menopausal transition phase.<sup>61</sup>

## **Guidelines and Society Statements**

The principal uncertainty for nonhormone therapies is effectiveness, whereas for hormone therapies it is the balance of benefits and harms. In May 2012, the U.S. Preventive Services Task Force (USPSTF) issued an update to their 2005 guideline titled *Hormone Replacement Therapy for the Prevention of Chronic Conditions in Postmenopausal Women*, in which the use of hormones for the prevention of chronic conditions was not recommended. This updated systematic review included research published through November 2011, but the report did not consider treatment of menopausal symptoms.<sup>37</sup>

The 2010 North American Menopause Society (NAMS) position statement on hormone therapy concluded, "Recent data support the initiation of [hormone therapy] around the time of menopause to treat menopause-related symptoms; to treat or reduce the risk of certain disorders, such as osteoporosis or fractures in select postmenopausal women; or both. The benefit-risk ratio for menopausal [hormone therapy] is favorable for women who initiate [hormone therapy] close to menopause but decreases in older women and with [greater] time-since-menopause in previously untreated women."<sup>62</sup>

The 2007 International Menopause Society (IMS) recommendations state, "The safety of [hormone therapy] largely depends on age. Women younger than 60 years old should not be concerned about the safety profile of [hormone therapy]. New data and reanalyses of older studies by women's age show that, for most women, the potential benefits of hormone therapy given for a clear indication are many and the risks are few when initiated within a few years of menopause."<sup>63</sup> Neither the NAMS position statement nor the IMS recommendations were accompanied by systematic reviews, yet both express considerable certainty and are somewhat at odds with trends in hormone therapy use.<sup>51</sup>

The Endocrine Society recently performed an extensive review of evidence surrounding postmenopausal hormone therapy, published as a scientific statement.<sup>64</sup> Efforts to systematically review and synthesize the literature were described, although methods used in the review (e.g.,

search strategies and the process for rating evidence) were not detailed. Reviewers graded the quality of the evidence supporting use of menopausal hormone therapy as “high” for ameliorating vasomotor symptoms and vaginal atrophy, preventing bone loss, decreasing colon cancer risk, and increasing the risk of thromboembolism and gallbladder disease.

Position statements on compounded therapies have also been issued. The NAMS does not generally recommend compounded combined hormone therapy and suggests that compounded hormone products include a patient package insert identical to that required for products that have government approval. The NAMS states that “in the absence of efficacy and safety data for bioidentical [compounded] hormone therapy, the generalized benefit-risk ratio data of commercially available hormone therapy products should apply equally to bioidentical [compounded] hormone therapy.”<sup>53</sup> Analogous views are held by American College of Obstetricians and Gynecologists (ACOG), The Endocrine Society, and the American Association of Clinical Endocrinologists (AACE). ACOG states that in addition to having the same safety issues as those associated with FDA approved hormone therapy, compounded hormones may have additional risks intrinsic to compounding.<sup>65</sup> The FDA maintains that while pharmacists engaging in traditional compounding provide a valuable service, anyone receiving compounded hormones should discuss options with their health care provider to determine if compounded drugs are the best option for their medical needs.<sup>66</sup>

## **Challenges in Synthesizing the Evidence**

From the perspectives of systematic review and evidence synthesis, there are a number of challenges in comparing different hormone therapies and comparing those therapies to alternatives:

- **Population:** Trial populations vary by factors such as age, ethnicity, time since menopause, length of time on hormone replacement therapy, BMI, and uterine status. For example, in a single trial, women with and without a uterus may be offered different treatment regimens.
- **Intervention:** The array of hormone and nonhormone therapies is broad and includes a number of biologic CAM and prescription agents, making synthesis difficult. Hormone therapies vary by preparation, type, and administration route. Compounded hormones are not standardized.
- **Outcomes:** There are numerous categories of menopausal outcomes: psychological, vasomotor, sexual function, sleep disturbances, and overall quality of life. Each of these categories can be measured by a variety of standardized scales, making synthesis challenging. Also, these outcomes are self-reported, and individuals assess levels of importance and severity of symptoms differently.
- **Timing:** Some harms are not immediately evident (e.g., breast cancer), and some benefits are not immediately evident (prevention of osteoporosis and fractures). Long followup times are necessary to adequately determine benefits and harms from these therapies.

Two large-hormone replacement therapy trials exemplify the complexities described above when collecting evidence for a systematic review on this topic. The WHI, which is a primary evidence base for harms from hormone replacement therapy, had a treatment population that overlaps but differs from the target population in this review. The WHI hormone trials excluded women with severe menopausal symptoms and enrolled primarily women older than those

recently menopausal. These population characteristics of the WHI trials are relevant when attempting to interpret the results. A more recent report from the WHI observational trial<sup>67</sup> found women experiencing early vasomotor symptoms were at the lowest risk of cardiovascular disease and cardiovascular events. Another large trial with combined menopausal hormone therapy,<sup>26</sup> the Women's International Trial of Long Duration Oestrogen after Menopause [WISDOM], was prematurely closed because of the findings of the WHI trial, resulting in a trial with only 1 year of followup.

## **Objectives**

For an individual menopausal woman considering hormonal or nonhormonal therapies, the questions of interest are: Given the presence of menopausal symptoms, what is the balance of benefits and harms of these therapies? Does the timing and duration of these therapies affect the balance? Accordingly, the objectives of this review include: systematically reviewing and synthesizing evidence evaluating the comparative effectiveness of treatments for menopausal symptoms, potential benefits other than symptom relief, and potential harms.

## **Population(s), Interventions, Comparators, Outcomes, Timing, and Setting**

### **Population(s)**

Women experiencing symptoms accompanying natural menopause (during perimenopausal or postmenopausal periods) or surgically induced menopause (during the postmenopausal period).

### **Interventions**

Three categories of interventions are included in the report: hormone therapies, nonhormone prescription therapies, and nonprescription therapies:

- Hormone therapies including estrogen therapy and estrogen/progestin (or estrogen/androgen) therapy administered by oral, transdermal, nasal, or vaginal route; combined estrogen-progestin and progestin-only contraceptives; compounded menopausal hormone therapy, often referred to as “bioidentical hormones”
- Nonhormone prescription therapies including antidepressants (SSRIs and SNRIs), eszopiclone, clonidine, methyl dopa, gabapentin, and pregabalin
- Nonprescription therapies including isoflavones, red clover, black cohosh, St. John's wort, ginseng, flax seed, vitamin E, dong quai, and DHEA

### **Comparators**

Placebo or direct comparison between therapies, such as varying hormone dose and formulation.

### **Outcomes**

- For Key Question 1 (KQ1) and Key Question 4 (KQ4):
- Final outcomes are menopausal symptom-related:
  - Vasomotor symptoms
  - Sleep disturbance

- Psychological symptoms
- Urogenital atrophy
- Sexual function
- Quality of life
- For Key Question 2 (KQ2) and Key Question 3 (KQ3):
- Final outcomes are other benefits and harms:
  - Coronary heart disease
  - Stroke
  - Thromboembolism
  - Breast cancer
  - Endometrial cancer
  - Ovarian cancer
  - Colorectal cancer
  - Gall bladder disease
  - Osteoporotic fractures
  - Agent-specific adverse events

## Timing

For KQ1 and KQ4, at least 12 weeks of followup for adequate assessment of hormone and nonprescription treatment effects is required for inclusion. For centrally acting agents (SSRI, SNRI, gabapentin, and pregabalin) minimum trial duration will be 4 weeks. This is based on evidence that efficacy in treating vasomotor symptoms with these agents is demonstrable by 4 to 8 weeks—and translates into similar efficacy at 12 weeks.<sup>68</sup> For KQ2 and KQ3, longitudinal studies on colorectal cancer, breast cancer, and ovarian cancer require a followup of 5 years or greater for inclusion. Longitudinal studies on coronary heart disease, stroke, thromboembolism, endometrial cancer, gall bladder disease, and osteoporotic fractures require a followup of one year or greater for inclusion.

## Setting

Primary care and community settings

## Key Questions

Question 1. What is the comparative effectiveness of different treatments for reducing symptoms of menopause (vasomotor symptoms, sleep disturbance, psychological symptoms, urogenital atrophy, and sexual dysfunction) and for improving quality of life? Individual agents will be compared to the extent permitted by the evidence.

Treatments of interest include:

- Hormone therapies
  - Oral estrogen only or combined with progestin (or androgen)
  - Transdermal estrogen or combined with progestin
  - Vaginal estrogen

- Combined estrogen-progestin and progestin-only contraceptives (for women desiring contraception)
- Compounded menopausal hormone therapy

Evidence evaluating hormone therapies will be considered separately for women with and without a uterus. Women with breast cancer will be excluded.

- Nonhormone therapies
  - Prescription
    - Antidepressants—SSRIs and SNRIs
    - Eszopiclone
    - Clonidine
    - Methyldopa
    - Gabapentin/pregabalin
  - Nonprescription complementary and alternative therapies
    - Isoflavones, including red clover (*Trifolium pratense*)
    - Black cohosh (*Cimicifuga racemosa*)
    - St. John's wort (*Hypericum perforatum*)
    - Ginseng
    - Flax seed
    - Vitamin E
    - Dong quai (*Angelica sinensis*)
    - Dehydroepiandrosterone

Question 2. What are the effects of hormone therapy preparations on coronary heart disease, stroke, or thromboembolism; gall bladder disease; osteoporotic fractures; or endometrial, breast, colorectal, or ovarian cancers? Exposure will be examined according to duration of use and initiation relative to age and onset of menopause. (For women desiring contraception, combined estrogen-progestin and progestin-only contraceptives are included.)

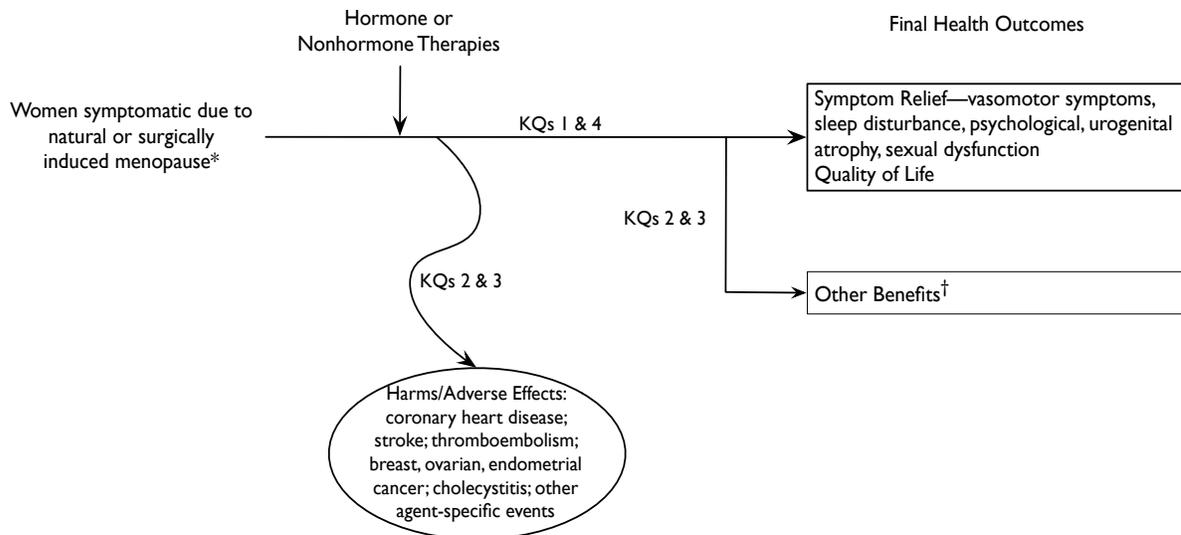
Question 3. What are the effects of nonhormone therapy preparations on coronary heart disease, stroke, or thromboembolism; gall bladder disease; osteoporotic fractures; or endometrial, breast, colorectal, or ovarian cancer? Exposure will be examined according to duration of use and initiation relative to age and onset of menopause. What are the significant agent-specific harms/adverse effects of nonhormone therapies?

Question 4. Does effectiveness and adverse effects vary among subgroups of participants defined by demographics, symptom severity, other medications, and comorbidities or according to agent, preparation, or dose?

## Analytic Framework

Figure 1 depicts the potential impact of both hormonal and nonhormonal treatments among women with menopausal symptoms. KQ1 and KQ4 illustrate how hormone and nonhormone therapies for menopausal symptoms may improve quality of life as well as reduce the occurrence or severity of the following symptoms: vasomotor symptoms, sleep disturbance, sexual dysfunction, urogenital atrophy, quality of life, and psychological symptoms. Other benefits of these treatments may include the prevention of osteoporotic fractures and colorectal cancer, as represented by the straight line of KQ2 and KQ3. The curved line of KQ2 and KQ3 represent potential consequential adverse effects among women using hormone and nonhormone therapies. These adverse effects include coronary heart disease, stroke, thromboembolism, breast cancer, endometrial cancer, ovarian cancer, and gall bladder disease.

Figure 1. Analytic framework



KQ = key question

\*Excludes women with breast cancer or receiving tamoxifen

†Osteoporotic fractures and colorectal cancer

## Methods

This comparative effectiveness review (CER) followed the methods suggested in the AHRQ Methods Guide for Effectiveness and Comparative Effectiveness Reviews (<http://www.effectivehealthcare.ahrq.gov/methodsguide.cfm>). Methods were applied as appropriate for the evidence available for each key question. For KQ1 and KQ4, evidence included randomized clinical trials. For KQ2, systematic reviews of randomized controlled trials were supplemented by observational studies when appropriate to assess applicability. Evidence sought for KQ3 included systematic reviews, meta-analyses, randomized clinical trials, and observational studies. The topic refinement process, literature search strategies, inclusion/exclusion criteria, data extraction and management procedures, evidence syntheses, and quality assessment methods are described below, specific to each key question.

### Topic Refinement and Review Protocol

The topic for this report was nominated in a public process. Input was sought from Key Informants representing clinicians (general medicine and gynecology), academicians, and patients during topic refinement. Key Questions were subsequently posted and public comment obtained. A Technical Expert Panel (TEP) was assembled including content and clinical experts. Public comments were reviewed along with input from Key Informants and the Technical Expert Panel (TEP) to assure that the questions are specific and explicit about what information is being reviewed. The key questions were posted for public comment and finalized by the EPC after review of the comments (<http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=1022>).

A review protocol was drafted by the EPC in consultation with the TEP and also posted for public comment. Comments were reviewed by the EPC, discussed with the TEP, and appropriate changes made to the protocol. The protocol was amended during the course of the review in two main respects. First, for KQ1 vasomotor symptom and quality of life outcomes, for the most common treatments a network meta-analysis was judged appropriate. Second, the USPSTF report<sup>37</sup> was released addressing KQ2 in its entirety, save issues of applicability. With the release of that report and the discrepant conclusions concerning associations observed from observational studies and randomized, controlled trials, evidence for effects was limited to randomized comparisons.

### Literature Search Strategy

Search strategies were developed (see Appendix A) by an expert librarian in collaboration with the trial team. No date limitations were applied. Only English-language articles were included.

The literature search was run on MEDLINE<sup>®</sup>, EMBASE<sup>®</sup>, Cochrane Controlled Trials Register, and AMED Allied and Complementary Medicine. The search included articles through March 2012. Duplicate records were deleted. The reference lists for systematic reviews and meta-analyses were screened to identify additional references which may not have been included in the original search. The search strings are provided in Appendix A. A single search strategy was used for all key questions, but different inclusion/exclusion criteria were used for the different key questions, details of which are outlined in the Inclusion and Exclusion Criteria section below.

## Grey Literature Search Strategy

Searches were performed in [clinicaltrials.gov](http://clinicaltrials.gov), the FDA Web site, and relevant conference abstracts (conferences identified by TEP members). Attempts to locate related publications were made and trial authors were contacted for unpublished results if two senior team members concurred that the evidence could impact results meaningfully (i.e., alter evidence GRADE). A text search for the following words was used to identify relevant conference abstracts: random, meta, systematic, testosterone, sertraline, citalo, fluoxetine, paroxetine, vilazodone, venlafax, eszopiclone, gaba, clonidine, methyl, mirt, myocardial, stroke, thromboembol, breast ca, endometrial ca, ovarian ca, colorectal ca, gall bladder disease, fracture.

References identified in the grey literature search were then screened using the same inclusion criteria as the original literature search and were incorporated into the review process when appropriate. Potentially unpublished evidence was also requested by the Scientific Resource Center from manufacturers.

Additional strategies were conducted to identify relevant literature on compounded or “bioidentical” hormone therapies. Based on the absence of clinical trials for compounded hormone therapy, specific position statements containing keywords: “compounded or bioidentical hormones” were identified, reviewed, and selected from the following professional societies:

North American Menopause Society<sup>53</sup>

American College of Obstetricians and Gynecologists<sup>65</sup>

The Endocrine Society<sup>69</sup>

American Association of Clinical Endocrinologists<sup>70</sup>

Special committee reports from the United States Senate<sup>71</sup> and U.S. FDA<sup>66</sup> were also identified for review. Finally, we reviewed an influential lay-press publication on bioidentical hormones to provide further perspective regarding the controversial topic of compounded hormone therapy.<sup>72</sup>

## Inclusion and Exclusion Criteria

### Population(s), Interventions, Comparators, Outcomes

- **Population(s)**

Women experiencing symptoms accompanying natural menopause (during perimenopausal or postmenopausal periods) or surgically induced menopause (during the postmenopausal period). **Exclusions:** women with breast cancer; trial populations that consisted of only participants with preexisting conditions such as fibromyalgia, rheumatoid arthritis, or cardiovascular disease.

- **Interventions**

Hormone therapy including estrogen therapy and estrogen-progestin (or estrogen-androgen) therapy administered by oral, transdermal, or vaginal route; combined estrogen-progestin and progestin-only contraceptives; compounded menopausal hormone therapy, often referred to as “bioidentical hormones” (Key Questions [KQs] 1 and 2). **Exclusions:** Women receiving tamoxifen.

Nonhormone therapies are listed under Key Question 1.

- **Comparators**

Placebo or direct comparison between therapies, including hormone dose and formulation.

- **Outcomes**

- No intermediate outcomes are included.
- Final outcomes - menopausal symptom-related:
  - Vasomotor symptoms
  - Sleep disturbance
  - Psychological symptoms
  - Urogenital atrophy
  - Sexual function
  - Quality of life
- Final outcomes - other benefits and harms:
  - Coronary heart disease
  - Stroke
  - Thromboembolism
  - Breast cancer
  - Endometrial cancer
  - Ovarian cancer
  - Colorectal cancer
  - Gall bladder disease
  - Osteoporotic fractures
  - Agent-specific adverse events

- **Timing**

For hormone and nonhormone therapies, exposure to treatment will be at least 12 weeks from the baseline assessment. For centrally acting agents such as SSRIs, SNRIs, gabapentin, and pregabalin, trial duration will be at least four weeks from baseline assessment.

- **Setting**

Primary care and community (biologic complementary and alternative therapies).

## **Study Designs—Inclusion and Exclusion Criteria**

### **Key Question 1—Symptom Relief**

We included randomized controlled trials (RCTs) with placebo or an active comparator. Anticipating sufficient RCTs for this key question, nonrandomized studies were not included. RCTs should have at least 25 patients randomized per arm who are studied for at least 12 weeks for hormone and nonhormone therapies, 4 weeks for centrally acting agents (SNRIs, SSRIs, gabapentin, pregabalin); these conditions are minimums consistent with trials used to define efficacy for vasomotor symptoms. Other meta-analyses and systematic reviews will not be included. Table 1 summarizes the inclusion and exclusion criteria.

**Table 1. All therapies inclusion/exclusion criteria for the relief of vasomotor symptoms, sleep disturbance, psychological symptoms, urogenital atrophy, sexual dysfunction, and quality of life**

Trial Design	Criteria
RCTs with placebo comparator or active comparator	Include <sup>a</sup>
Meta-analyses and systematic reviews	Exclude <sup>b</sup>
Observational studies	Exclude
Single arm/case series	Exclude
Case reports	Exclude
Minimum duration <sup>c</sup>	≥ 12 weeks
Sample size	≥25 participants randomized per arm

<sup>a</sup> Women with breast cancer are excluded.

<sup>b</sup> Bibliographies of meta-analyses and systematic reviews will be reviewed for any trials not identified in the literature search.

<sup>c</sup> Minimum duration for centrally acting agents such as SSRI, SNRI, gabapentin, and pregabalin, is 4 weeks. This is based on evidence that efficacy in treating vasomotor symptoms with these agents is demonstrable by 4 to 8 weeks.<sup>68,73</sup>

RCTs = randomized controlled trials

Several of the nonhormone therapies are consumed as part of a regular diet (soy, vitamin E, ginseng, for example) and are therefore often part of large population-based food consumption observational studies. For the purposes of this report, those studies were not included. Only studies in which the nonhormone therapies are treatments were included.

Therapies were required to be administered during the perimenopausal or menopausal years for study inclusion. If therapies were used only during the premenopausal years, those studies were excluded. If we were unable to determine if the nonhormone therapies were administered during the perimenopausal or menopausal years, for example studies reporting “ever” use, those studies were excluded.

## Key Question 2—Other Benefits/Harms Hormones

The associations of hormone therapies with the other benefits and harms considered here has been the subject of controversy, considerable research, and a motivation for conducting the WHI trials. Discrepant conclusions concerning these associations have been observed from observational studies and randomized controlled trials.<sup>41</sup> The discrepancies have been attributed to two primary reasons—selection bias and time-varying confounding.<sup>42-44</sup> While the association with cardiovascular outcomes has been most scrutinized, difficulties assessing causal effects of hormone therapy on the KQ2 outcomes from observational data appear to extend to other outcomes as well, including hip fractures<sup>42</sup> and colorectal cancer.<sup>44</sup> Relying on observational data employing standard analyses to examine these outcomes is problematic.<sup>43</sup> Accordingly, study selection to evaluate treatment effects (i.e., those causal) for KQ2 will be limited to systematic reviews of randomized, controlled trials.

SRs examining relevant outcomes (coronary heart disease, stroke, or thromboembolism; gall bladder disease; osteoporotic fractures; or endometrial, breast, colorectal, or ovarian cancer) will be considered if meeting the following criteria derived from the AMSTAR tool and AHRQ guidance: 1) at least two electronic sources were searched; key words and/or MeSH® terms stated; 2) study inclusion/exclusion criteria reported; 3) study quality (potential bias) of included trials assessed and documented. However, during the course of the CER, Nelson et al completed a review for the USPSTF on the effects of menopausal hormone therapy for chronic disease prevention<sup>37</sup> which met all criteria and addressed outcomes included in KQ2. Accordingly, it was used as the basis for KQ2.

It is important to note that the approach adopted was not primarily to appraise conclusions of the identified review, but to use the review to identify relevant trials meeting our inclusion criteria and appraise and synthesize evidence from them, including assigning a strength of evidence.

Given the natural history of osteoporosis, as well as breast, ovarian, and colorectal cancer, minimum trial duration of 5 years was established as an inclusion criterion for longitudinal studies investigating those outcomes. A minimum sample size of 250 women per trial was imposed to allow valid assessment of event rates. Outcomes were identified in consultation with the TEP to capture those most consequential. They were not intended to be an exhaustive list.

We anticipated evidence for KQ2 to ultimately derive in whole or in part from the WHI trials. These trials enrolled an older sample overlapping the target population of this CER.<sup>74</sup> Owing to this difference, applicability of evidence requires scrutiny. This step is in addition to those outlined in AHRQ guidance<sup>75</sup> (which notes “the exact process needs to be flexible and will likely evolve”) and adopted by the review team owing to the controversy surrounding applicability of WHI results to the CER target population. To assess applicability for KQ2 we examined our search to identify trials and observational studies enrolling peri- and recently menopausal women and consulted a clinical content expert. Informative studies were selected based on recommendations from the content expert in consultation with the review team. Results from these studies were included in the applicability discussion.

### **Key Question 3—Nonhormone Other Benefits/Harms**

For nonhormone prescription treatments, we limited our review to studies using the drugs to treat menopausal symptoms (and not for other indications for which the interventions may be commonly used) to increase the applicability of the review to the population of women with menopausal symptoms.

For nonhormone nonprescription treatments, any study design identifying agent-specific harms was included. Due to scope issues, we limited the list of included agents as prioritized in consultation with the TEP. The list is not exhaustive – see Key Question 1 for included agents.

The evidence base for agent-specific adverse events for nonhormone therapies consisted of articles included in Key Question 1 which also reported adverse events, as well as meta-analyses, systematic reviews, and observational studies. Reference lists in the systematic reviews and meta-analyses were reviewed, to identify randomized, controlled trials and observational studies meeting inclusion criteria (Table 2 and Table 3).

**Table 2. Nonhormone therapies trial inclusion/exclusion for agent-specific adverse events**

Trial Design	Prescription therapies	Nonprescription therapies
RCTs with placebo comparator or with active comparator	Include	Include
Meta-analyses and systematic reviews	Include	Include
Observational studies	Include	Include
Single arm and case series	Exclude	Include
Case reports	Exclude	Include
Minimum duration	≥12 weeks	None
Sample size	≥25 participants randomized per arm	None

RCTs = randomized controlled trials

**Table 3. Nonhormone therapies trial inclusion/exclusion—coronary heart disease, stroke, or thromboembolism; gall bladder disease; osteoporotic fractures; or endometrial, breast, colorectal, or ovarian cancer**

Trial Design	Prescription therapies	Nonprescription therapies
RCTs with placebo comparator or with active comparator	Include	Include
Meta-analyses and systematic reviews	Include	Include
Observational studies	Include	Include
Single arm and case series	Exclude	Exclude
Case reports	Exclude	Exclude
Minimum duration	5 years <sup>a</sup> 1 year <sup>b</sup>	5 years <sup>a</sup> 1 year <sup>b</sup>
Sample size	>250	>250

<sup>a</sup> Longitudinal studies of colorectal, breast, or ovarian cancers; and fracture outcomes (does not apply to case-control studies).

<sup>b</sup> All other outcomes (does not apply to case-control studies).

RCTs = randomized controlled trials

## Key Question 4—Subgroups

Subgroups (age, BMI, prior use of therapies, vasomotor severity of symptoms, time since menopause, uterine status, therapy schedule, comorbidities [smoking, anxiety, premenstrual syndrome or postnatal depression]) were selected from included trials in KQ1. Women with breast cancer were excluded.

## Key Question 1 and Key Question 4 Duplicate Populations

Duplicate populations already described in an included article not reporting additional outcomes of interest (KQ1 and KQ4) were excluded.

## Study Selection Process

Articles from the literature search were transferred into EndNote<sup>®</sup> (Thomson Reuters, New York, NY) and then into DistillerSR (Evidence Partners Inc., Manotick, ON, Canada) for trial selection. A pilot training set of 50 titles was screened by two team members. Titles alone did not provide sufficient screening information and the review proceeded with title/abstract screening. A set of 50 titles/abstracts was used to train the team members. In the title/abstract

screening phase, all references underwent dual review for inclusion in the full-text review. Disagreements were resolved by an independent team member.

Citations marked for inclusion during the title/abstract screening phase were retrieved for full text review. A dual screening process was conducted to determine inclusion/exclusion status from the full text. Disagreements between reviewers were resolved by an independent team member. Articles were excluded if they did not meet the criteria specific for each key question for appropriate trial design, minimum number of participants, and minimum length of followup. Reasons for exclusion were recorded in the DistillerSR database (Appendix B).

## **Data Extraction and Management**

### **Key Question 1 and Key Question 4**

Data were defined in a data dictionary and abstracted into tables created in DistillerSR (Appendix C). Two training sets of three articles each were abstracted by all team members. Meetings were held after each training set of articles was abstracted, to discuss potential abstraction discrepancies. The data dictionary and abstraction forms were modified based on input from all team members. After finalizing the data dictionary, abstraction forms, and abstraction instructions, data abstraction was conducted. Abstraction was performed by one team member, and verified by a different team member. Inconsistencies identified were resolved by consensus with publication review. For crossover trials only the first phase was included.

Included in abstracted data were the following (see data dictionary Appendix C) for complete listing):

- Trial Characteristics: author, year, country, number of trial sites, trial design, total number randomized, intervention, surgical or natural menopause, disclosures and conflicts of interest, funding, primary and secondary outcomes, and if required for inclusion into trial frequency or intensity of climacteric symptoms
- Trial Arm Characteristics: number of participants, age, ethnicity, BMI, time since menopause, tobacco use, and treatment specifics such as type of treatment, dosage, and mode of administration
- Outcomes: scale or measurement; results from baseline, 12-weeks, and final assessments; depending on how the results were reported, mean scores, mean changes, percent reductions, standard deviations, 95% confidence intervals, preintervention/postintervention comparisons, and between group comparisons.
- Many trials included in KQ1 reported outcomes using more than one scale or metric for each domain (with up to 7 arms per trial). For example, psychological symptoms reported may have included depression, anxiety, and a global measure; vasomotor symptoms may have been reported as frequency, severity, and with a menopausal symptom instrument. Selecting outcome metrics to abstract a priori could potentially introduce bias if one was chosen not uniformly or even commonly reported. In addition, data reported with one metric/scale for the same outcome might not provide sufficient quantitative data to estimate an effect while another did. Therefore, we abstracted (digitizing figures when necessary) up to 3 metrics/scales per KQ1 outcome from each trial.

Treatment dosages were recorded for all agents. For analytical purposes, estrogen doses were classified: ultralow, low, standard, and high. With oral treatments, the dosing category

definitions were based on those used in the 2009 Cochrane review on hormone replacement therapy and endometrial hyperplasia.<sup>76</sup> For example, dose categories for oral conjugated equine estrogens were: ultralow (0.15 to 0.3 mg); low (0.4 mg); standard (0.625 mg); and high (1.25 mg). For other modes of administration, such as transdermal and spray, dosing categorizations were established in consultation (i.e., primarily) with the clinical content expert. For a complete list of estrogen dose categories, by type of estrogen and mode of administration, refer to Appendix D.

When only graphical outcomes were presented, figures were digitized. Data were exported to and analyzed with R.<sup>16</sup> Data were abstracted into separate datasets. For example, we constructed two study level data sets: study characteristics and study quality ratings; a data set including characteristics for each study arm; and for KQ1 6 datasets or one for each outcome. With few exceptions, trial-level and summary evidence tables were created by manipulating, analyzing, and formatting data in R, then exporting to Microsoft<sup>®</sup> Excel. The only errors for tables produced in this manner are due to either abstraction or possible coding errors with transcription errors being eliminated.

## Key Question 2

Data from trials identified through the Nelson report for the USPSTF<sup>37</sup> were abstracted, including treatment type, treatment dose, length of followup, and results.

## Key Question 3

With a small literature base for the effect of nonhormone therapies on long-term conditions, quantitative synthesis was not possible. Descriptive summaries of the available evidence were generated. Summary tables were created and contained the following information: condition, treatment, trial design, trial descriptions, and results.

Adverse events reported for nonhormone therapies included a wide variety of symptoms. Events were categorized according to the International Federation of Pharmaceutical Manufacturers and Associations<sup>12</sup> recommended scheme: blood and lymphatic system; cardiac; congenital, familial, and genetic disorders; ear and labyrinth disorders; eye; endocrine disorders; gastrointestinal; general disorders and administration site conditions; hepatobiliary disorders; immune system disorders; infections and infestations; investigations; injury, poisoning, and procedural complications; metabolism/nutritional; musculoskeletal; neoplasms benign, malignant, and unspecified (including cysts and polyps); nervous system; psychiatric disorders; renal/urinary; respiratory, thoracic, and mediastinal disorders; skin and subcutaneous tissue; and vascular.

Data were abstracted into adverse events tables including: author, year, country, treatment, dose, trial population size, total adverse events, and percentage of events for each category.

## Evidence Tables

The body of evidence for KQ1 (and contributing to KQ4) was large including multiple comparators and trials reporting multiple outcomes. Following exploratory and descriptive analyses, we organized 7 sets of evidence tables according to 9 generally exclusive categories of comparators: 1) hormone vs. placebo; 2) SSRI/SNRI vs. placebo; 3) other prescription agents vs. placebo; 4) nonprescription agents vs. placebo; 5) hormone, nonprescription, placebo comparisons; 6) hormone vs. hormone; 7) nonprescription vs. hormone; 8) nonprescription vs.

nonprescription; and 9) SSRI/SNRI vs. nonprescription. Seven sets of evidence tables were generated including: 1) descriptive trial data; 2) patient age, body mass index, smoking history; 3) ethnicity/race; 4) uterine status, mean at menopause, years since menopause, prior hormone therapy; 5) outcomes reported; 6) treatment specifics including category, dose, route, generic and trade name, and estrogen dose if estrogen given (for each treatment arm); and 7) study quality elements and overall ratings. Only for the treatment specifics were trial arms specified which ranged from 2 to 6 (the single 7-arm trial footnoted). For each of the 63 tables, studies were ordered chronologically. These tables appear in Appendix E. (Sample R Code for constructing tables available from the authors.)

## Quality Assessment of Individual Studies

In adherence with the EPC Program *Methods Guide for Effectiveness and Comparative Effectiveness Reviews* (hereafter *Methods Guide*),<sup>13</sup> quality (bias) assessment was performed by applying the criteria of the U.S. Preventive Services Task Force.<sup>14</sup> An assessment was performed by two independent reviewers. Studies were given ratings of good, fair, and poor.<sup>14</sup> Discordant quality assessments were resolved with input from a third reviewer. A modified version of AMSTAR, a validated tool, was used for quality assessment of meta-analyses and systematic reviews.<sup>77</sup>

When interpreting study quality ratings, it is important to note the study design along with the rating. Features such as randomization and control arms in RCTs inherently reduce risk of bias, while observational studies generally have more sources of bias.<sup>78</sup> A “fair” rating for an RCT is not equivalent to a “fair” rating for an observational study. We therefore added the qualifier “observational study” next to the good, fair, and poor ratings in the quality assessment tables for the cohort and case control studies.

Even with appropriate analysis, the ability of observational studies to identify unconfounded associations and causal effects<sup>79</sup> or ascertain harms<sup>80</sup> can be highly variable. Moreover, all observational data are considered lesser (low) strength of evidence.<sup>24</sup> The perspective here is that a qualitative appraisal of observational studies that scrutinizes both the design and analytic approaches used to evaluate any causal effects is informative alongside a more quantitative one (i.e. checklist). For the more qualitative approach, we will adapt the method described by Thompson et al.<sup>81</sup>

## Randomized, Controlled Trials

The following criteria were used to assess the study quality of randomized, controlled trials: assembly of comparable groups; blinding of researchers and subjects; adequate concealment of group assignment; maintenance of comparable groups; differential loss to followup; equal and reliable measurements; clearly defined interventions; important outcomes considered and defined; and intention-to-treat analysis.

Based on these criteria, ratings for randomized, controlled trials were defined as:

Good: Meets all criteria; comparable groups are assembled and maintained throughout study (followup at least 80%); reliable and valid measurement instruments used and applied equally between groups; interventions clearly defined; important outcomes defined; and intention-to-treat analysis performed.

Fair: Generally comparable groups assembled initially, but questions remain about differences in followup; measurement instruments acceptable and generally applied equally;

some but not all important outcomes considered; some but not all potential confounders accounted for.

Poor: Groups assembled initially are not close to being comparable or maintained; unreliable or invalid measurement instruments used; key confounders are given little or no attention.

## **Cohort Studies**

The following criteria were used to assess the study quality of cohort studies: assembly of comparable groups; maintenance of comparable groups; differential loss to followup; equal and reliable measurements; important outcomes considered and defined; and statistical adjustment for potential confounders.

Based on these criteria, ratings for cohort studies were defined as:

Good: Meets all criteria; comparable groups are assembled and maintained throughout study (followup at least 80%); reliable and valid measurement instruments used and applied equally between groups; interventions clearly defined; important outcomes defined; and appropriate statistical adjustment for confounders.

Fair: Generally comparable groups assembled initially, but questions remain about differences in followup; measurement instruments acceptable and generally applied equally; some but not all important outcomes considered; some but not all potential confounders accounted for.

Poor: Groups assembled initially are not close to being comparable or maintained; unreliable or invalid measurement instruments used; key confounders are given little or no attention.

## **Case Control Studies**

The following criteria were used to assess study quality of case control studies: accurate ascertainment of cases; nonbiased selection of cases and controls; response rate; equal application of diagnostic tests to each group; accurate and equal measure of exposure to each group; and attention to potential confounders.

Based on these criteria, ratings for case control studies were defined as:

Good: Appropriate ascertainment of cases and nonbiased selection of controls; response rate  $\geq 80\%$ ; diagnostic procedures and measurements accurate and applied equally; and appropriate attention to potential confounders.

Fair: No major selection or diagnostic bias among groups; response rate  $<80\%$ ; attention to some but not all potential confounders.

Poor: Major selection or diagnostic biases; response rate  $<50\%$ ; inaccurate or unequal exposure measurements; or inattention to potential confounders.

## **Data Synthesis**

### **Overall Approaches and Meta-Analyses for Direct Comparisons**

The approach adopted for evidence synthesis was inclusive to incorporate as much evidence as possible. The rationale for this approach has four primary underpinnings. First, while symptom severity varies, the experience of menopause is universal. Second, defining homogeneous populations of women within the evidence base of trials identified is potentially problematic due to varying patient characteristics, as well as reporting. For example, years since menopause was reported in 32.3 percent of trials. Thirdly, trials employed a variety of different

patient-reported outcome instruments; while some are more commonly used than others there are clearly few standards. To apply an inclusion criteria stipulating use of particular instrument(s) could arguably introduce bias. Lastly, combining outcomes obtained on different metrics requires calculating standardized effect measures. Obtaining effects and some estimate of variance from trials reported in a myriad of ways is challenging. For example, as outlined below, outcomes can be reported in a host of different ways, each allowing calculation of an effect and variance. Excluding trials reporting a nonsignificant result from a pooled analysis would introduce bias and requires imputation. Further, in the end, one must also consider potential reporting bias. There are therefore numerous potential sources of uncertainty over and above those typically encountered in meta-analyses. While reported and even discussed, the p-values accompanying pooled estimates should be considered cautiously as their calculation does not incorporate some sources of statistical uncertainty; arguable most all should be penalized and a lower level of type I error applied than is convention. For example, normality of outcome metrics cannot be completely verified. For vasomotor symptoms we examined qq plots according to metric which supported normality for most, but confirming for those metrics used in a few trials was not possible. Additionally, while data extraction was verified and each reverified for potential outliers (SMDs >1.0 or < -1.0) in preliminary analyses, data extraction for use in SMDs is difficult.<sup>46</sup> Many p-values used to calculate variances were not reported as exact by as <0.05 or <0.01 so serving as upper bounds. We accordingly adopted a purposeful, pragmatic, but cautious approach to sifting, analyzing, and interpretation of KQ1 evidence. For example, clearly identifiable outliers were excluded from main pooled estimates (as apparent on forest, funnel, and radial plots) with results also provided including those estimates. Outliers had implausibly large or small estimated standardized effects (e.g., > 1.5 or < -1.5). Pooling was also performed with and without lesser influential observations; and attempted to include in the network meta-analyses (vasomotor symptoms and QoL consistent effects).

## Use of Standardized Effect Measures

Studies evaluated patient reported outcomes using a variety of outcome instruments and metrics. Standardized effect measures were calculated and pooled according to the EPC Program *Methods Guide*.<sup>15</sup> Calculating the standardized mean difference (SMD), which is (effect1-effect2)/standard deviation, allows for comparison of results across studies using different measures. Analyses were performed in R<sup>16</sup> using the meta,<sup>17</sup> compute.es,<sup>18</sup> and ggplot2<sup>19</sup> packages.

We estimated effects for each arm to calculate SMDs as follows: 1) from reported pre-post change and standard deviation (or error), 2) if baselines were similar using end of treatment means and standard deviation if reported, 3) if baselines differed with baseline and end of treatment standard deviations reported calculated change and estimated standard deviation (assuming 0.5 correlation between initial and final standard deviations), 4) using p-values with baseline and end of treatment value or reported change for arm-specific effect, 5) using between-arm differences and p-values. If an effect was reported as nonsignificant but the trial was to be pooled, a nonsignificant p-value was imputed for pooling so not to selectively exclude nonsignificant results. A small number of trials reported dichotomous outcomes; when feasible they were transformed to continuous measures and quantified as change scores for meta-analysis.

Interpretation of standardized effect measures is challenging as they lack intuitive meaning. However, SMDs can be re-expressed as either odds ratios to aid interpretation; given a control group response of approximately 20% to 60%, odds ratios obtained from standardized effects

appropriately represent relative effects.<sup>82</sup> Different methods allow converting SMDs to odds ratios we adopted Hasselblad and Hedges’ approach,<sup>82</sup>—odds ratios of 0.7, 0.6, 0.5, 2, 3, and 4 corresponding to approximate effect sizes of -0.2, -0.3, -0.4, 0.3, 0.6, and 0.75 respectively. For the network results calculated odds ratios are presented, elsewhere the reader can use these approximate conversions. Furukawa<sup>83</sup> provides an NNT conversion reproduced here (Table 4).

**Table 4. NNTs according to SMD and control group response rates (e.g., achieving some minimal important improvement).**

		Response Rate Control Group (percent)								
		10	20	30	40	50	60	70	80	90
SMD	0.2	25.2	16.5	13.7	12.7	12.6	13.4	15.2	19.5	32.5
SMD	0.5	8.5	6	5.3	5.1	5.2	5.7	6.8	9.1	16
SMD	0.8	4.6	3.5	3.2	3.3	3.5	3.9	4.8	6.7	12.3
SMD	1	3.5	2.8	2.6	2.7	2.9	3.4	4.2	6	11.3

## Pooling

Analyses were performed in R<sup>16</sup> using the meta<sup>17</sup> package. Appropriateness for pooling was judged on the basis of trial characteristics together with subject matter knowledge. Because the goal of any pooling is to estimate unconditional effects,<sup>20</sup> random-effects models were used. The magnitude of statistical heterogeneity was examined by using tau<sup>2</sup> owing to limitations of the I<sup>2</sup> metric and because between-trial variances are more intuitively interpreted on the effect estimate scale.<sup>21</sup> Evidence of possible publication bias was explored by using funnel plots. At the protocol stage, we anticipated examining subgroup-specific effects according vasomotor symptom severity, years since menopause (age), ethnicity, and comorbidities (smoking, obesity). Given inconsistent and incomplete reporting of these variables such analyses were not conducted. In addition, other than for KQ2 trial reporting did not allow evaluating results separately for women with and without a uterus. Outcomes were summarized and reported in the order specified by therapies in the KQs.

## Minimal Clinically Important Differences

To discuss the outcomes in the context of clinical relevance, attempts were made to find established thresholds for the minimal clinically important difference for each outcome. PubMed and Google Scholar were searched for minimal clinically important differences (MCID) for the following: Greene scale, MENQOL, MQOL, WHQ, Kupperman Index, hot flushes, night sweats, Hamilton Depression scale, SF-36, CES-D, McCoy scale, Menopause Rating Scale, Visual Analog Scale and WHI sleep scale. Search terms for MCID included “MCID,” “MID,” “minimal important difference,” “clinical important difference,” “clinically important difference,” “minimal difference,” “clinical difference” and “important difference.” Search terms for outcomes included “Greene scale,” “Greene,” “MENQOL,” “MQOL,” “menopause QOL,” “menopause quality of life,” “WHQ,” “WHQ scale,” “Kupperman Index,” “Kupperman,” “night sweats,” “vasomotor,” “Hamilton,” “HAMD,” “SF-36,” “RAND-36,” “CES-D,” “McCoy sex scale,” “McCoy scale,” “McCoy sex,” “Menopause Rating Scale,” “MRS,” “Visual Analog Scale,” “VAS,” “WHI scale,” “WHI,” and “menopause.” Articles retrieved from the search that

had a postmenopausal patient population were then searched for the MCIDs using the find function and MCID search terms. If MCIDs were not found in articles with a postmenopausal population, then articles with any patient population were searched. Table 5 summarizes the MCID for each outcome or scale.

**Table 5. Minimal clinically important difference for various scales**

Article	Scale	MCID	Note
Huntley, 2003 <sup>84</sup>	Kupperman total score	final score $\leq 15$	Cites Kupperman, 1959 <sup>85</sup>
Kupperman, 1959 <sup>85</sup>	Kupperman total score	final score $\leq 15$	
Morrison, 2004 <sup>86</sup>	Hamilton-Depression	-3 points	
Gelfand, 2003 <sup>87</sup>	MENQOL summary score	0.5 difference between groups	
Zollner, 2005 <sup>88</sup>	MENQOL subscales	1 point change	Cites Hilditch, 2008 <sup>89</sup>
Hilditch, 2008 <sup>89</sup>	MENQOL subscales	1 point change	
Lewis, 2005 <sup>90</sup>	MENQOL subscales MENQOL summary	1 point change	
Wyrwich, 2003 <sup>91</sup>	SF-36 general health SF-36 mental health	Small change: 10 Moderate change: 20 Large change: 30 State change: 5	
Samsa, 1999 <sup>92</sup>	SF-36	3-5 point	
Levine, 2005 <sup>93</sup>	WHI Insomnia Scale	1/2 a SD change	
DeRogatis, 2009 <sup>94</sup>	Satisfying sexual episodes	+1 episode/4-week period	
Simon Lee, 2003 <sup>95</sup>	VAS (0-100)	mean reduction of 30	
Tashjian, 2009 <sup>96</sup>	VAS (0-10)	1.4 change	

## Indirect Comparisons with Mixed Treatment Comparisons Techniques

A random-effects network meta-analysis was performed pooling standardized effects in a Bayesian model described by Chaimani (<http://www.mtm.uoi.gr/3.continuousmodeldescription.pdf>). Models were fitted in OpenBUGS 3.2.2 using noninformative priors and convergence assessed using the Brooks-Gelman-Rubin plot and statistic (no value exceeded 1.002 in the model). A burn-in of 10,000 samples was discarded and subsequent 50,000 analyzed. Rankings were estimated for the probability a treatment was most effective, next most effective, and so on. Effect estimates and accompanying 95 percent credible intervals were obtained from the samples. To evaluate consistency we compared available pairwise estimates to the network results.<sup>22</sup> We examined all pairwise comparisons individually in random effects models and graphically using forest funnel plots.

## Outcome Measures

### Key Questions 1 and 4

Instrument details to assist interpretation are included in respective results sections.

Outcomes for KQ1 and KQ4 were categorized into the following menopausal symptom categories: vasomotor symptoms, sleep disturbance, psychological symptoms, urogenital atrophy, sexual function, and quality of life. Outcomes were self-reported, from daily diaries or derived from validated survey instruments. There existed a wide variety of potential outcome measures for each of the categories, so abstraction was limited to the more common outcomes. The following outcomes, by category, were abstracted for analyses:

Vasomotor symptoms: self-reported hot flushes, night sweats, and severity of hot flushes; vasomotor subscores from instruments such as the Greene Climacteric Scale (GCS), the Kupperman Menopausal Index (KI), Women's Health Questionnaire (WHQ), and the Menopause-specific Quality of Life (MENQOL)

Sleep disturbance: self-reported insomnia and sleep problems; Women's Health Initiative Insomnia Rating Scale, and sleep subscales from GCS, KI, or MENQOL

Psychological symptoms: anxiety, depression, and global measures; subscales from the larger menopause-related survey instruments such as KI, GCS, MENQOL, or from psychological survey instruments such as Beck and Hamilton

Urogenital atrophy: self-reported vaginal dryness; urogenital atrophy or vaginal atrophy subscale scores from KI, GCS, and MENQOL

Sexual function: dyspareunia, satisfying sexual episodes, number of sexual episodes; McCoy Sex Scale, and sexual function subscales from GCS, KI, WHQ, and MENQOL

Quality of life: total scores from GCS, KI, MENQOL

Some investigators devised their own scales rather than using the above standardized scales. We included outcomes that used these other scales as well.

## Key Questions 2 and 3

Outcomes included heart disease (myocardial infarction, angina), stroke, or thromboembolism; cholecystitis; osteoporotic fractures; or endometrial, breast, colorectal, or ovarian cancer.

## Strength of the Body of Evidence

Strength of evidence (SOE) assessments were based on the Evidence-based Practice Center (EPC) approach,<sup>23</sup> which is conceptually similar to the GRADE system.<sup>24</sup> Two reviewers graded the strength of evidence, resolving disagreements by consensus. Details for the strength of evidence approach are also available at the AHRQ Effective Healthcare site, [http://effectivehealthcare.ahrq.gov/repFiles/2009\\_0805\\_grading.pdf](http://effectivehealthcare.ahrq.gov/repFiles/2009_0805_grading.pdf).

We adopted a point-based approach to SOE ratings in which each assessment started at high (3 points) and downgraded by one point each for: high risk of bias, inconsistent or unknown consistency, imprecise or unknown precision, indirect body of evidence, and suspected reporting bias. Domain ratings were entered into a spreadsheet that provided a summary SOE for each outcome. If the summary SOE remained 3 with no downgrades, strength of evidence was rated high; if the summary SOE equaled 2, strength of evidence was rated moderate; if the summary SOE equaled 1, strength of evidence was rated low; if the summary SOE was zero or lower, strength of evidence was rated insufficient.

We imposed one departure from the SOE evidence domains outlined in Table 6. In the presence of a large number of trials ( $n \geq 10$ ), even when a majority was rated poor quality, risk of bias was assigned medium. The rationale was that poor quality was typically due consistently low rating for three items—equal, reliable, and valid measurements; clearly defined

interventions; all important outcomes considered. Notwithstanding, when there were >10 trials (usually more) with consistent effects, and no suspected reporting bias, we concluded that low trial quality did not justify a lower strength of evidence.

For Key Question 1, when sufficient trials allowed for evidence synthesis, strength of evidence was determined by outcome (vasomotor symptoms, quality of life, psychological symptoms, sexual function, urogenital atrophy, and sleep dysfunction) and by comparators. *For outcomes and comparator groups without poolable data represented by single trials, strength of evidence was deemed insufficient and not reported.*

For Key Question 2, strength of evidence was determined by outcome (breast cancer; gall bladder disease; colorectal cancer; coronary heart disease, stroke, and thromboembolism; endometrial cancer; osteoporotic fractures, and ovarian cancer), and by treatment regimen (either estrogens alone or estrogens with progestins).

For Key Question 3, strength of evidence was determined by outcome (breast cancer; gall bladder disease; colorectal cancer; coronary heart disease, stroke, and thromboembolism; endometrial cancer; osteoporotic fractures, and ovarian cancer), and by treatment regimen (antidepressants, isoflavones, and vitamin E).

For Key Question 4, strength of evidence was determined by outcome (vasomotor symptoms, quality of life, psychological symptoms, sexual function, urogenital atrophy, and sleep dysfunction), by subgroup (age, body mass index, race, severity of symptoms, time since menopause, and uterine status), and by treatment regimen (estrogens, other prescription treatments, and nonprescription treatments). *For outcomes and comparator subgroups represented by single trials, strength of evidence was deemed insufficient and not reported.*

**Table 6. Strength of evidence rating domains.**

Domain	Level	Criteria
Risk of bias	General	Degree to which studies have high likelihood of protection against bias; derived from assessment of the risk of bias in individual studies; incorporates both study design and conduct. Grading this domain requires assessment of aggregate quality of studies within each major study design and integration into overall risk of bias score. Limitations of design for reducing bias in addressing a key question should be taken into account. If studies differ substantially in risk of bias, may give greater weight to those studies with low risk of bias.
	Low	At least 1 good quality RCT or nonrandomized comparative study
	Medium	At least 1 fair quality RCT; OR 1 fair quality nonrandomized comparative study; AND 1 additional study of good or fair quality
	High	Does not meet minimum requirements for low or medium risk of bias
Consistency	General	Degree to which studies are similar in effect sizes; degree to which studies have same direction of effect (even in presence of statistical heterogeneity)
	Consistent	Effect sizes have same direction. When multiple RCTs were available and the risk of bias was low, the range of effects needed to be narrow.
	Inconsistent	Effect sizes are in different directions.
	Unknown	Single study evidence base
Directness	General	A single direct link between intervention and health outcome; intervention and comparator(s) compared head-to-head within a study
	Direct	Direct head-to-head comparison of interventions within a study or assesses a final health outcome
	Indirect	Not a direct head-to-head comparison of interventions within a study

		or assesses an intermediate outcome
Precision	General	Degree of certainty surrounding an effect estimate
	Precise	Uncertainty around an effect compatible with only one of these: clinically important superiority, inferiority, or noninferiority. In absence of meta-analysis, individual studies consistently report precise and/or statistically significant results.
	Imprecise	Uncertainty around an effect compatible with both clinically important superiority and inferiority. In absence of meta-analysis, individual studies do not consistently report precise and/or statistically significant results.
Reporting Bias	Suspected	A substantial difference in the pooled fixed effect estimate between small and large studies, such that small study effect reflects an exaggerated benefit or harm, or a qualitative assessment of the risk based on reviewers' consensual judgment of the likely impact of reporting bias on the included evidence.
	Undetected	All other scenarios

## Applicability

Applicability is defined as the extent to which treatment effects observed in published studies reflect expected results when treatments are applied to these populations in the real world.<sup>97</sup> Details on assessing applicability are available on the AHRQ Effectiveness Healthcare site, <http://www.effectivehealthcare.ahrq.gov/ehc/products/272/603/Methods%20Guide--Atkins--01-03-2011KM.pdf>.

The population of interest for this CER is women experiencing symptoms accompanying natural menopause (during perimenopausal or postmenopausal periods) or surgically induced menopause (during the postmenopausal period). Potential factors which may affect applicability in this body of evidence include:

- Study populations may consist of all menopausal women, regardless of presence of symptoms
- Study populations may combine results on menopausal women with and without a uterus
- Study populations may consist of menopausal women with different levels of symptom severity
- Study populations may have a larger proportion of older menopausal women

Limitations in the applicability of individual studies were identified. When there were questions applying results from randomized controlled trials for KQ2, we reviewed observational studies from the original literature search seeking more comparable populations. As suggested by the AHRQ Methods Guide, when applicability issues occurred, they were highlighted and clearly discussed following the evidence tables.

## Peer Review and Public Commentary

Key Informants are the end-users of research, including participants and caregivers, practicing clinicians, relevant professional and consumer organizations, purchasers of health care, and others with experience in making health care decisions. Within the EPC program, the Key Informant role is to provide input into identifying the Key Questions for research that will

inform health care decisions. The EPC solicited input from Key Informants when developing questions for systematic review or when identifying high-priority research gaps and needed new research. Key Informants were not involved in analyzing the evidence or writing the report and have not reviewed the report, except as given the opportunity to do so through the peer or public review mechanism.

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

Technical Experts comprise a multidisciplinary group of clinical, content, and methodological experts who provide input in defining populations, interventions, comparisons, or outcomes as well as identifying particular studies or databases to search. They are selected to provide broad expertise and perspectives specific to the topic under development. Divergent and conflicted opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Therefore trial questions, design, and/or methodological approaches do not necessarily represent the views of individual technical and content experts. Technical Experts provide information to the EPC to identify literature search strategies and recommend approaches to specific issues as requested by the EPC. Technical Experts do not do analysis of any kind nor contribute to the writing of the report and have not reviewed the report, except as given the opportunity to do so through the public review mechanism.

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

Peer reviewers are invited to provide written comments on the draft report based on their clinical, content, or methodological expertise. Peer review comments on the preliminary draft of the report are considered by the EPC in preparation of the final draft of the report. Peer reviewers do not participate in writing or editing of the final report or other products. The synthesis of the scientific literature presented in the final report does not necessarily represent the views of individual reviewers. The dispositions of the peer review comments are documented and will, for CERs and Technical briefs, be published 3 months after the publication of the Evidence report.

Potential Reviewers also disclose any financial conflicts of interest greater than \$10,000 and any other relevant business or professional conflicts of interest. Invited Peer Reviewers may not have any financial conflict of interest greater than \$10,000. Peer reviewers who disclose potential business or professional conflicts of interest may submit comments on draft reports through the public comment mechanism.

[Text to be added pending receipt of comments.]



synthesis (either network meta-analysis or pairwise comparisons) for those trials with data that was amenable to pooling; a strength of evidence assessment for the evidence that was synthesized; a summary of the trials that were not amenable to a quantitative synthesis; and key points.

KQ2 and KQ3 results are presented by condition: breast cancer; gall bladder disease; colorectal cancer; coronary heart disease, stroke, and thromboembolism; endometrial cancer; osteoporotic fractures; and ovarian cancer. KQ3 includes an additional discussion of adverse events.

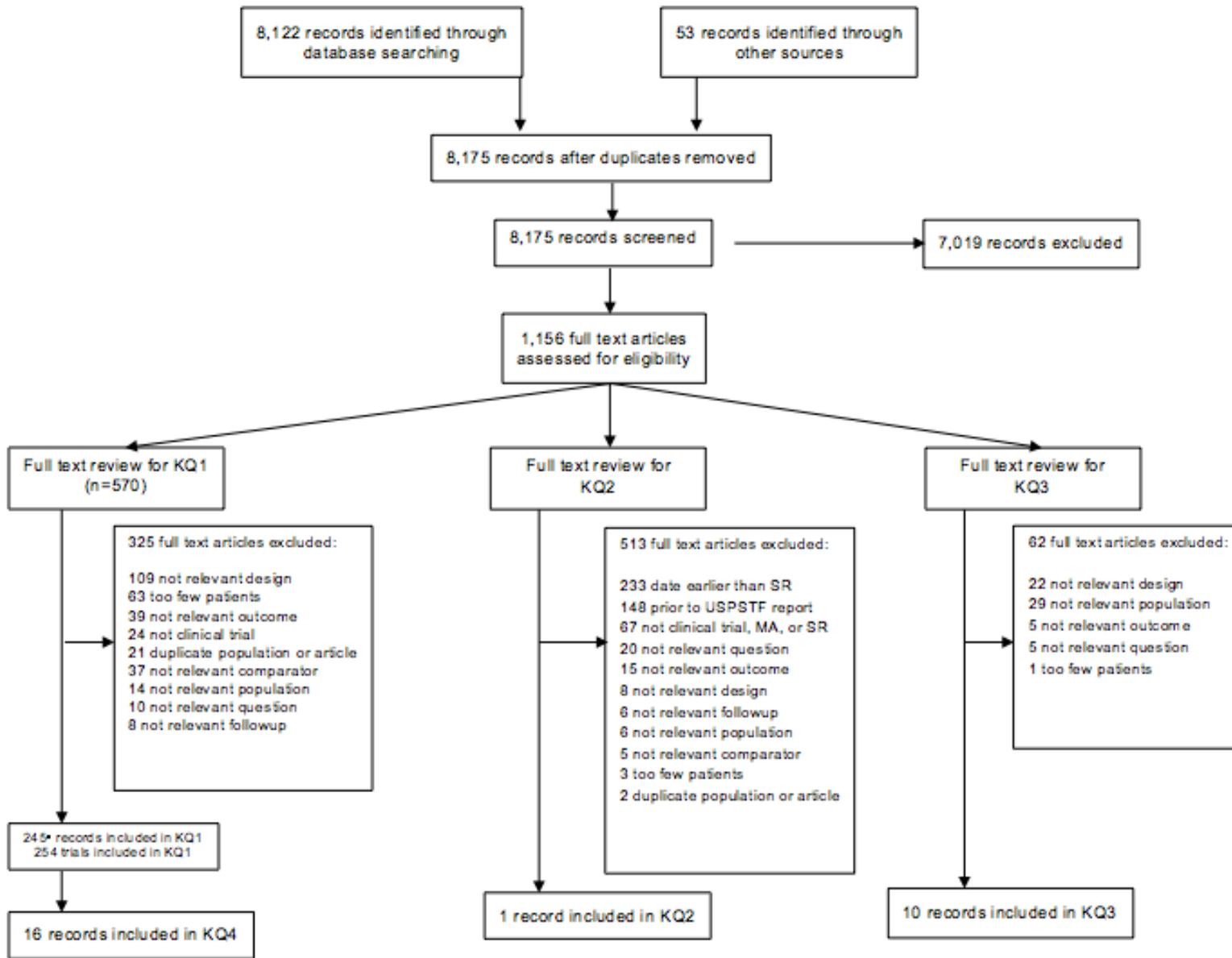
KQ4 results are organized by the six outcome categories, as listed in the KQ1 description.

## **Results of Literature Searches**

The literature search identified 8,122 records, with an additional 53 records identified through the grey literature search and hand searching of bibliographies. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)<sup>98</sup> diagram shown in Figure 2 depicts the flow of search screening and study selection. From the total 8,175 abstracts screened, 1,156 full text articles were assessed for inclusion. For Key Question 1, 570 full text articles were screened, with 245 records included. Nine of those records presented results for two distinct trials, so those publications were given two unique reference numbers and were counted as two trials, for a total of 254 trials included in KQ1. For Key Question 2, a systematic review by Nelson et al.<sup>37</sup> published in May 2012, contained the most current literature review addressing the same outcomes in this key question. This systematic review therefore became the primary source for Key Question 2. For Key Question 3, 72 articles were screened, with 11 studies included: seven RCTs, two cohort studies, and two case control studies. Sixteen trials from Key Question 1 included subgroup analyses of interest and were the evidence base for Key Question 4.

The list of excluded studies with reasons for exclusion is presented in Appendix B.

Figure 2. PRISMA diagram



MA: meta-analysis; SR: systematic review

<sup>a</sup> 9 records presented results from two distinct patient populations and were divided into 2 trials each

# Key Question 1. Effectiveness of Different Treatments for Postmenopausal Symptoms

## Description of Included Studies

Two hundred and fifty-four trials were included in this key question, providing results for the following outcomes: vasomotor symptoms (187 trials), quality of life (108 trials), psychological symptoms (90 trials), sexual function (76 trials), urogenital atrophy (63 trials), and sleep dysfunction (48 trials). Some trials contributed results to more than one outcome.

Evidence synthesis was dependent on the number of trials with comparators and outcomes that could be appropriately pooled. When the number of trials allowed for a synthesis of outcomes by comparator group, either meta-analyses or pairwise comparisons were performed. Strength of evidence was then determined. When there were not enough trials for certain comparators and outcomes, synthesis was not possible and strength of evidence was not determined. Descriptions of these trials are provided.

Results for Key Question 1 are presented by outcome. Within each of these six categories, there are the following sections: a summary table of the included trials; a presentation of the quantitative synthesis (network meta-analysis and/or pairwise comparisons) for trial data amenable to pooling; a strength of evidence rating for synthesized evidence; a summary of the trials that were not amenable to a quantitative synthesis; and key points.

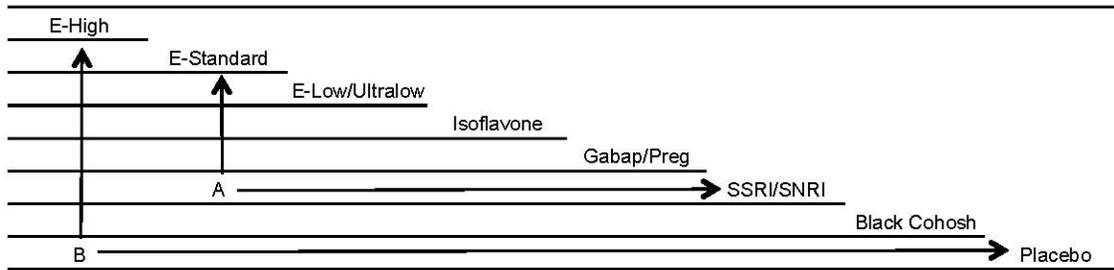
## Navigating Key Question 1 Results

Owing to the use of different outcome scales all results were quantified in a standardized effect metric or a standardized mean difference (SMD) with the goal of quantifying difference in change with treatment. As noted in the methods, with control-group event rates of 20 to 60 percent SMDs can be expressed as odds ratios with SMDs of -0.2, -0.3, -0.4, 0.3, 0.6, and 0.75 corresponding to odds ratios of 0.7, 0.6, 0.5, 2, 3, and 4 respectively. A typical placebo response rate of women with vasomotor symptoms is approximately 25 percent.<sup>99</sup> While it is difficult to ascertain a precise absolute difference in response for outcomes, the SMDs can be translated into NNTs for a given placebo response rate (see Methods).<sup>83</sup> The most useful of the KQ1 results is are likely conclusions pertaining to relative treatment efficacies.

Except for sexual function, results are displayed first as a grid or matrix displaying comparisons among multiple treatments or agents. When a network meta-analysis was performed (vasomotor symptoms and quality of life) the all comparisons are represented as estimated by the model—direct and indirect. For pairwise results only direct comparisons are displayed. Table 8 shows how comparisons are presented in the grid or matrix form. Forest plots for pairwise comparisons can be found in the appendices. When a network meta-analysis was performed, a table of rank efficacy for treatments is shown. Finally, a graphical representation is provided as a caterpillar plot that summarizes all pooled estimates or forest plots, which can be found in appendices. Note that for the network meta-analyses, the plot incorporates all possible comparisons between agents in the analyses, whereas for others, only pairwise pooled (not single-trial) comparisons are shown.

Strength of evidence ratings are provided in the text for comparisons that included evidence from more than a single trial. The specific domain ratings corresponding to each rating are provided in tabular form at the end of each outcome section.

**Table 8. Comparison matrix example. Estimate B represents E-High (high-dose estrogen) compared with placebo; estimate A represents E-Standard (standard-dose estrogen) compared with an SSRI/SNRI**



E: estrogen; Gabap: gabapentin; Preg: pregabalin; SSRI: selective serotonin reuptake inhibitor; SNRI: serotonin and norepinephrine reuptake inhibitor.

# Vasomotor Symptoms

## Included Trials

Treatment effects on vasomotor symptoms were reported in 187 trials (73.6 percent with 141 trials specifying vasomotor symptoms as a primary outcome), including 48,041 patients at more than 3,400 sites. Twenty-three trials were multinational whereas 164 were conducted in 28 countries including Austria, Ecuador, Estonia, Greece, Japan, Norway, Singapore, Spain, Switzerland, Ukraine, Sweden, Thailand, China, Hong Kong, India, Brazil, Finland, France, the Netherlands, South Korea, Taiwan, Denmark, Germany, Turkey, Australia, Canada, United Kingdom, Italy, and the United States (in order of increasing numbers of trials with 61 United States trials).

The mean ages of women enrolled in individual trials ranged from 43.8 to 66.7 years (not reported in 25 trials). The average number years since menopause (3.9 years overall) was reported in only 63 trials. Race or ethnicity was reported in 65 trials (Table 9). In contrast, the presence or absence of a uterus in women was noted in 141 trials and most of these (n=81) enrolled women in either category. Mean body mass index ranged from 17.3 to 29.3 kg/m<sup>2</sup> (from 107 trials).

Approximately two-thirds of trials randomized patients to 2 arms with the remainder to multiple arms. Followup ranged from 4 weeks (for trials of centrally acting agents including SSRIs, SNRIs, gabapentin, and pregabalin) to more than 5 years with a mean of 25.2 weeks. The most commonly studied agents were estrogens (106 trials) administered by various routes and isoflavones (31 trials). Agents examined in fewer trials included SSRIs, SNRIs, eszopiclone, clonidine, methyl dopa, gabapentin, pregabalin, isoflavones, black cohosh, St. Johns wort, ginseng, flax seed, vitamin E, dong quai, DHEA, other herbal ingredients, and combinations of nonprescription agents

Vasomotor symptoms were ascertained and reported in different ways and in 103 trials (55.1 percent) using more than one metric. The most common method was to quantify hot flushes that occurred daily or weekly (and both) but sometimes monthly. Daily occurrence was analyzed if reported, followed by weekly and then monthly. Other instruments and metrics included hot flush severity, night sweats, indices combining frequency and severity of hot flushes, visual analogue scales, graphic rating scales, patients experiencing greater than 50 or 80 percent improvement, and vasomotor scale components (e.g., Greene Climacteric Scale, MENQOL, WHQ, Kupperman Menopausal Index). The vasomotor domains of specific scales are as follows:

- Greene Climacteric Scale includes one hot flush and one night sweat item each rated 0 (none) to 3 (severe).
- WHQ includes one hot flush and one night sweat item rated as 0 (not at all) to 3 (definitely).
- MENQOL vasomotor domain includes hot flushes, night sweats, and sweating items scaled from 0 (not at all bothered) to 6 (extremely bothered).
- Kupperman Menopausal Index includes one hot flush item, scaled from 0 (none) to 3 (severe).

Some measure of hot flush frequency was reported in 110 trials (58.8 percent), hot flush severity in 52 (27.8 percent), night sweats in 22 (11.8 percent), combined hot flush and night sweats in 19 (10.2 percent), Greene vasomotor scale in 25 (13.4 percent), Kupperman vasomotor in 16 (8.6 percent), MENQOL vasomotor in 19 (10.2 percent), WHQ vasomotor in 9 (4.8

percent), and another measure in 32 (17.1 percent). We included in these analyses the most commonly reported outcome metric (hot flush frequency) followed by next most common (severity) and so on. Overall, 135 (72.2 percent) trials reported hot flush frequency, severity, and or night sweats. While all trials assessed the effects of treatment on vasomotor symptoms, for purposes of sensitivity analysis in the network meta-analysis we identified those trials specifying vasomotor symptoms as a primary outcome or symptom presence as an inclusion criterion.

Most trials were rated as poor quality (n=150, 80.2 percent); 15 were fair quality and 22 were good quality. Table 9 displays further detail summarizing trial and patient characteristics. The funding source was not stated for 30.5 percent trials, 50.3 percent appeared completely industry sponsored, 12.3 percent had some industry funding, and 12.8 percent reported funding by public sources.

**Table 9. Characteristics of trials assessing efficacy of treatment on vasomotor symptoms**

	Characteristic	Value
Trial Characteristics	Number of trials	187
	Total number of patients	48,041
	Number of sites from trials that specified (n=145)	3,400 1 to 502 (mean 23; median 5)
	Trials described only as multicenter	19 (10.2)
	Multicenter trials	113 (60.4)
	Two-arm trials	122 (65.2)
	Multi-arm trials	65 (34.8)
	Patients per trial	50 to 2,459 (mean 257; median 156)
	Range of followup (weeks)	4 to 260 (mean 25.5; median 12)
	Funding	Industry only
Public only		24 (12.8)
Industry and public		12 (6.4)
Not stated		57 (30.5)
Comparator Category	Placebo vs. hormone	75 (40.1)
	Antidepressant vs. placebo or other antidepressant	11 (5.9)
	Placebo vs. other prescription	6 (3.2)
	Placebo vs. nonprescription	56 (29.9)
	Placebo vs. hormone vs. nonprescription	2 (1.1)
	Hormone vs. hormone	29 (15.5)
	Hormone vs. nonprescription	3 (1.6)
	Nonprescription vs. antidepressant	1 (0.5)
Study Quality	Good	22 (11.8)
	Fair	15 (8.0)
	Poor	150 (80.2)
Patient Demographics	Mean age (years)	43.8 to 66.7 (NR 25)
	Age range (years)	26.0 to 85.0 (NR 143)
	Years since menopause	3.9 (0.6 to 16.5) (NR 124)
	Current smokers (%)	0.0 to 44.0 (NR 145)
	Mean BMI (kg/m <sup>2</sup> )	17.3 to 29.3 (NR 71)
	White (%)	0.0 to 100.0
	Black (%)	0.0 to 58.8
	Hispanic (%)	0.0 to 16.6
	Asian (%)	0.0 to 100.0
	Other (%)	0.0 to 26.6
Uterus Status	All intact	50 (26.7)
	All absent	10 (5.3)
	Mixed	81 (43.3)
	Range, percentage intact among trials with mixed	22.5 to 99
	Not reported	46 (24.6)

Note: Demographics were not reported in all studies.

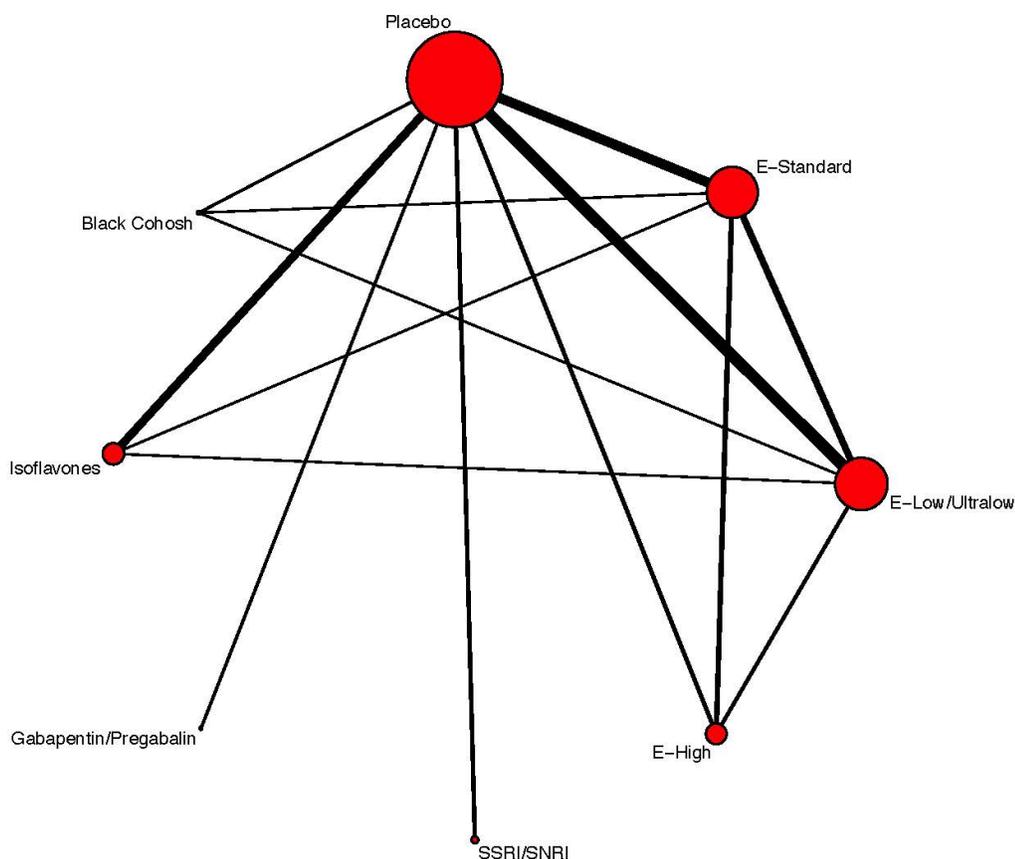
NR—not reported.

## Evidence Synthesis

### Network Meta-analysis

The treatments of greatest clinical interest were studied in multiple trials — estrogens (high-, standard-, and low/ultralow-dose), isoflavones, gabapentin and pregabalin, SSRI/SNRIs, and black cohosh. Comparisons between one or more non-placebo treatments were reported for all but gabapentin and pregabalin. To examine comparative efficacy of these treatments, results were pooled in a network meta-analysis including results from 135 trials (71.8 percent). Figure 3 displays the network of included comparisons. Data were most extensive for estrogens (n=128 comparisons) followed by isoflavones (n=32), SSRI/SNRIs (n=11), gabapentin and pregabalin (n=4), and black cohosh (n=5) (comparisons exceed trial total owing to multi-arm trials). Results from one trial comparing black cohosh with an SSRI (fluoxetine)<sup>100</sup> were excluded from the network comparison because results were highly inconsistent with fitted network.

**Figure 3. Network of comparisons included in vasomotor analyses — line thickness (and circle area) proportional to the number of comparisons**



E: estrogen; SSRI: selective serotonin reuptake inhibitor; SNRI: serotonin and norepinephrine reuptake inhibitor.

Table 10 displays estimated standardized effect sizes and 95 percent credible intervals from the fitted model ordered according to efficacy ranking. Negative SMDs represent fewer reported symptoms. In the bottom row are SMDs comparing each treatment with placebo, the next row up are SMDs comparing each treatment with black cohosh and so on. Of all comparators, estrogens appeared most effective in relieving vasomotor symptoms with high dose estrogen having the largest standardized effect size compared to placebo. Statistical differences were not apparent among estrogen doses. The magnitudes of effect for SSRI/SNRIs, isoflavones, gabapentin or pregabalin, and black cohosh were substantially lower. Estrogens were generally superior to all other agents in the network.

**Table 11. Approximate conversion of standardized effect sizes (SMDs) to reduction in daily hot flashes in trials assessing moderate to severe hot flashes**

Daily HF Reduction	-0.7	-1.3	-1.7	-2.1	-2.4	-2.5	-2.8	-3.0	-3.3	-3.6	-3.9	-4.0
SMD	-0.10	-0.15	-0.20	-0.25	-0.30	-0.35	-0.41	-0.45	-0.55	-0.67	-0.82	-0.90

Table 12 displays efficacy rankings obtained from the network analyses. Estrogens were the highest ranked; SSRI/SNRIs, isoflavones, and gabapentin/pregabalin were similar; with black cohosh and placebo ranked last. Table 13 shows pairwise effect estimates and Figure 4 shows a caterpillar plot displaying all vasomotor symptoms comparisons included in the network analysis and 95% credible intervals.

Comparison of the network and pairwise results were generally highly consistent. Exceptions were for black cohosh and isoflavones compared with estrogens (Appendix F Table 4)—all represented by single trials. As a sensitivity analysis, we deleted all black cohosh comparisons from the network and yielded similar estimates for other comparisons in the network (see Appendix F). We also fitted the network model excluding the 16 trials not specifying vasomotor symptoms as a primary outcome or symptom presence as an inclusion criterion. Results were effectively identical to those obtained from the larger set of trials (Appendix E).

### Estrogen Compared With Placebo

Within the network, there were 89 pairwise comparisons of placebo with estrogen—six with high-dose estrogen (four from poor-quality trials), 37 with standard dose (30 from poor-quality trials), and 46 with low/ultralow dose (40 from poor-quality trials). The magnitudes of pooled standardized effect sizes for all dose categorizations of estrogen are large and the estimates are precise. While a majority of trials was rated as poor quality, the consistency over a large number of comparisons argues that the strength of evidence that estrogens (of any dose) improve hot flush symptoms is rated as high.

### Estrogen Compared With Estrogen

Comparisons among estrogens included 11 comparisons of high-dose with standard dose estrogens, four high-dose with low- or ultralow-dose estrogens, and 21 standard with low- or ultralow-dose estrogens. Results were derived from 33 trials, of which five were rated as good or fair quality. Pooled estimates showed no difference between dose categories: high versus standard (SMD: -0.19; 95 percent CI: -0.46 to 0.08;  $\tau^2=0.20$ ; 11 comparisons); high versus low or ultralow (SMD: -0.12; 95 percent CI: -0.47 to 0.24;  $\tau^2=0.11$ ; four comparisons); and standard versus low or ultralow (SMD: -0.10; 95 percent CI: -0.22 to 0.02;  $\tau^2=0.05$ ; 21

comparisons). The strength of evidence that improvement in vasomotor symptoms does not differ by estrogen dose is rated as high.

### **Isoflavones Compared With Placebo**

There were 29 pairwise comparisons of isoflavones with placebo (23 from low-quality and six from fair- or high-quality trials). Limiting the pairwise analysis to fair- and good-quality trials yielded a larger, although less precise and more heterogeneous, pooled standardized effect size than when all comparisons were included: -0.78 (95 percent CI: -1.51 to -0.06;  $\tau^2=0.78$ ; six comparisons) versus -0.41 (95 percent CI: -0.58 to -0.25;  $\tau^2=0.15$ ; 29 comparisons). The strength of evidence that isoflavones improve hot flush symptoms is rated as moderate.

### **Gabapentin or Pregabalin Compared With Placebo**

Four comparisons of gabapentin or pregabalin with placebo from poor quality trials were included. The standardized effect estimates from the network and pairwise analyses were similar in magnitude but differed in precision: -0.38 (95 percent CI: -0.82 to 0.05) versus -0.33 (95 percent CI: -0.33 to -0.22;  $\tau^2=0$ ; four comparisons) respectively. The strength of evidence that gabapentin or pregabalin improve hot flush symptoms is rated as moderate.

### **SSRIs or SNRIs Compared With Placebo**

There were 10 comparisons of SSRIs or SNRIs (including escitalopram, venlafaxine, desvenlafaxine, citalopram, fluoxetine, and paroxetine) with placebo (six poor and four fair or good quality). The standardized effect estimates from the network and pairwise analyses were similar in magnitude yet differed in precision: -0.36 (95 percent CI: -0.63 to -0.09) versus -0.40 (95 percent CI: -0.54 to -0.26;  $\tau^2=0.03$ ; 10 comparisons) respectively. Limiting the analysis to the four comparisons from high-quality trials yielded a similar result (-0.33; 95 percent CI: -0.45 to -0.22;  $I^2=0$ , 4 comparisons). The strength of evidence that SSRIs or SNRIs improve hot flush symptoms is rated as high.

### **Black Cohosh Compared With SSRI**

Oktem et al.<sup>100</sup> compared black cohosh with fluoxetine for treatment of menopausal symptoms—120 randomized patients with 85 (70.1%) patients evaluated at 12 weeks. Trial quality was rated poor. Using a “monthly hot flush score” the authors reported black cohosh superior to fluoxetine SMD of -1.15 (95% CI: -1.63 to -0.68). (As noted earlier, results from this trial were not included in the network owing to inconsistency).

### **Black Cohosh Compared With Placebo**

Three trials included three comparisons of black cohosh with placebo (two poor and one good quality). The standardized effect estimates from the network and pairwise analyses were similar in magnitude yet differed in precision: -0.31 (95 percent CI: -0.71 to 0.08) versus -0.26 (95 percent CI: -0.43 to -0.09;  $\tau^2=0$ ; three comparisons) respectively. The strength of evidence that black cohosh improves hot flush symptoms is rated as low.

**Table 10. Vasomotor symptoms estimates of comparative efficacy as standardized effect sizes and 95 percent credible intervals from network meta-analysis. Treatments are ordered left to right from most to least efficacious. Highlighted effects are those where the credible interval does not overlap zero. The negative effects reflect improvement (lower on the symptom scale) for the agent on the left versus comparators to its right from intersecting treatments listed on the diagonal**

E-High		E-Standard		E-Low/Ultralow		SSRI/SNRI		Isoflavone		Gabap/Preg		Black Cohosh		Placebo	
-0.11 (-0.29 to 0.08)		-0.04 (-0.18 to 0.10)		-0.23 (-0.53 to 0.06)		-0.04 (-0.35 to 0.28)		-0.04 (-0.50 to 0.41)		-0.12 (-0.71 to 0.47)		-0.26 (-0.68 to 0.15)			
-0.38 (-0.71 to -0.05)	-0.27 (-0.57 to 0.03)														
-0.42 (-0.67 to -0.17)	-0.31 (-0.51 to -0.11)	-0.27 (-0.47 to -0.08)													
-0.46 (-0.93 to 0.00)	-0.35 (-0.80 to 0.09)	-0.32 (-0.75 to 0.12)	-0.08 (-0.59 to 0.42)												
-0.58 (-1.04 to -0.13)	-0.47 (-0.90 to -0.05)	-0.43 (-0.86 to -0.01)	-0.20 (-0.70 to 0.29)	-0.16 (-0.61 to 0.28)											
-0.84 (-1.04 to -0.65)	-0.74 (-0.86 to -0.61)	-0.70 (-0.81 to -0.58)	-0.47 (-0.74 to -0.19)	-0.43 (-0.59 to -0.27)	-0.38 (-0.81 to 0.04)										

E: estrogen; Gabap: gabapentin; Preg: pregabalin; SSRI: selective serotonin reuptake inhibitor; SNRI: serotonin and norepinephrine reuptake inhibitor.

**Table 11. Approximate conversion of standardized effect sizes (SMDs) to reduction in daily hot flashes in trials assessing moderate to severe hot flashes**

Daily HF Reduction	-0.7	-1.3	-1.7	-2.1	-2.4	-2.5	-2.8	-3.0	-3.3	-3.6	-3.9	-4.0
SMD	-0.10	-0.15	-0.20	-0.25	-0.30	-0.35	-0.41	-0.45	-0.55	-0.67	-0.82	-0.90

**Table 12. Vasomotor symptoms rankings of comparative efficacy, standard deviations, and 95% credible intervals (integer values because they arise from a distribution of integers)**

Treatment	Mean Rank	SD	95% CrI
E-High	1.3	0.6	(1-3)
E-Standard	2.3	0.7	(1-4)
E-Low/Ultralow	2.7	0.7	(1-4)
SSRI/SNRI	4.7	1.0	(2-7)
Isoflavone	5.1	0.8	(4-6)
Gabap/Preg	5.2	1.4	(2-8)
Black Cohosh	7.2	0.9	(5-8)
Placebo	7.6	0.5	(7-8)

SD: standard deviation; CrI: credible interval; E: estrogen; SSRI: selective serotonin reuptake inhibitor; SNRI: serotonin and norepinephrine reuptake inhibitor.

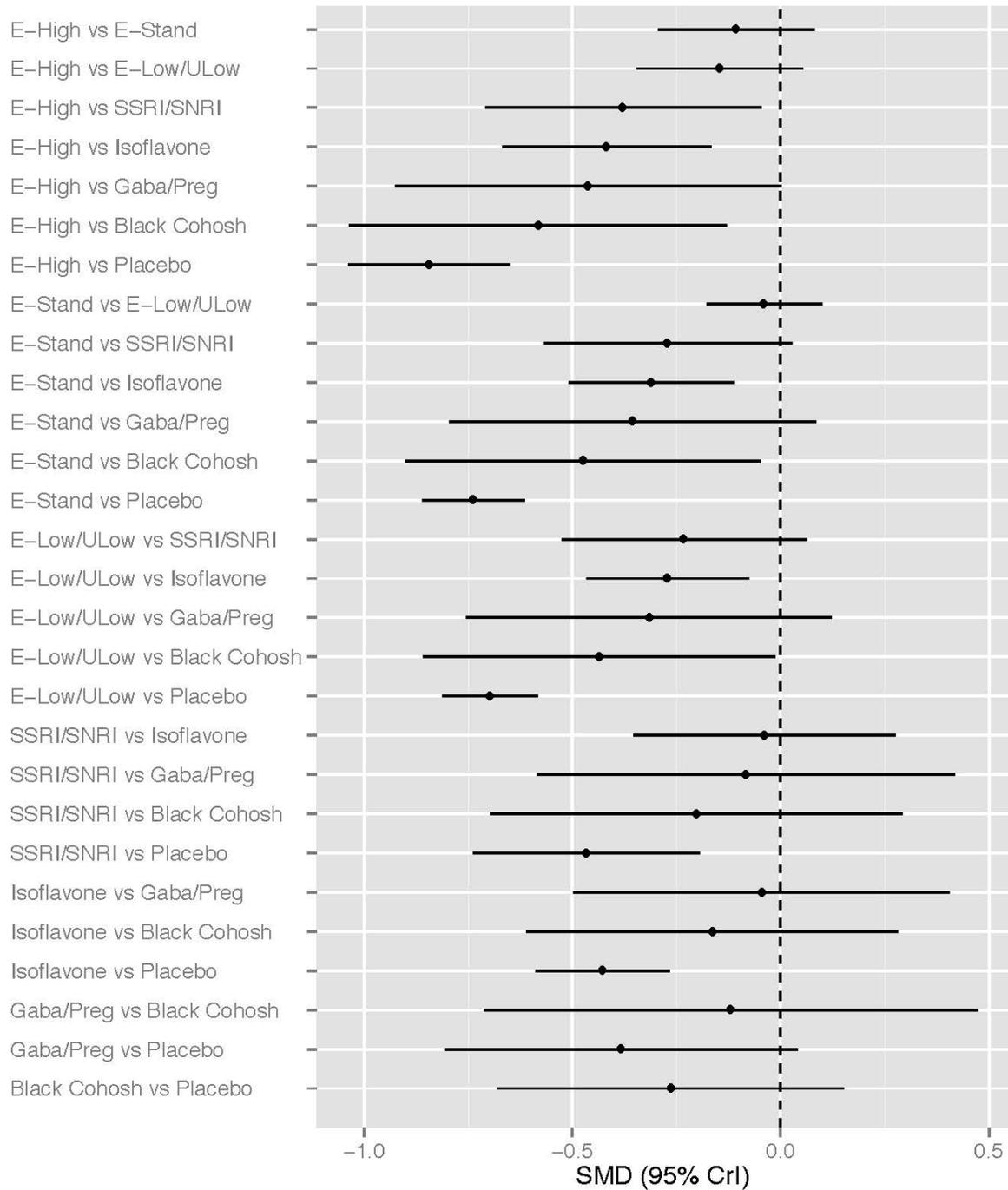
**Table 13. Vasomotor symptoms pairwise effect estimates (pooled random effect estimates or single-trial effects if only data available)**

E-High							
-0.15 (-0.40 to 0.09) $\tau^2=0.18$ n=13	E-Standard						
-0.16 (-0.39 to 0.07) $\tau^2=0.07$ n=7	-0.10 (-0.22 to 0.02) $\tau^2=0.05$ n=21	E-Low/Ultralow					
				SSRI/SNRI			
		0.16 (-0.32 to 0.64) n=1	-0.71 (-1.23 to -0.18) n=1			Isoflavones	
				Gabap/Preg			
		-1.0 (-1.43 to -0.57) n=1	-0.22 (-0.72 to 0.27) n=1	1.15 (0.68 to 1.63) n=1			Black Cohosh
-0.72 (-0.99 to -0.44) $\tau^2=0.14$ n=9	-0.79 (-0.92 to -0.66) $\tau^2=0.12$ n=36	-0.70 (-0.83 to -0.58) $\tau^2=0.15$ n=46	-0.40 (-0.54 to -0.26) $\tau^2=0.03$ n=10	-0.41 <sup>a</sup> (-0.58 to -0.25) $\tau^2=0.15$ n=29	-0.33 (-0.45 to -0.22) $\tau^2=0.00$ n=4	-0.26 (-0.43 to -0.09) $\tau^2=0$ n=3	Placebo

<sup>a</sup> Excluding two outliers isoflavones -0.29 (-0.39 to -0.19),  $\tau^2=0.03$ , n=27

E: estrogen; SSRI: selective serotonin reuptake inhibitor; SNRI: serotonin and norepinephrine reuptake inhibitor.

**Figure 4. Caterpillar plot displaying all vasomotor symptoms comparisons included in the network analysis and 95% credible intervals**



E: estrogen; Ulow: ultralow; Gabap: gabapentin; Preg: pregabalin; SSRI: selective serotonin reuptake inhibitor; SNRI: serotonin and norepinephrine reuptake inhibitor; SMD: standardized mean difference; CrI: credible interval.

## Other Trials Pooled

### Ginseng Compared With Placebo

Three trials (two of poor quality) with a total of 513 participants compared ginseng with placebo.<sup>101, 102</sup> yielding a pooled SMD of -0.41 (95 percent CI: -0.83 to 0.02; tau<sup>2</sup>=0.10). The strength of evidence that ginseng improves vasomotor symptoms is rated as low.

## Trials Not Pooled

### Progestin and Other Hormones Compared With Placebo

Five trials (Table 14) were identified that compared progestin in different doses, either with estrogen<sup>103, 104</sup> or alone,<sup>105-107</sup> for relief of vasomotor symptoms. Three of the trials administered progestin through a cream,<sup>105-107</sup> one through a patch,<sup>103</sup> and one orally.<sup>104</sup> Among the trials using cream, one found significant vasomotor symptom relief with low doses of progestin,<sup>107</sup> with a standard mean difference of less than -1.0 (p<0.001). The other two progestin cream trials report no symptom relief.<sup>105, 106</sup> Rozenberg et al. reported that both sequential and continuous administrations of transdermal estrogens/progestins were as effective as a combination estrogen patch and oral progestins.<sup>103</sup> Gambacciani et al. reported significant improvement in vasomotor symptoms among several combinations of estrogens/progestins.<sup>104</sup> Because trials studied different therapy combinations, the strength of evidence was not rated.

**Table 14. Trials comparing placebo with progestins, reporting vasomotor outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	SMD (95% CI); or p-value
Benster, 2009 <sup>105</sup>	Placebo	—	43	Cream	24	Fair	—
	Progestin	5	46	Cream			-0.18 (-0.59 to 0.23)
	Progestin	20	44	Cream			-0.20 (-0.62 to 0.22)
	Progestin	40	43	Cream			-0.24 (-0.66 to 0.18)
	Progestin	60	45	Cream			-0.39 (-0.81 to 0.03)
Wren, 2003 <sup>106</sup>	Placebo	—	42	Cream	12	Poor	—
	Progestin	32	38	Cream			-0.41 (-0.85 to 0.03)
Leonetti, 1999 <sup>107</sup>	Placebo	—	47	Cream	52	Poor	—
	Progestin	20	43	Cream			< -1.0; p<0.001
Rozenberg, 1997 <sup>103</sup>	Estradiol + NETA	0.05 E + 1 P	153	Oral <sup>a</sup>	52	Poor	—
	Estradiol + NETA	0.05 E + 0.17 P <sup>b</sup>	154	Patch			0.00 (-0.22 to 0.22)
	Estradiol + NETA	0.05 E + 0.35 P <sup>b</sup>	158	Patch			-0.08 (-0.30 to 0.13)
	Estradiol + NETA	0.05 E + 0.17 P <sup>c</sup>	153	Patch			0.00 (-0.22 to 0.22)
	Estradiol + NETA	0.05 E + 0.35 P <sup>c</sup>	156	Patch			0.01 (-0.21 to 0.23)
Gambacciani, 2005 <sup>104</sup>	Estradiol + trimegestone	1 E + 0.125 P	432	Oral	104	Poor	—
	Estradiol + norethisterone	1 E + 0.5 P	242	Oral			0.04 (-0.12 to 0.19)
	Estradiol + norethisterone	2 E + 1 P	176	Oral			-0.11 (-0.28 to 0.07)
	Estradiol + norethisterone						

<sup>a</sup> The reference group was randomized 1:1 to receive an estrogen patch and the progestin orally either by 20 mg daily dydrogesterone or 1 mg for 2 weeks norethisterone

<sup>b</sup> Estradiol and NETA combined

<sup>c</sup> Estradiol and NETA sequential

SMD: standardized mean difference; CI: confidence interval; E: estrogen; NETA: norethisterone acetate; P: progestin; NS: not significant; FU: followup; Wks: weeks.

## Other Prescription Agents Compared With Placebo

One trial compared eszopiclone, a sedative, with placebo for the relief of vasomotor symptoms (Table 15).<sup>108</sup> In this randomized, double-blind, placebo-controlled crossover trial, half the participants (n=30) received eszopiclone patches for four weeks, followed by a two-week washout period, and then four weeks of placebo patches. The other half of the participants (n=29) received the placebo patches first, followed by the eszopiclone patches. There was no difference between eszopiclone and placebo in the relief of vasomotor symptoms.<sup>108</sup>

One trial compared clonidine with placebo and reported mean change in weekly hot flushes (Table 15).<sup>109</sup> In this double-blind, placebo-controlled crossover trial, treatment lasted four weeks. Treatment with clonidine resulted in 19.2 fewer hot flushes per week while 13.1 fewer hot flushes per week were reported during the placebo phase. The standardized effect size was -0.08 (95 percent CI: -0.50 to 0.34).

**Table 15. Trials comparing placebo with other prescription agents, reporting vasomotor outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	SMD (95% CI)
Joffe, 2010 <sup>108</sup>	Placebo	—	29	Oral	4	Poor	—
	Eszopiclone	3	30	Oral			NS
Clayden, 1974 <sup>109</sup>	Placebo	—	43	Oral	4	Poor	—
	Clonidine	0.05-0.15	42	Oral			-0.08 (-0.50 to 0.34)

SMD: standardized mean difference; CI: confidence interval; NS: not significant; FU: followup; Wks: weeks.

## Different Modes of Estrogen Administration

Eight trials<sup>110, 111 49, 112-117</sup> compared different modes of estrogen administration employing similar doses (Table 16) (see Appendix D for dose categorization by route of administration) and so were not included in other comparative analyses. Absent a factorial design, discerning treatment differences according to route of administration in trials comparing differing doses and routes is potentially problematic. In contrast, the comparisons below provide direct evidence on route. All of the trials identified had confidence intervals including zero, supporting the conclusion that estrogens are equally effective regardless of mode of administration. The strength of evidence that estrogens improve vasomotor symptoms by different modes of administration is rated as moderate.

**Table 16. Trials comparing estrogens of similar dose but by different route, reporting vasomotor outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	SMD (95% CI)
Odabasi, 2007 <sup>110</sup>	Estradiol + Progesterone	0.3 E + 90 P	32	Spray	12	Poor	—
	Estradiol + Progesterone	0.05 E + 90 P	29	Patch			-0.75 (-1.26 to -0.23)
Serrano, 2006 <sup>111 49</sup>	CEE + MPA	0.625 E + 10 P	55	Oral	52	Poor	—
	Estradiol + MPA	0.05 E + 10 P	59	Patch			0.20 (-0.17 to 0.56)
Davis, 2005 <sup>112</sup>	Estradiol	0.05	60	Patch	16	Poor	—
	Estradiol	0.3	60	Spray			0.06 (-0.30 to 0.41)
Ozsoy, 2002 <sup>113</sup>	Estradiol + MPA	2 E + 5 P	100	Oral	24	Poor	—
	Estradiol + MPA	0.3 E + 5 P	101	Spray			0.0 (-0.28 to 0.28)
Lopes, 2001 <sup>114</sup>	Estradiol + dydrogesterone	0.05 E + 10 P	185	Patch	12	Poor	0.10 (-0.11 to 0.30)
	Estradiol + dydrogesterone	0.3 E + 10 P	176	Spray			
Mattsson, 2000 <sup>115</sup>	Estradiol + dydrogesterone	2 E + 10 P	342	Oral	24	Good	—
	Estradiol +	0.3 E + 10 P	317	Spray			0.0 (-0.15 to 0.15)

dydrogesterone							
Studd, 1995 <sup>116</sup>	CEE + dydrogesterone Estradiol + dydrogesterone	0.625 E + 20 P 0.05 E + 20 P	104 100	Oral Patch	12	Poor	— 0.23 (-0.05 to 0.50)
Parsey, 2000 <sup>117</sup>	Estradiol CEE	0.025 0.3	95 98	Patch Oral	12	Poor	— 0.04 (-0.24 to 0.32)

SMD: standardized mean difference; CI: confidence interval; E: estrogen; CEE: conjugated equine estrogen; MPA: medroxyprogesterone acetate; P: progestin; NS: not significant; FU: followup; Wks: weeks.

## Other Nonprescription Agents Compared with Placebo

Twenty-one trials compared nonprescription treatments (other than isoflavones and ginseng) with placebo for the relief of vasomotor symptoms (Table 17). Nonprescription treatments included various herbal or plant extracts,<sup>118-131</sup> black cohosh,<sup>132-134</sup> St. John's wort,<sup>132, 134, 135</sup> DHEA,<sup>136</sup> and other nutritional supplements.<sup>137, 138</sup> Eight of the trials showed improvements in vasomotor symptoms compared with placebo, though none of these eight trials tested the same nonprescription agent. The variety of treatments and dosages among these 21 trials did not allow for pooling effects.

**Table 17. Trials comparing nonprescription agents with placebo, reporting vasomotor outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	SMD (95% CI)
Haines, 2008 <sup>118</sup>	Placebo Dang gui and huang qi	— 3000	39 45	Oral Oral	26	Poor	— 0.18 (-0.25 to 0.60)
Garcia, 2010 <sup>119</sup>	Placebo Nutrafem® <sup>a</sup>	— 300	28 103	Oral Oral	12	Poor	— -0.65 (-1.07 to -0.22)
van der Sluijs, 2009 <sup>120</sup>	Placebo Plant extracts <sup>b</sup>	— 3820	46 46	Oral Oral	16	Good	— 0.12 (-0.28 to 0.53)
van Die, 2009 <sup>135</sup>	Placebo St. John's wort	— 900	50 50	Oral Oral	16	Good	— 0.13 (-0.26 to 0.52)
Yang, 2007 <sup>121</sup>	Placebo Pine extract	— 200	75 80	Oral Oral	24	Poor	— -0.47 (-0.78 to -0.15)
Chung, 2007 <sup>132</sup>	Placebo Black cohosh/St. John's wort	— 84	35 42	Oral Oral	12	Poor	— -0.73 (-1.19 to -0.27)
Mucci, 2006 <sup>122</sup>	Placebo Isoflavones, lactobacilli, magnolia bark	— 60	45 44	Oral Oral	24	Poor	— -0.72 (-1.14 to -0.29)
Heger, 2006 <sup>123</sup>	Placebo Rheum raphonticum	— 4	55 54	Oral Oral	12	Poor	— -0.64 (-1.03 to -0.26)
Winther, 2005 <sup>124</sup>	Placebo Femal® <sup>c</sup>	— 80	32 32	Oral Oral	13	Good	— -0.14 (-0.63 to 0.34)
Verhoeven, 2005 <sup>133</sup>	Placebo Isoflavones/black cohosh	— 50	64 60	Oral Oral	12	Good	— -0.15 (-0.50 to 0.20)
Davis, 2001 <sup>125</sup>	Placebo 12 Chinese herbs	— NA	27 29	Oral Oral	12	Poor	— 0.43 (-0.10 to 0.96)
Hirata, 1997 <sup>126</sup>	Placebo Dong quai	— 4500	36 35	Oral Oral	24	Poor	— 0.22 (-0.24 to 0.68)
Chenoy, 1994 <sup>127</sup>	Placebo Primrose oil	— 4000	28 28	Oral Oral	26	Poor	— 0.67 (0.14 to 1.20)
Hsu, 2011 <sup>128</sup>	Placebo Dioscorea alata	— 24	25 25	Oral Oral	52	Poor	— -0.41 (-0.97 to 0.14)
Uebelhack,	Placebo	—	150	Oral	16	Good	—

2006 <sup>134</sup>	Black cohosh/St. John's wort	3.75	151	Oral			-0.85 (-1.08 to -0.61)
Dodin, 2005 <sup>137</sup>	Placebo	—	94	Oral	52	Fair	—
	Flaxseed	40,000	85	Oral			0.29 (-0.01 to 0.58)
Barnhart, 1999 <sup>136</sup>	Placebo	—	30	Oral	12	Poor	—
	DHEA	50	30	Oral			-0.22 (-0.72 to 0.28)
Andrikoula, 2011 <sup>138</sup>	Placebo	—	34	Oral	12	Poor	—
	Nutritional supplement <sup>d</sup>	NA	36	Oral			0.22 (-0.24 to 0.69)
Auerbach, 2012	Placebo	—	38	Oral	12	Poor	—
	Pomegranate seed oil	0.254	43	Oral			-1.53 (-2.02 to -1.04)
Chang, 2011 <sup>130</sup>	Placebo	—	33	Oral	12	Fair	—
	EstroG-100® <sup>e</sup>	NA	31	Oral			-1.74 (-2.31 to -1.17)
Xia, 2012 <sup>131</sup>	Placebo	—	36	Oral	12	Good	—
	Jiawei Qing'e Fang	3500	36	Oral			-0.32 (-0.78 to 0.14)

SMD: standardized mean difference; CI: confidence interval; DHEA: dehydroepiandrosterone; FU; followup; Wks: weeks.

<sup>a</sup> combination of Mung beans, Eucommia bark

<sup>b</sup> combination of black cohosh, er xian tang, zhi bai di huang wan

<sup>c</sup> combination of pure pollen, pollen/pistil extract

<sup>d</sup> combination of 21 vitamins and minerals

<sup>e</sup> combination of cynanchum wilfordii, phlomis umbrosa, angelica gigas

## Estrogen Compared With a Nonprescription Agent

One trial (Table 18) compared estrogen, with or without progestin, with a nonprescription treatment, pueraria mirifica, for the relief of vasomotor symptoms.<sup>139</sup> Pueraria mirifica is a highly estrogenic herb found in Thailand. Both treatments reduced hot flushes as measured by the Greene Climacteric Scale (0 = none, 1 = mild, 2 = moderate, 3=severe). After three months of followup, pueraria mirifica reduced the average Greene score from 2.1 to 0.55 and estrogen treatment reduced the score from 2.1 to 0.35.

**Table 18. Trials comparing estrogen with a nonprescription agent, reporting vasomotor outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	SMD (95% CI)
Chandeying, 2007 <sup>139</sup>	CEE + MPA	0.625 E + 2.5 P	30	Oral	24	Poor	—
	Pueraria mirifica	50	30	Oral			NS

SMD: standardized mean difference; CI: confidence interval; CEE: conjugated equine estrogen; MPA: medroxyprogesterone acetate; E: estrogen; P: progestin; FU: followup; Wks: weeks.

## Nonprescription Agents Compared

Three trials (Table 19) compared nonprescription agents for relief of vasomotor symptoms. In one trial, two different dosages of pueraria mirifica were equally effective in relieving vasomotor symptoms,<sup>140</sup> and in another trial, isoflavones compared with isoflavones and magnolia bark were equally effective in relieving vasomotor symptoms.<sup>141</sup> In a trial comparing vitamin E with isoflavones, isoflavones significantly improved vasomotor symptoms compared with vitamin E. After one year followup, 41.9 percent of the isoflavone group report no more hot flushes and 16.1 percent of the vitamin E group report no more hot flushes (p<0.05).<sup>142</sup>

**Table 19. Trials comparing nonprescription agents, reporting vasomotor outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	SMD (95% CI)
Agosta, 2011 <sup>141</sup>	Isoflavones	60	301	Oral	12	Poor	—
	Isoflavones/magnolia bark	60	335	Oral			NS
Virojchaiwong, 2011 <sup>140</sup>	Pueraria mirifica	25	26	Oral	26	Poor	—
	Pueraria mirifica	50	26	Oral			-0.22 (-0.76 to 0.32)

Zervoudis, 2008). <sup>142</sup>	Vitamin E Isoflavones	500 UI NR	31 31	Oral Oral	52	Poor	p<0.05
----------------------------------	--------------------------	--------------	----------	--------------	----	------	--------

SMD: standardized mean difference; CI: confidence interval; NS: not significant; UI: international unit; NR: not reported; FU: followup; Wks: weeks.

## Trials Without Quantifiable or Poolable Data

Seven trials did not have data that could be analyzed using standardized effect sizes methods. Results of these trials would not have affected the overall outcomes presented above.

In a 12-week, placebo-controlled trial, Hedrick et al. compared three levels of estradiol gel 0.1% at doses of 1.0 (n=125), 0.5 (n=123), and 0.25 (n=122) mg/day with placebo (n=125) for the relief of vasomotor symptoms.<sup>143</sup> Outcomes of interest were number and severity of moderate-to-severe vasomotor symptoms, reported as median change from baseline. After 12 weeks, women in all three treatment groups showed significant improvements in number and severity of moderate-to-severe vasomotor symptoms compared with placebo.<sup>143</sup>

A trial comparing oral (n=35), gel (n=25), and patch (n=28) administrations of estrogen with or without progestin collected information on complete symptom relief of vasomotor symptoms. The authors reported that all three modes of administration were successful in relieving vasomotor symptoms.<sup>144</sup>

A secondary analysis from the UltraLow-dose Transdermal Estrogen Assessment Trial (ULTRA) reported change in vasomotor symptoms using categories of “markedly improved,” “somewhat improved,” and “no improvement.”<sup>145</sup> Diem et al. report that women in the treatment group did not show improvements in hot flushes compared with the placebo group.<sup>145</sup>

Hidalgo et al.<sup>146</sup> compared one or two daily doses of a compound containing isoflavones 60 mg, primrose oil 440 mg, and vitamin E 10 mg in 1080 women. Similar reductions in hot flush frequency occurred in both groups over 6 months, but no variance estimate was reported or calculable.

Kerwin et al.<sup>147</sup> compared the variability of women’s responses in reduction of hot flushes with sertraline or placebo in a crossover trial. They reported an average “modest” response with either treatment.

Plotnikoff et al.<sup>148</sup> evaluated two doses of TU-025 keishibukuryogan with placebo in 178 post-menopausal women. Over 12 weeks, the Mayo hot flush scores declined by 34 percent (placebo), 40 percent (low-dose TU-025), and 38 percent (high-dose TU-025). No difference was reported for the omnibus comparison of arms.

## Strength of Evidence Ratings Vasomotor Symptoms

**Table 20. Strength of evidence ratings domains for vasomotor symptoms<sup>a</sup>**

Comparisons	Comparators <sup>a</sup>		Risk of Bias	Consistency	Directness	Precision	Reporting Bias	SOE	Downgrading Rationale
89	<b>Estrogen</b>	Placebo	M	C	D	P	U	High	—
36	<b>Estrogen</b>	<b>Estrogen (different dose)</b>	M	C	D	P	U	High	—
29	<b>Isoflavone</b>	Placebo	M	I	D	P	U	Mod	Trials different effect direction

4	<b>Gabapentin/ pregabalin</b>	Placebo	H	C	D	P	U	Mod	Poor trial quality
10	<b>SSRI/SNRI</b>	Placebo	M	C	D	P	U	High	6 poor- and 4 fair/good-quality trials
3	<b>Black cohosh</b>	Placebo	H	C	D	I	U	Low	2 poor- and 1 good-quality trial; wide pooled CI
3	<b>Ginseng</b>	Placebo	H	C	D	I	U	Low	3 poor-quality trials; CI overlapping 0
8	<b>Estrogen mode a</b>	<b>Estrogen mode b</b>	H	C	D	P	U	Mod	7 poor- and 1 good-quality trial

Risk of Bias: High (H), Medium (M), Low (L); Consistency: Inconsistent (I), Unknown (U), Consistent (C); Directness: Indirect (I), Direct (D); Precision: Imprecise (I), Unknown (U), Precise (P); Reporting Bias: Suspected (S), Undetected (U).

a Bold font of comparator indicates the more effective treatment; if both comparators are bold, the treatments are equivalently effective

SOE: strength of evidence; Mod: moderate; CI: confidence interval.

## Key Points

- 187 trials including 48,041 women examined treatment of vasomotor symptoms with prescription agents (estrogen, SSRIs, SNRIs, gabapentin, pregabalin, progestins, eszopiclone, and clonidine) and nonprescription agents (isoflavones, black cohosh, vitamin E, flax seed, St. John's wort, ginseng, and a variety of herbs and other agents).
- Most trials were rated as poor quality and 62.5 percent were funded in whole or in part by industry.
- Amelioration of vasomotor symptoms was measured in a number of different patient-reported outcomes—most trials commonly used some metric of hot flashes.
- Strength of evidence of relative effectiveness of agents in relieving vasomotor symptoms
  - There is **high** strength of evidence that estrogen is the most effective agent in relieving vasomotor symptoms. Combined results of trials that included a total of more than 19,000 women showed that the SMD is -0.7 or lower compared with placebo. There is **high** strength of evidence that different doses of estrogen are equally effective.
  - There is **moderate** evidence that isoflavones (N=3,246) and gabapentin or pregabalin (N=1,347) are more effective than placebo; and **high** strength of evidence for SSRIs or SNRIs (N=4,160). These agents are less effective than estrogen in relieving vasomotor symptoms, with a combined SMD of -0.40 or less.
  - There is **low** strength of evidence that black cohosh (N=572) and **insufficient** evidence that ginseng (N=513) are effective compared with placebo. The combined SMD for black cohosh is -0.26. For all other agents, data were not quantifiable and most agents were studied in single trials.
- Analyses comparing effectiveness of treatments show estrogens alleviate vasomotor symptoms best, with the following mean rankings: high dose estrogens (1.3), standard dose estrogens (2.3), and low dose estrogens (2.7). The nonhormone treatments were ranked much lower: SSRI/SNRI (4.7), isoflavones (5.1), gabapentin/pregabalin (5.2), and black cohosh (7.2).

## Quality of Life

### Included Trials

Of the 254 trials, 108 (42.5 percent) reported quality of life outcomes (57 trials specified quality of life as a primary outcome). Fifty-four trials examined hormone treatment effects on quality of life, with the following comparators: placebo (36 trials), other hormones (16 trials), and nonprescription treatments (two trials). Fifty trials examined nonprescription treatment effects on quality of life, with the following comparators: placebo (44 trials), other nonprescription treatments (three trials), hormones (two trials), and antidepressants (one trial). Nonprescription treatments included isoflavones, ginseng, black cohosh, DHEA, herbal extracts, and vitamins and minerals. Seven trials compared antidepressants' effect on quality of life compared with placebo (six trials) and nonprescription treatments (one trial). Desvenlafaxine, escitalopram, and fluoxetine were the antidepressants included in the trials.

The 108 trials originated from 30 different countries and 17 trials were multinational. The most common countries included the United States (n=16), Italy (n=9), Australia (n=5), Germany (n=5), and Turkey (n=5). Other countries contributing evidence include the United Kingdom, the Netherlands, Denmark, Canada, and Brazil. The trials were conducted in 2,271 sites. Length of followup ranged from 8 to 187 weeks.

Quality-of-life outcomes were reported using a variety of scales, both general health related quality of life scales and menopause-specific quality of life scales. A majority of the trials used menopausal quality of life scales (n=96), which focus on physical and psychological symptoms relating to menopause. Several trials used general health related quality of life measures which include broader domains, such as the Short Form-36 (SF-36, sometimes referred to as Rand-36), EuroQOL, Utian QOL, and 15D (n=12). The most common scales in the included trials were: Kupperman Menopausal Index (n=56), Greene Climacteric Scale (n=17), Menopause Rating Scale (MRS) (n=8), Menopause-specific Quality of Life (MENQOL) (n=7), and SF-36 (n=5). The following are brief descriptions of scales which were used in more than five trials:

- The Kupperman Index is a numerical index that scores 11 menopausal symptoms: hot flushes, paresthesia, insomnia, nervousness, melancholia, vertigo, weakness, arthralgia or myalgia, headache, palpitations, and formication. Each symptom is rated from 0 to 3 according to severity, where 0 = no symptoms and 3 = most severe. The scores are weighted and a total sum is calculated. The maximum score is 51 points, with a higher score indicating a worse quality of life.
- The Greene Climacteric Scale includes 21 questions covering five domains: anxiety, depression, somatic symptoms, vasomotor symptoms, and sexual function. Each question is answered on a four-point Likert scale (0 – “not at all”; 1 – “a little”; 2 – “quite a bit”; 3 – “extremely”). The answers to all 21 questions are summed to give a total quality of life measure, in which a higher score indicates a worse quality of life.
- MENQOL consists of 29 questions covering four domains: vasomotor, psychosocial, physical, and sexual. The scoring for each question is 1 – “No”, 2 – “Yes, but not at all bothered” through 8 – “Yes, extremely bothered.” The scores for each question are summed for a total quality-of-life score, in which the higher score indicates a worse quality of life.
- MRS scores 11 menopausal symptoms: hot flushes, heart discomfort, sleep problems, depressive mood, irritability, anxiety, physical and mental exhaustion, sexual problems, bladder problems, vaginal dryness, and joint and muscular discomfort.

Each item is scored from 0 – “none” to 4 – “extremely severe.” The scores are summed for a total quality-of-life score, in which a higher score indicates a worse quality of life.

- SF-36, or Rand-36, is a general quality-of-life scale, not created specifically for menopausal women. This scale consists of 36 questions covering the following eight domains: physical functioning, role limitations caused by physical health problems, role limitations caused by emotional problems, social functioning, emotional well-being, energy/fatigue, pain, and general health perceptions. The answer to each question is transformed linearly to a 0-100 score and then all items in one domain are averaged. This scale can be used to produce outcomes on a total quality of life, subscores for each of the domains, a physical health subscore, or a mental health subscore.

Study quality was generally rated as poor (75.0 percent), with 12 fair- and 15 high-quality trials. Industry funding was indicated in 47 trials and public funding was reported in 17 trials. Table 21 describes additional trial and patient characteristics.

**Table 21. Characteristics of trials assessing efficacy for quality of life outcomes**

	Characteristic	Number (%)
Trial Characteristics	Number of trials	108
	Total number of patients	56,497
	Total sites from trials that specified (n=76)	2,271 1 to 502 (mean: 25; median: 2)
	Trials described only as multicenter	11 (10.2)
	Multicenter trials	61 (56.5)
	Two arm trials	79 (73.1)
	Multi-arm trials	29 (26.9)
	Patients per trial	50 to 16,608 (mean: 523; median:146)
	Range of followup (weeks)	8 to 187 (mean: 28.7; median: 23)
	Funding	Industry only
Public only		17 (15.7)
Industry and public		9 (8.3)
Not stated		35 (32.4)
Comparator Category	Placebo vs. hormone	36 (33.3)
	Antidepressant vs. placebo or other antidepressant	6 (5.6)
	Placebo vs. other prescription	0 (0.0)
	Placebo vs. nonprescription	44 (40.7)
	Placebo vs. hormone vs. nonprescription	0 (0.0)
	Hormone vs. hormone	16 (14.8)
	Hormone vs. nonprescription	2 (1.9)
	Nonprescription vs. antidepressant	1 (0.9)
	Nonprescription vs. nonprescription	3 (2.8)
Study Quality	Good	15 (13.9)

	Fair	12 (11.1)
	Poor	81 (75.0)
Patient Demographics	Mean age (years)	43.8 to 66.8 (NR: 10)
	Age range (years)	29.0 to 85.0 (NR: 85)
	Years since menopause	3.5 (0.6 to 18.6 (NR: 72)
	Current smokers (%)	0.0 to 41.2 (NR: 88)
	Mean BMI (kg/m <sup>2</sup> )	17.3 to 30.1 (NR: 37)
	White (%)	0.0 to 100.0
	Black (%)	0.0 to 58.8
	Hispanic (%)	0.0 to 66.1
	Asian (%)	0.0 to 100.0
	Other (%)	0.0 to 11.5
Uterus Status	All intact	30 (27.8)
	All absent	7 (6.5)
	Mixed	35 (32.4)
	Range, percentage intact among trials with mixed	22.5 to 96.9
	Not reported	36 (33.3)

Note: Demographics were not reported in all studies.

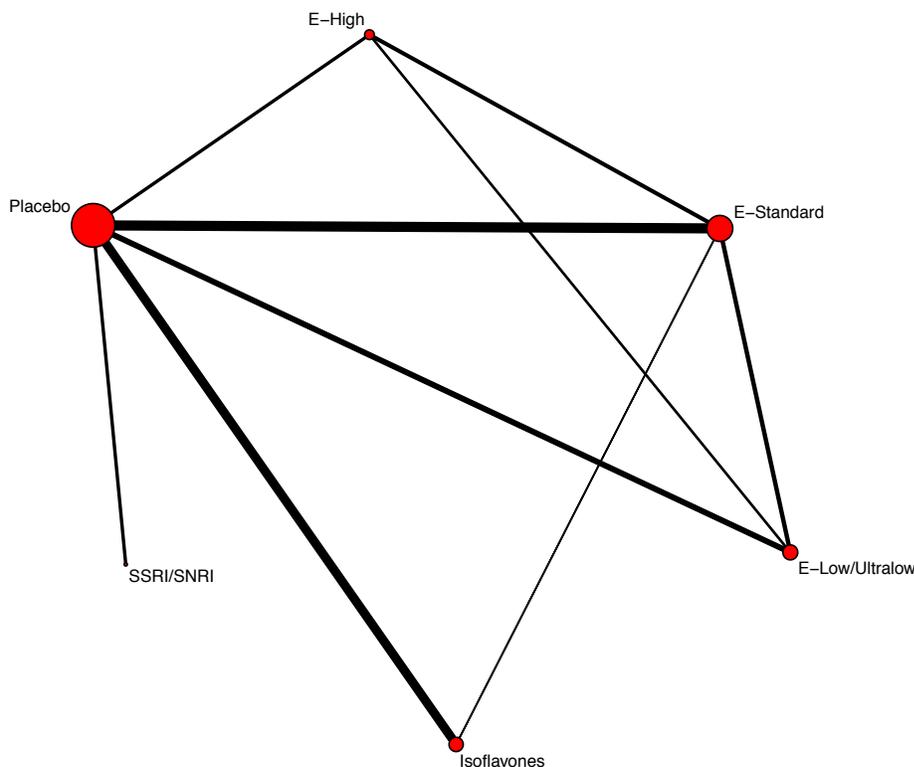
NR: not reported.

## Evidence Synthesis for Quality of Life

### Network Meta-analysis

The treatments judged of greatest clinical interest and that were studied in multiple trials—estrogens (high-, standard-, and low/ultralow-dose), isoflavones, and antidepressants—were compared in a network meta-analysis. The network analysis was conducted twice, once including all eligible trials and a second which included only those trials whose outcomes were measured using menopause-specific quality of life scales. Figure 5 displays the network and comparisons included. Data were most extensive for estrogens (n=54 comparisons), followed by isoflavones (n=18 comparisons), and SSRI/SNRIs (n=5 comparisons). Results from three trials judged to be numerical outliers were excluded and examined in sensitivity analyses for pairwise comparisons (one estrogen versus placebo trial<sup>149</sup>, and two isoflavones versus placebo trials<sup>150, 151</sup>). Additionally, four trials examining black cohosh were not incorporated in the network meta-analysis because their effects were inconsistent with others included (three comparisons with placebo<sup>132, 134, 152</sup> and one with fluoxetine<sup>100</sup>).

**Figure 5. Network of comparisons included in quality of life analyses—line thickness and circle area are proportional to the number of comparisons**



E: estrogen; SSRI: selective serotonin reuptake inhibitor; SNRI: serotonin and norepinephrine reuptake inhibitor.

Table 22 displays estimated standardized effect sizes and 95 percent credible intervals from the fitted model using all trials which had data that could be pooled. In the bottom row are SMDs comparing each treatment with placebo, the penultimate row are SMDs comparing each treatment with isoflavones, and so on. The highest quality of life scores were reported in women taking estrogens, and while best with a standard dose, differences between doses were not distinguishable statistically. Compared with placebo, SSRI/SNRIs were associated with scores of lesser magnitude (credible interval overlapping 0) and isoflavones with the lowest scores. The second network analysis, excluding the trials using general health related quality of life scales, resulted in comparable effect sizes and credible intervals which did not substantively change the results in Table 22.

Figure 6 displays the SMDs estimated in the network as a caterpillar plot. Table 23 lists how the comparative treatments ranked with accompanying uncertainty; lower ranking representing better reported quality of life scores. While there is considerable overlap in the credible intervals, estrogens appear to be superior to all other agents in the network and all agents were better than placebo. Finally, Table 24 displays pooled effects from pairwise meta-analyses. There was little discrepancy with the network analysis indicating the network-estimated direct and indirect effects likely accurate representations.<sup>22</sup> (See Appendix F.)

## **Estrogen Compared With Placebo**

Within the network, there were 38 pairwise comparisons of placebo with estrogen—four with high-dose estrogen (three from poor-quality trials), 22 with standard dose (17 from poor-quality trials), and 12 with low/ultralow dose (eight from poor-quality trials). The standardized effect sizes for high, standard, and low doses of estrogen were 0.70 (95 percent CI: 0.40 to 1.01;  $\tau^2=0.06$ ; four comparisons), 0.71 (95 percent CI: 0.55 to 0.88;  $\tau^2=0.13$ ; 22 comparisons), and 0.40 (95 percent CI: 0.24 to 0.56;  $\tau^2=0.05$ ; 12 comparisons). A large effect size was estimated from one trial comparing standard-dose estrogen with placebo;<sup>149</sup> excluding it diminished the pooled SMD to 0.61 (Table 24). The funnel plot of the standard-dose estrogen–placebo comparison also exhibited asymmetry (Appendix G), but appeared attributable to three large trials<sup>26, 153, 154</sup> with reporting bias less likely. The mean ages of women in those trials were at the upper end of the distribution (62.8 to 63.6 years); excluding those trials yielded a symmetric funnel plot and an SMD of 0.76 (95 percent CI: 0.50 to 1.01;  $\tau^2=0.27$ ; 18 comparisons). The magnitudes of pooled standardized effect sizes for all dose categorizations of estrogen are large and the estimates are precise. While a majority of trials was rated as poor quality, the consistency over a large number of comparisons argues that the strength of evidence that estrogens of any dose improve quality-of-life scores among menopausal women is rated as high.

## **Estrogen Compared With Estrogen**

Six trials compared high-dose estrogens with standard-dose estrogens, two trials compared high-dose estrogens with low-dose estrogens, and eight trials compared standard-dose estrogens with low-dose estrogens (13 of 16 trials were rated as poor quality). Pooled estimates showed no differences between dose categories: standard versus high (SMD: 0.07; 95 percent CI: -0.05 to 0.18;  $\tau^2=0.01$ ; six comparisons); high versus low (SMD: -0.10; 95 percent CI: -0.29 to 0.09;  $\tau^2<0.0001$ ; two comparisons); and standard versus low (SMD: 0.13; 95 percent CI: -0.04 to 0.29;  $\tau^2=0.03$ ; eight comparisons). A substantial number of trials, although mostly of poor quality, consistently found no differences and had precise confidence intervals; the strength of evidence that reported quality-of-life scores do not differ by estrogen dose is rated as high.

## **Estrogen Compared With Isoflavones**

One trial compared standard-dose estrogens with isoflavones (SMD: 0.22; 95 percent CI: -0.25 to 0.70). The single trial was judged as poor quality, consistency is unknown, and the measure is direct, but imprecise. The strength of evidence that estrogens improve quality-of-life scores compared with isoflavones among menopausal women is rated as insufficient.

## **Isoflavones Compared With Placebo**

There were 20 pairwise comparisons of isoflavones with placebo (17 from poor-quality and three from good-quality trials). The standardized effect size was 0.38 (95 percent CI: 0.15 to 0.61;  $\tau^2=0.24$ ; 20 comparisons). However, two outliers were evident in forest, funnel, and radial plots,<sup>150, 151</sup> excluding those results diminished the pooled SMD substantially to 0.17 (95 percent CI: 0.06 to 0.29). The strength of evidence that isoflavones improve quality-of-life scores among menopausal women is rated as moderate.

## **SSRI/SNRI Compared With Placebo**

There were five pairwise comparisons of antidepressants with placebo (all from poor-quality trials). The standardized effect size was 0.27 (95 percent CI: 0.18 to 0.36;  $\tau^2=0.02$ ; five

comparisons). While the trials were rated as poor quality, results were consistent among the trials and the estimates were precise. The strength of evidence that antidepressants improve quality of life among menopausal women is rated as moderate.

**Table 22. Quality-of-life estimates of comparative efficacy as standardized effect sizes and 95 percent credible intervals from network meta-analysis. Treatments are ordered left to right from most to least efficacious. Highlighted effects are those where the credible interval does not overlap zero. The effects reflect improvement (lower on the scale) for the agent on the left versus comparators to its right from intersecting treatments listed on the diagonal.**

E-Standard										
0.04 (-0.19 to 0.27)	E-High									
0.16 (-0.04 to 0.35)	0.11 (-0.15 to 0.38)	E-Low/Ultralow								
0.34 (-0.01 to 0.70)	0.30 (-0.10 to 0.70)	0.19 (-0.18 to 0.56)	SSRI/SNRI							
0.40 (0.17 to 0.63)	0.36 (0.07 to 0.66)	0.25 (-0.01 to 0.50)	0.06 (-0.31 to 0.43)	Isoflavone						
0.61 (0.46 to 0.75)	0.56 (0.33 to 0.80)	0.45 (0.27 to 0.64)	0.26 (-0.06 to 0.59)	0.20 (0.02 to 0.39)	Placebo					

E: estrogen; SSRI: selective serotonin reuptake inhibitor; SNRI: serotonin and norepinephrine reuptake inhibitor.

**Table 23. Quality-of-life rankings of comparative efficacy, standard deviations, and 95% credible intervals (integer values because they arise from a distribution of integers).**

Treatment	Mean Rank	SD	95% CrI
E-Standard	3.4	0.6	(3-5)
E-High	3.9	0.8	(3-6)
E-Low/Ultralow	4.9	0.7	(3-6)
SSRI/SNRI	6.2	1.0	(4-8)
Isoflavones	6.6	0.6	(5-7)
Placebo	7.9	0.3	(7-8)

SD: standard deviation; CrI: credible interval; E: estrogen; SSRI: selective serotonin reuptake inhibitor; SNRI: serotonin and norepinephrine reuptake inhibitor.

**Table 24. Quality of life pairwise effect estimates (pooled random effect estimates or single trial effects if only data available).**

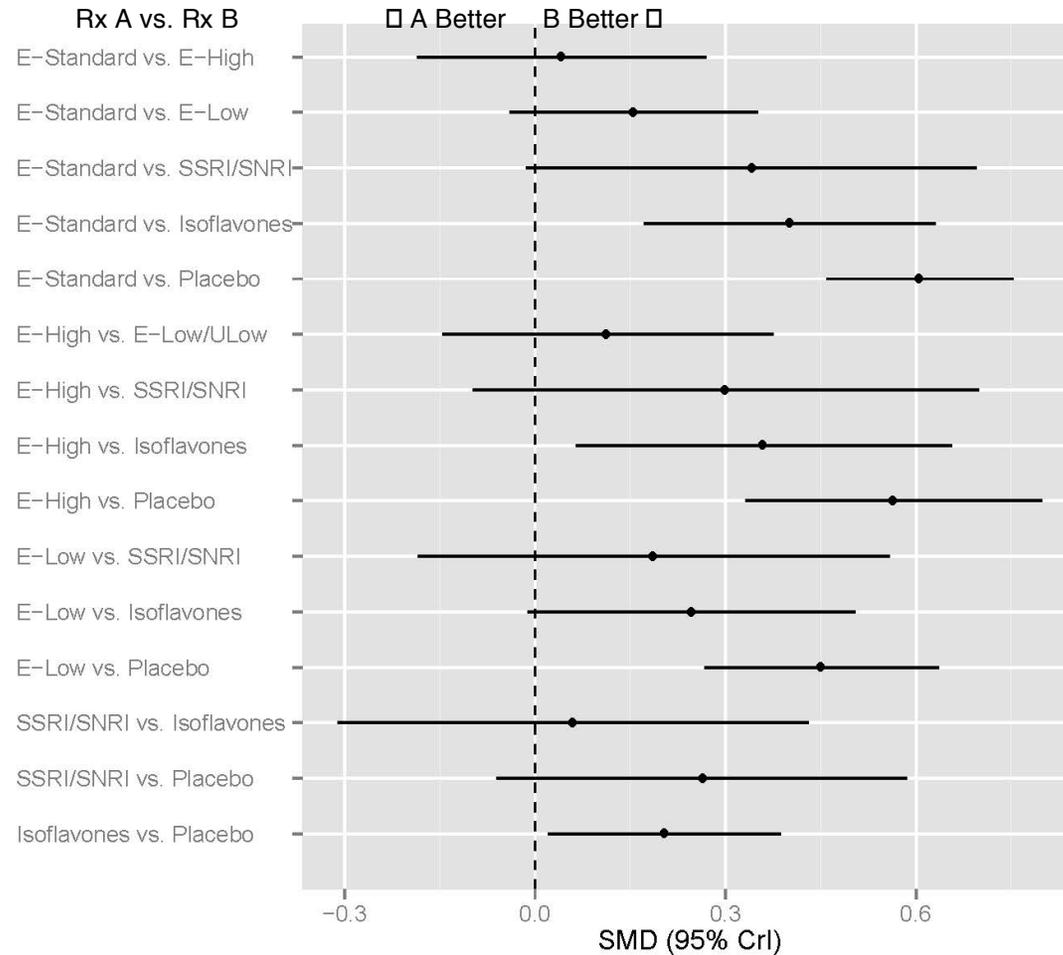
E-Standard					
0.07					
(-0.05 to 0.18)					
tau <sup>2</sup> =0.01; n=6					
E-High					
0.13					
(-0.04 to 0.29)					
tau <sup>2</sup> =0.03; n=8					
-0.10		E-Low/Ultralow			
(-0.29 to 0.09)		tau <sup>2</sup> <0.0001; n=2			
SSRI/SNRI					
0.22					
(-0.25 to 0.70)					
n=1					
Isoflavones					Placebo
0.63					
(0.47 to 0.78) <sup>a</sup>					
tau <sup>2</sup> =0.11; n=21					
0.70		0.40		0.27	
(0.40 to 1.01)		(0.24 to 0.56)		(0.18 to 0.36)	
tau <sup>2</sup> =0.06; n=4		tau <sup>2</sup> =0.05; n=12		tau <sup>2</sup> =0.00; n=5	
				0.17	
				(0.06 to 0.29) <sup>b</sup>	
				tau <sup>2</sup> =0.02; n=18	

<sup>a</sup> Including Baksu et al.<sup>149</sup> – SMD: 0.71 (95% CI: 0.55 to 0.88); tau<sup>2</sup>=0.13; 22 comparisons.

<sup>b</sup> Including Hidalgo et al.<sup>150</sup> and Han et al.<sup>151</sup> – SMD: 0.38 (95% CI: 0.15 to 0.61); tau<sup>2</sup>=0.24; 20 comparisons.

E: estrogen; SSRI: selective serotonin reuptake inhibitor; SNRI: serotonin and norepinephrine reuptake inhibitor.

**Figure 6. Caterpillar plot displaying all quality-of-life comparisons included in the network analysis and 95% credible intervals**



SD: standard deviation; CrI: credible interval; E: estrogen; SSRI: selective serotonin reuptake inhibitor; SNRI: serotonin and norepinephrine reuptake inhibitor; SMD: standardized mean difference; CrI: credible interval.

## Other Trials Pooled

### Black Cohosh Compared With Placebo

The three black cohosh trials (two of poor quality), with a total of 664 participants, reported significant improvements in quality-of-life scores for the treatment groups.<sup>132, 134, 152</sup> The pooled SMD was 0.40 (95 percent CI: 0.18 to 0.63; tau<sup>2</sup>=0.02). The strength of evidence that black cohosh improves quality of life scores is rated as low.

### DHEA Compared With Placebo

The three DHEA trials (two of poor quality), with a total of 365 participants, reported inconsistent results. Two trials of oral DHEA compared with placebo did not find significant differences in quality of life among study groups.<sup>136, 155</sup> One trial compared three different doses of DHEA in vaginal ovules with placebo and found improvements in quality-of-life scores with all three doses compared with placebo.<sup>156</sup> The strength of evidence that DHEA improves quality of life scores was rated insufficient.

### Ginseng Compared With Placebo

Three trials (two of poor quality) with a total of 513 participants, compared ginseng with placebo.<sup>101, 102</sup> yielding a pooled SMD of 0.19 (95 percent CI: 0.01 to 0.36; tau<sup>2</sup>=0.0). The strength of evidence that ginseng improves quality of life scores is rated as low.

## Trials Not Pooled

### Different Modes of Estrogen Administration

Seven trials compared similar estrogen doses administered through different modes.<sup>110, 112-115, 157, 158</sup> (See Appendix D for dose categorization by mode of administration). Three trials compared estrogen spray with estrogen patch, two compared oral estrogen with estrogen spray, one compared oral estrogen with estrogen patch, and one compared estrogen patches administered sequentially or combined. These trials were not included in the meta-analyses. Absent a factorial design, discerning treatment differences according to mode of administration in trials with differing doses and modes of administration is potentially problematic. None of the seven trials showed differences between the modes of administration. These results support the conclusion that mode of administration does not determine estrogen effectiveness. The strength of evidence that quality of life scores do not differ by mode of estrogen administration is rated as moderate.

**Table 25. Trials comparing different modes of estrogen administration, reporting quality-of-life outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	SMD (95% CI)
Odabasi, 2007 <sup>110</sup>	Estradiol + progestin	0.3 E + 90 P	32	Spray	12	Poor	—
	Estradiol + progestin	0.05 E + 90 P	29	Patch			NS
Davis, 2005 <sup>112</sup>	Estradiol	0.05	60	Patch	16	Poor	—
	Estradiol	0.3	60	Spray			NS
Ozsoy, 2002 <sup>113</sup>	Estradiol + MPA	2.0 E + 5.0 P	100	Oral	24	Poor	—
	Estradiol + MPA	0.3 E + 5.0 P	101	Spray			-0.28 (-0.56 to 0.00)

Lopes, 2001 <sup>114</sup>	Estradiol + progestin	0.05 E + 10 P	185	Patch	12	Poor	— -0.15 (-0.35 to 0.06)
	Estradiol + progestin	0.3 E + 10 P	176	Spray			
Mattsson, 2000 <sup>115</sup>	Estradiol + progestin	2.0 E + 10 P	342	Oral	24	Good	— -0.11 (-0.26 to 0.05)
	Estradiol + progestin	0.3 E + 10 P	317	Spray			
Lubbert, 1997 <sup>157</sup>	Estradiol	0.05	1232	Patch <sup>a</sup>	12	Poor	— NS
	Estradiol	0.05	1227	Patch <sup>b</sup>			
Polvani, 1991 <sup>158</sup>	CEE + MPA	0.625 E + 10 P	170	Oral	26	Poor	— -0.03 (-0.23 to 0.18)
	Estradiol + MPA	0.05 E + 10 P	203	Patch			

<sup>a</sup> Combined.

<sup>b</sup> Sequential.

SMD: standardized mean difference; CI: confidence interval; E: estrogen; P: progestin; CEE: conjugated equine estrogen; MPA: medroxyprogesterone acetate; FU: followup; Wks: weeks.

## Different Doses of Same Nonprescription Treatments

Two trials compared different doses of the same nonprescription treatments and reported quality-of-life outcomes.<sup>140, 146</sup> One trial compared two doses of isoflavones and reported significant improvements in quality of life in both groups and no between-group difference.<sup>146</sup> The other trial compared two doses of pueraria mirifica and also reported significant improvements in quality of life in both groups, with no difference between doses.<sup>140</sup>

**Table 26. Trials comparing different doses of the same nonprescription treatment, reporting quality-of-life outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	SMD (95% CI)
Hidalgo, 2006 <sup>146</sup>	Isoflavones	60	478	Oral	26	Poor	— 0.13 (0.00 to 0.26)
	Isoflavones	120		Oral			
Virojchaiwong, 2011 <sup>140</sup>	Pueraria mirifica	25	26	Oral	26	Poor	— 0.07 (-0.48 to 0.61)
	Pueraria mirifica	50	26	Oral			

SMD: standardized mean difference; CI: confidence interval; FU: followup; Wks: weeks.

## SSRI/SNRIs Compared

One trial compared two different antidepressants, desvenlafaxine and escitalopram, and reported quality-of-life outcomes.<sup>159</sup> The trial was of good quality and reported that both antidepressants improved quality-of-life scores significantly, without a difference between groups.

**Table 27. Trials comparing antidepressants, reporting quality-of-life outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	SMD (95% CI)
Soares, 2010 <sup>159</sup>	Desvenlafaxine	100-200	224	oral	8	good	— 0.16 (-0.02 to 0.35)
	Escitalopram	10-20	237	oral			

SMD: standardized mean difference; CI: confidence interval; FU: followup; Wks: weeks.

## Nonprescription Agents Compared with Placebo

Fourteen trials compared nonprescription treatments with placebo. Three trials compared DHEA with placebo,<sup>136, 155, 156</sup> two trials compared herbal extracts with placebo,<sup>120, 130</sup> and two trials compared flaxseed with placebo.<sup>137, 160</sup> St. John's wort,<sup>135</sup> rheum raphonticum,<sup>123</sup> pollen extract,<sup>124</sup> isoflavones combined with black cohosh,<sup>133</sup> dong quai,<sup>126</sup> a vitamin/mineral mixture,<sup>138</sup> and dioscorea alata<sup>128</sup> were compared with placebo in one trial each.

Of these trials comparing nonprescription agents with placebo, only one small trial (N=64) showed improvements in quality of life (SMD: 0.66; 95 percent CI: 0.16 to 1.15). Participants

were treated with a mixture of the following herbal extracts: *Cynanchum wilfordii*, *Phlomis umbrosa*, and *Angelica gigas*.<sup>130</sup>

**Table 28. Trials comparing nonprescription treatments, reporting quality-of-life outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	SMD (95% CI)
van der Sluijjs, 2009 <sup>120</sup>	Placebo Plant extracts <sup>a</sup>	— 3820	46 46	Oral Oral	16	Good	— -0.15 (-0.56 to 0.26)
Lewis, 2006 <sup>160</sup>	Placebo Flaxseed	— *	33 33	Oral Oral	16	Good	— NS
van Die, 2009 <sup>135</sup>	Placebo St John's wort	— 900	50 50	Oral Oral	16	Good	— -0.27 (-0.66 to 0.12)
Heger, 2006 <sup>123</sup>	Placebo Rheum rhaponticum	— 4	55 54	Oral Oral	12	Poor	— 0.38 (0.00 to 0.76)
Winther, 2005 <sup>124</sup>	Placebo Femal® <sup>b</sup>	— 80	32 32	Oral Oral	13	Good	— 0.15 (-0.34 to 0.64)
Verhoeven, 2005 <sup>133</sup>	Placebo Isoflavones/Black cohosh	— 50 I + 100 BC	64 60	Oral Oral	12	Good	— 0.01 (-0.34 to 0.37)
Hirata, 1997 <sup>126</sup>	Placebo Dong Quai	— 4,500	36 35	Oral Oral	24	Poor	— -0.06 (-0.52 to 0.41)
Hsu, 2011 <sup>128</sup>	Placebo Dioscorea alata	— 24	25 25	Oral Oral	52	Poor	— 0.31 (-0.25 to 0.86)
Labrie, 2009 <sup>156</sup>	Placebo DHEA DHEA DHEA	— 3.25 6.5 13.0	53 53 56 54	Ovule Ovule Ovule Ovule	12	Poor	— 0.65 (0.26 to 1.04) 0.35 (-0.03 to 0.72) 0.42 (0.04 to 0.80)
Panjari, 2009 <sup>155</sup>	Placebo DHEA	— 50	43 46	Oral Oral	26	Good	— 0.16 (-0.26 to 0.58)
Dodin, 2005 <sup>137</sup>	Placebo Flaxseed	— 40,000	94 85	Oral Oral	52	Fair	— 0.15 (-0.14 to 0.44)
Barnhart, 1999 <sup>136</sup>	Placebo DHEA	— 50	30 30	Oral Oral	12	Poor	— -0.05 (-0.56 to 0.46)
Andrikoula, 2011 <sup>138</sup>	Placebo Nutritional supplement <sup>c</sup>	— *	34 36	Oral Oral	12	Poor	— 0.12 (-0.35 to 0.59)
Chang, 2011 <sup>130</sup>	Placebo EstroG-100® <sup>d</sup>	— *	33 31	Oral Oral	12	Fair	— 0.66 (0.16 to 1.15)

SMD: standardized mean difference; CI: confidence interval; NS: not significant; DHEA: dehydroepiandrosterone; FU: followup; Wks: weeks.

<sup>a</sup> combination of black cohosh, er xian tang, zhi bai di huang wan

<sup>b</sup> combination of pure pollen, pollen/pistil extract

<sup>c</sup> combination of 21 vitamins and minerals

<sup>d</sup> combination of cynanchum wilfordii, phlomis umbrosa, angelica gigas

## Trials without Quantifiable or Poolable Data

Below is a description of six trials that did not have data that could be analyzed by the standardized effect size method or pooled because of the reporting metric. Results of these trials would not have affected the overall outcomes presented above.

The Estonian Postmenopausal Hormone Therapy Trial compared 0.625 mg combined estrogens plus 2.5 mg of medroxyprogesterone acetate daily orally with placebo.<sup>161</sup> Quality of life was measured using the EQ-5D developed by the EuroQol group. No baseline measures were reported. Post-treatment median EQ-5D scores showed no significant difference in quality of life among the trial groups.

A multicenter randomized trial (N=122) compared black cohosh with placebo.<sup>162</sup> Quality of life was reported as median Kupperman Index scores. The placebo group experienced a 17 percent improvement in quality of life and the black cohosh group experienced a 26 percent improvement.

A randomized blinded trial (N=152) compared two different doses of black cohosh and reported median Kupperman Index scores as a measure of quality of life.<sup>163</sup> Both black cohosh doses improved quality-of-life scores equally.

A randomized blinded trial (N=80) compared a combination of isoflavones, lignans, and *Cimicifuga racemosa* with placebo and reported median Kupperman Index scores as a measure of quality of life.<sup>164</sup> After three 28-day cycles of treatment, the quality of life in the treatment group improved significantly compared with the placebo group.

One trial compared a nonprescription treatment, *pueraria mirifica*, with a hormone treatment, conjugated equine estrogens, and reported quality-of-life outcomes in 60 women.<sup>139</sup> The study was of poor quality and did not show a significant difference in quality-of-life scores between the treatment groups.

A randomized open trial (N=70) compared isoflavones with placebo and reported median Kupperman Index scores as a measure of quality of life.<sup>165</sup> The authors reported that the quality of life significantly improved in the isoflavone group compared with the placebo group.

A randomized, double-blind trial (N=81) compared pomegranate seed oil, 30 mg taken twice daily, with placebo.<sup>129</sup> Quality-of-life outcomes were reported using the Menopause Rating Scale (MRS). The authors reported no significant difference in MRS scores between the trial groups after 12 weeks of followup.

## Strength of Evidence Ratings—Quality of Life

Table 29. Strength of evidence ratings domains for quality-of-life

Comparisons	Comparators <sup>a</sup>		Risk of Bias	Consistency	Directness	Precision	Reporting Bias	SOE	Downgrading Rationale
38	<b>Estrogen</b>	vs. Placebo	M	C	D	P	U	High	—
16	<b>Estrogen</b>	vs. <b>Estrogen (different dose)</b>	M	C	D	P	U	High	—
20	<b>Isoflavones</b>	vs. Placebo	M	C	D	P	S	Mod	After excluding outliers, effect size diminished by more than half
5	<b>SSRI/SNRI</b>	vs. Placebo	H	C	D	P	U	Mod	All trials rated poor quality
3	<b>Black Cohosh</b>	vs. Placebo	H	C	D	I	U	Mod	2 poor-quality trials
3	<b>DHEA</b>	vs. Placebo	H	I	D	I	U	Insuff	2 poor-quality trials; significant and insignificant results
3	<b>Ginseng</b>	vs. Placebo	H	C	D	I	U	Low	2 poor-quality trials; wide CI with lower bound 0.01
7	<b>Estrogen mode a</b>	vs. <b>Estrogen mode b</b>	H	C	D	P	U	Mod	6 poor quality trials

<sup>a</sup> Bold font of comparator indicates the more effective treatment; if both comparators are bold, the treatments are equivalently effective

Risk of Bias: High (H), Medium (M), Low (L); Consistency: Inconsistent (I), Unknown (U), Consistent (C); Directness: Indirect (I), Direct (D); Precision: Imprecise (I), Unknown (U), Precise (P); Reporting Bias: Suspected (S), Undetected (U); SOE: strength of evidence; Mod: moderate; CI: confidence interval; SSRI: selective serotonin reuptake inhibitor; SNRI: serotonin and norepinephrine reuptake inhibitor; DHEA: dehydroepiandrosterone; Insuff: insufficient.

### Key Points

- 108 trials including 56,497 women reported some measure of quality of life or well-being in women treated with prescription agents (estrogen, SSRIs, SNRIs) and nonprescription agents (isoflavones, black cohosh, vitamin E, flax seed, ginseng, and a variety of herbs and other agents).
- Three-fourths of trials were rated as poor quality and 51.8 percent were funded in whole or in part by industry.
- Results were reported from a variety of scales—a majority used menopause-specific scales.
- Data were most extensive for estrogens, isoflavones, SSRIs, and SNRIs. In a network meta-analysis, estrogens were associated with the highest reported scores (strength of evidence **high**; without difference between estrogen doses **moderate**; without difference between mode of estrogen administration **moderate**). Lower scores were seen with

SSRIs or SNRIs (strength of evidence **moderate**), and isoflavones (strength of evidence **moderate**).

- Strength of evidence of relative effectiveness of agents in improving measures of quality of life:
  - There is **high** strength of evidence that estrogen is the most effective agent in improving measures of quality of life. Combined results of trials that included a total of more than 35,000 women showed an SMD of 0.63 compared with placebo. There is **moderate** evidence that effects are similar for different estrogen dosages. Compared with standard-dose estrogen, direct comparisons (N=2214) of high-dose estrogen showed a combined SMD of -0.07 and low/ultralow-dose estrogen (N=2595) showed an SMD of 0.13.
  - Evidence strength is **moderate** that isoflavones, SSRIs or SNRIs, or black cohosh, improve quality of life measures. Combined results (N=2141) of trials of isoflavones showed an SMD of 0.17 compared with placebo. For SSRIs or SNRIs (N=3547), the combined SMD was 0.27 compared with placebo. For black cohosh (N=364), the combined SMD was 0.40 compared with placebo.
  - The strength of evidence ratings pertaining to other agents and comparators were either **low or insufficient**.
- Analyses comparing effectiveness of treatments show estrogens improve quality of life symptoms best, with the following mean rankings: standard dose estrogens (3.4), high dose estrogens (3.9), and low dose estrogens (4.9). The nonhormone treatments were ranked much lower: SSRI/SNRI (6.2) and isoflavones (6.6).

# Psychological Symptoms

## Included Trials

Of the 254 trials included in this CER, 90 (35.4 percent) trials reported psychological outcomes in three domains: global, anxiety, and depression (45 trials specified at least one as a primary outcome). Many trials reported outcomes in more than one domain: global (n=50), anxiety (n=42), and depression (n=51). Forty-eight trials examined hormone treatment effects compared with: placebo (30 trials), other hormones (13 trials), other prescription treatments (three trials), and nonprescription treatments (two trials). Thirty-six trials examined nonprescription treatment effects compared with: placebo (31 trials), hormones (two trials), other nonprescription agents (two trials), and antidepressants (one trial). Nine trials examined antidepressants employing a placebo comparator in eight trials, and a nonhormone comparator in one trial.

The 90 trials originated from 23 different countries and 11 trials were multinational. Trials were conducted in the United States (n=16), United Kingdom (n=7), and six each from Turkey, Italy, and Canada. Other countries contributing evidence to this section include but are not limited to: Australia, Germany, Taiwan, Hong Kong, Denmark, Brazil, Austria, Sweden, India, and Finland. The trials were conducted in 1,909 sites. Length of followup ranged from four to 156 weeks.

Psychological outcomes were reported using a variety of scales. The most common scales were: Greene (10 anxiety, 10 depression, 13 global), WHQ (eight anxiety, seven depression, one global), MENQOL (13 global), Beck (four anxiety, eight depression), Hamilton (five anxiety, seven depression), SF-36 (nine global), and Kupperman (four anxiety, four depression). Additional scales used include CES-D, Hospital Anxiety and Depression Scale, Psychological General Well-Being, MRS, Profile of Mood States, and the Bond and Lader Mood Rating Scale. The following are brief descriptions of the most commonly used scales:

- The Greene anxiety subscale consists of six items, with scores ranging from 0 to 18.<sup>166</sup> Questions include heart beating quickly and strongly, feeling tense or nervous, difficulty sleeping, excitable, attacks of panic, and difficulty concentrating. The Greene depression subscale consists of five items, with scores ranging from 0 to 15. Questions include feeling tired or lacking in energy, loss of interest in most things, feeling unhappy or depressed, crying spells, and irritability. Total psychological scores range from 0 to 33. Higher scores indicate more severe symptoms.
- The WHQ can be administered as a 23- or 37-item instrument. The 37-item version includes four items in the anxiety assessment: I get very frightened or panic feelings for apparently no reason at all, I feel anxious when I go out of the house on my own, I get palpitations or a sensation of “butterflies” in my stomach or chest, and I feel tense or “wound up.” The depression score includes seven items: I feel miserable and sad, I have lost interest in things, I still enjoy the things I used to, I feel life is not worth living, I have a good appetite, I am more irritable than usual, and I have feelings of well-being. Total scores on subscales are 0 to 1 (some scales reversed according to the construct probed).
- The MENQOL psychosocial score is derived from seven items (scored 1 to 8): being dissatisfied with my personal life; feeling anxious or nervous; experiencing poor memory (no or yes); accomplishing less than I used to; feeling depressed, down, or blue; being impatient with other people; and feelings of wanting to be alone.<sup>89</sup>

- The Beck anxiety inventory and Beck depression inventory each include 21 items and total scores range from 0 to 63. The Beck anxiety inventory lists symptoms common to anxiety such as numbness, heart pounding, trembling, shaking, indigestion, and flushing.<sup>167</sup> The Beck depression inventory assesses mood, satisfaction, appetite, sleep, weight, and sexual activity.<sup>168</sup>
- The Hamilton scales are completed by a health care professional following an examination of the patient. This scale measures both mental distress as well as physical complaints related to anxiety and depression.<sup>169, 170</sup> The Hamilton anxiety score consists of 14 items with a total score of 0 to 56. The depression scale consists of 21 items with a total score of 0 to 52.
- The SF-36 mental health score consists of five items. The items assess nervousness, cheerfulness, peacefulness, depression, and happiness. Scores are summed, then normalized to a 0-100 scale.<sup>171</sup>
- Kupperman measures insomnia, nervousness, and melancholia.<sup>172</sup> Total scores range from 0 to 16 summed. Hospital Anxiety & Depression Scale (HADS) includes 14 items (seven depression and seven anxiety). The Psychological General Well Being is a 22-item derivative of the General Well Being Index Menopause Rating Scale.

In many cases, the presence of climacteric symptoms and/or anxious depressive disorders was required for inclusion in the study. However, women were often excluded if taking psychoactive drugs, had too high of a score on the assessment tool, or had suicidal thoughts. Table 30 further describes the trial and patient characteristics.

**Table 30. Characteristics of trials assessing efficacy for psychological symptoms**

	Characteristic	Value
	Number of trials	90
	Total number of patients	48,894
	Total sites from trials that specified site # (n=76)	1,909 1 to 502 (mean: 26; median: 2)
Trial Characteristics	Trials described only as multicenter	4 (4.4)
	Multicenter trials	46 (51.1)
	Two arm trials	70 (77.8)
	Multi-arm trials	20 (22.2)
	Patients per trial	50 to 16,608 (mean: 543; median:119)
	Range of followup (weeks)	4 to 156 (mean: 26.7; median: 16)
Funding	Industry only	35 (38.9)
	Public only	15 (16.7)
	Industry and public	9 (10.0)
	Not stated	31 (34.4)
Comparator Category	Placebo vs. hormone	30 (33.3)
	Antidepressant vs. placebo or other antidepressant	8 (8.9)
	Placebo vs. other prescription	3 (3.3)
	Placebo vs. nonprescription	31 (34.4)

	Placebo vs. hormone vs. nonprescription	0 (0.0)
	Hormone vs. hormone	13 (14.4)
	Hormone vs. nonprescription	2 (2.2)
	Nonprescription vs. nonprescription	2 (2.2)
	Nonprescription vs. antidepressant	1 (1.1)
Study Quality	Good	14 (15.6)
	Fair	8 (8.9)
	Poor	68 (75.6)
Patient Demographics	Mean age (years)	46.8 to 75.6 (NR: 12)
	Age range (years)	29 to 85 (NR: 75)
	Years since menopause	3.9 (0.8 to 18.6) (NR: 61)
	Current smokers (%)	0.0 to 41.2 (NR: 67)
	Mean BMI (kg/m <sup>2</sup> )	17.3 to 30.1 (NR: 34)
	White (%)	0.0 to 100.0
	Black (%)	0.0 to 15.1
	Hispanic (%)	0.0 to 9.0
	Asian (%)	0.0 to 100.0
	Other (%)	0.0 to 23.3
Uterus Status	All intact	25 (27.8)
	All absent	6 (6.7)
	Mixed	36 (40.0)
	Range, percentage intact among trials with mixed	25% to 94.3%
	Not reported	23 (25.6)

NR: not reported

Note: Demographics were not reported in all studies.

## Evidence Synthesis for Psychological Symptoms

Standard mean differences were calculated to allow comparison of outcomes across different psychological symptom scales. Analyses were separated by psychological domain: anxiety, depression, and global mental health. Pooling of the following comparators was performed: estrogens (high, standard, and low doses) versus placebo, isoflavones versus placebo, antidepressants versus placebo, gabapentin/pregabalin versus placebo, and black cohosh versus placebo. An analysis for estrogen doses combined was not performed because of suggested differences in effect between doses. Table 31 displays psychological outcomes pairwise effect estimates. Figure 7 is a caterpillar plot displaying all depression comparisons and 95 percent confidence intervals; Figure 8 is a caterpillar plot of all anxiety outcomes; and Figure 9 is a caterpillar plot of all global mental health comparisons. Forest plots are displayed in Appendix H.

## High-Dose Estrogens Compared With Placebo

### Depression

Four trials compared high dose estrogens with placebo and reported depression as an outcome (three poor quality trials)<sup>159, 173-175</sup> Pooling analyses found high dose estrogens significantly improved depression compared with placebo (SMD: -0.64; 95 percent CI: -0.94 to -0.33;  $\tau^2=0.06$ ; 4 comparisons). The strength of evidence that high dose estrogens improve depressive symptoms compared with placebo is rated as moderate.

### Anxiety

Two trials compared high dose estrogens with placebo and reported an anxiety outcome (one poor-quality trial).<sup>173, 174</sup> Pooled results found significantly improved symptoms of anxiety compared with placebo (SMD: -0.35; 95 percent CI: -0.58 to -0.13;  $\tau^2=0.01$ ; two comparisons). The strength of evidence that high dose estrogens improve anxiety symptoms compared with placebo is rated as low.

### Global Mental Health

One trial compared high-dose estrogens with placebo and reported a global mental health score as an outcome (fair quality).<sup>174</sup> The high dose estrogen arm showed significant improvement in global mental health scores compared with placebo (SMD: -0.42; 95 percent CI: -0.65 to -0.19; 1 comparison).

## Standard-Dose Estrogens Compared With Placebo

### Depression

Eleven trials compared standard dose estrogens with placebo and reported depression as an outcome (nine poor-quality trials).<sup>26, 154, 176-184</sup> Pooled analyses showed modest but significant improvements in depressive symptoms in the estrogen arm (SMD: -0.19; 95 percent CI: -0.31 to -0.07;  $\tau^2=0.02$ ; 11 comparisons). However, four the trials showed SMDs between -0.06 and 0.12.<sup>26, 154, 178, 183</sup> The strength of evidence that standard dose estrogens improve depressive symptoms compared with placebo is rated as moderate.

### Anxiety

Eight trials compared standard dose estrogens with placebo and reported an anxiety outcome (seven poor-quality trials).<sup>26, 176, 179-181, 183-185</sup> Pooled analyses did not detect a difference between standard dose estrogens and placebo (SMD: -0.16; 95 percent CI: -0.34 to 0.03;  $\tau^2=0.04$ ; 8 comparisons). Three trials found nonsignificant effects in the opposite direction.<sup>26, 181 65, 184</sup> The strength of evidence that standard-dose estrogens improve anxiety symptoms compared with placebo is rated as insufficient.

### Global Mental Health

Nine trials compared standard dose estrogens with placebo and reported global mental health as an outcome (seven poor-quality studies).<sup>153, 154, 176, 177, 181, 184, 186-188</sup> Pooled analyses did not detect a difference in global mental health between standard dose estrogens and placebo (SMD: -0.03; 95 percent CI: -0.10 to 0.04;  $\tau^2=0.00$ ; nine comparisons). The strength of evidence that standard dose estrogens improve global mental health scores compared with placebo is rated as insufficient.

## Low/Ultralow-Dose Estrogens Compared With Placebo

### Depression

Three trials compared low/ultralow dose estrogens with placebo and reported depression as an outcome (all poor quality).<sup>189-191</sup> Pooled analyses did not find an improvement in depression for those in the treatment arm (SMD: -0.04; 95 percent CI: -0.40 to 0.31;  $\tau^2=0.02$ ; 3 comparisons). The strength of evidence that low/ultralow dose estrogens improve depression compared with placebo is rated as insufficient.

### Anxiety

Three trials compared low/ultralow-dose estrogens with placebo and reported an anxiety outcome (all of poor quality).<sup>188, 190, 191</sup> Pooled analyses found no difference with low/ultralow estrogen (SMD: -0.19; 95 percent CI: -0.41 to 0.02;  $\tau^2=0.00$ ; three comparisons). The strength of evidence that low/ultralow dose estrogens improve depression compared with placebo is rated as low.

### Global Mental Health

Seven trials compared standard dose estrogens with placebo and reported global mental health outcomes (two good, two fair, and three poor-quality trials).<sup>87, 186-188, 192-194</sup> Pooled analyses found a significant difference in global mental health between low/ultralow-dose estrogens and placebo (SMD: -0.24; 95 percent CI: -0.45 to -0.02;  $\tau^2=0.06$ ; seven comparisons). The strength of evidence that standard dose estrogens improve global mental health scores compared with placebo is rated as high.

## Isoflavones Compared With Placebo

### Depression

Eight trials compared isoflavones with placebo and reported depression as an outcome (seven poor-quality trials).<sup>150, 195-201</sup> Pooled analyses showed a significant improvement in depression among the group treated with isoflavones compared with the placebo group (SMD: -0.41; 95 percent CI: -0.69 to -0.13;  $\tau^2=0.12$ ; eight comparisons). Four of the trials, including the two largest<sup>196, 197</sup> showed standardized effect sizes close to 0, whereas three of the smallest<sup>150, 199, 200</sup> showed large standardized effect sizes (-0.66 to -1.20) indicating possible reporting bias. The strength of evidence that isoflavones improve depression compared with placebo is rated as low.

### Anxiety

Seven trials compared isoflavones with placebo and reported anxiety as an outcome (six poor-quality trials).<sup>150, 195, 196, 199-202</sup> Pooled estimates showed a significant improvement in anxiety among the group treated with isoflavones compared with the placebo group (SMD: -0.53; 95 percent CI: -0.87 to -0.20;  $\tau^2=0.16$ ; seven comparisons). The strength of evidence that isoflavones improve anxiety compared with placebo among menopausal women is rated as low.

### Global Mental Health

Six trials compared isoflavones with placebo and reported global mental health as an outcome (four good-quality, two poor-quality trials).<sup>160, 196, 197, 201, 203, 204</sup> Pooled estimates show no significant difference in global mental health among the two trial groups (SMD: -0.12; 95

percent CI: -0.26 to 0.01;  $\tau^2=0.00$ ; six comparisons). The strength of evidence that isoflavones improve global mental health compared with placebo among menopausal women is rated as low.

## **SSRI/SNRI Compared With Placebo**

### **Depression**

Three trials compared antidepressants with placebo and reported depression as an outcome (all poor-quality trials).<sup>205-207</sup> One trial used fluoxetine,<sup>205</sup> and two trials used desvenlafaxine.<sup>206, 207</sup> Pooled estimates show that antidepressants improved depression in the treatment group compared with the placebo group (SMD: -0.40; 95 percent CI: -0.59 to -0.22;  $\tau^2=0.01$ ; 3 comparisons). From these three trials, the strength of evidence that antidepressants improve depression compared with placebo was rated moderate.

### **Anxiety**

Two trials compared antidepressants with placebo and reported anxiety as an outcome (both poor-quality trials).<sup>206, 207</sup> Both trials used desvenlafaxine. Pooled estimates show that antidepressants improved anxiety in the treatment group compared with the placebo group (SMD: -0.31; 95 percent CI: -0.53 to -0.08;  $\tau^2=0.01$ ; two comparisons). The strength of evidence that antidepressants improve anxiety compared with placebo is rated as low.

### **Global Mental Health**

Four trials compared antidepressants with placebo and reported mental health as an outcome (all poor-quality trials).<sup>207-210</sup> One trial studied venlafaxine,<sup>208</sup> one citalopram,<sup>209</sup> and two trials desvenlafaxine.<sup>207, 210</sup> Pooled estimates showed that antidepressants improved global mental health in the treatment group compared to the placebo group (SMD: -0.39; 95 percent CI: -0.63 to -0.15;  $\tau^2=0.04$ ; four comparisons). The strength of evidence that antidepressants improve overall mental health compared with placebo among is rated as moderate.

## **Gabapentin/Pregabalin Compared With Placebo**

### **Global Mental Health**

Two trials compared gabapentin/pregabalin with placebo and reported global mental health as an outcome (both poor-quality trials).<sup>73, 211</sup> Pooled estimates show that gabapentin/pregabalin did not significantly improve depression compared with placebo (SMD: -0.22; 95 percent CI: -0.46 to 0.03;  $\tau^2=0.00$ ; two comparisons). The strength of evidence that gabapentin/pregabalin improve global mental health compared with placebo is rated as insufficient.

## **Estrogen Compared With Nonprescription**

### **Anxiety**

One randomized trial compared black cohosh with an ultralow-dose estrogen/progestin patch and reported anxiety outcomes.<sup>212</sup> Both treatments significantly improved anxiety ( $p<0.001$  for both arms of the trial). There was no significant difference between the treatments (SMD: -0.09; 95 percent CI: -0.57 to 0.39).

## **Black Cohosh Compared with Placebo**

### **Global Mental Health**

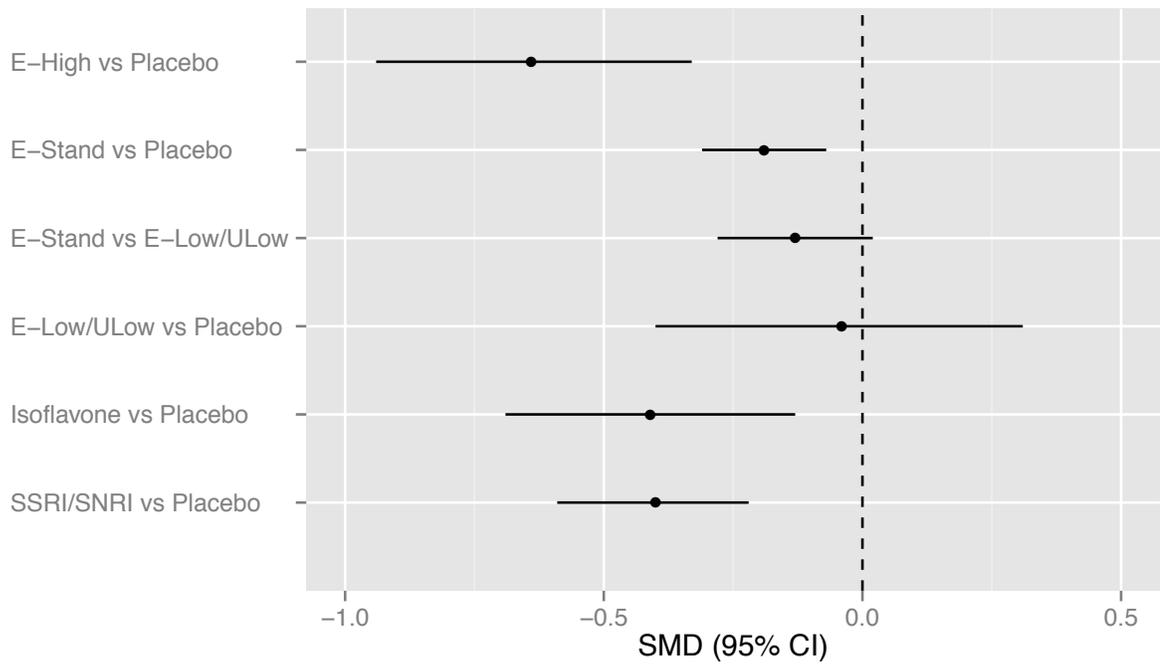
One trial compared black cohosh with placebo and reported a global mental health outcome.<sup>152</sup> Black cohosh was reported to improve global mental health (SMD: -0.59, 95 percent CI: -1.09 to -0.10).

**Table 31. Psychological outcomes pairwise effect estimates (pooled random effect estimates or single trial effects if only data available)**

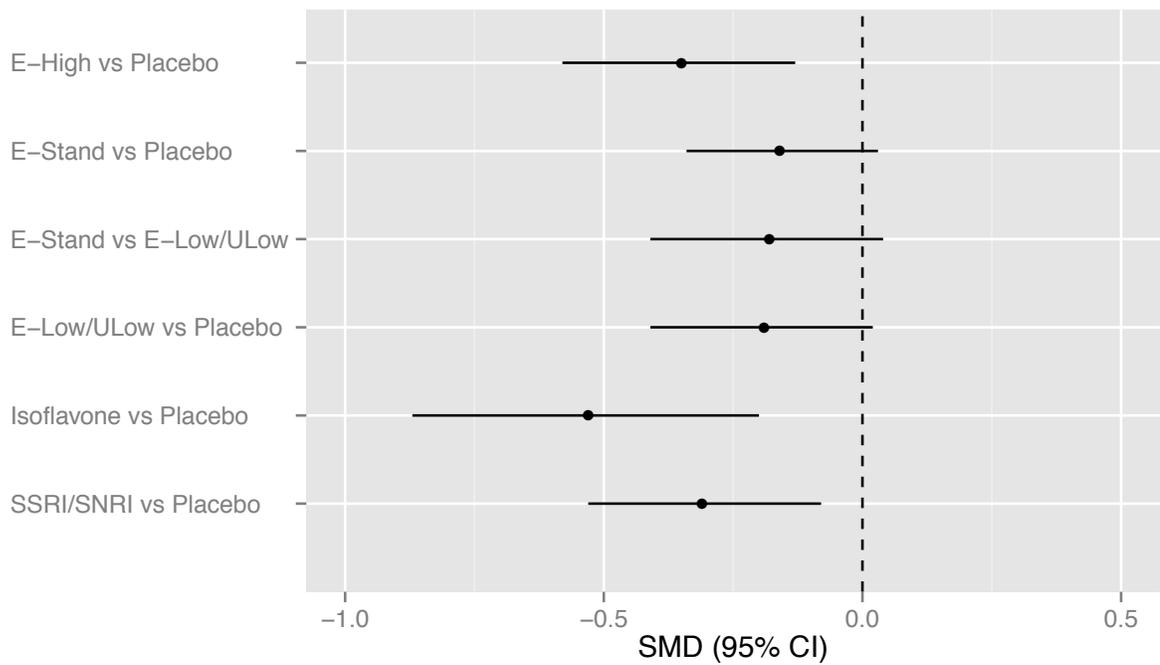
		E-High									
Depression		NS									
Anxiety		-0.09 (-0.28 to 0.10) n=1		E-Standard							
Global		0.17 (-0.03 to 0.38) n=1									
Depression		0.00 (-0.31 to 0.31) n=1		-0.13 (-0.28 to 0.02) tau <sup>2</sup> =0.00; n=2							
Anxiety		0.32 (0.00 to 0.63) n=1		-0.18 (-0.41 to 0.04) tau <sup>2</sup> =0.00; n=2		E-Low/Ultralow					
Global		0.14 (-0.17 to 0.45) n=1		-0.02 (-0.29 to 0.24) tau <sup>2</sup> =0.08; n=6							
Isoflavone											
SSRI/SNRI											
Gabap/Preg											
Depression		0.09 (-0.39 to 0.58) n=1				-0.85 (-1.30 to -0.40) n=1					
Anxiety		-0.09 (-0.57 to 0.39) n=1								Black Cohosh	
Global						0.09 (-0.34 to 0.53) n=1					
Depression		-0.64 (-0.94 to -0.33) tau <sup>2</sup> =0.06; n=4		-0.19 (-0.31 to -0.07) <sup>a</sup> tau <sup>2</sup> =0.02; n=11		-0.04 (-0.40 to 0.31) tau <sup>2</sup> =0.02; n=3		-0.41 (-0.69 to -0.13) <sup>d</sup> tau <sup>2</sup> =0.13; n=8		-0.40 (-0.59 to -0.22) tau <sup>2</sup> =0.01; n=3	
Anxiety		-0.35 (-0.58 to -0.13) tau <sup>2</sup> =0.01; n=2		-0.16 (-0.34 to 0.03) <sup>b</sup> tau <sup>2</sup> =0.04; n=8		-0.19 (-0.41 to 0.02) tau <sup>2</sup> =0.00; n=3		-0.53 (-0.87 to -0.20) <sup>e</sup> tau <sup>2</sup> =0.16; n=7		-0.31 (-0.53 to -0.08) tau <sup>2</sup> =0.02; n=2	
Global		-0.42 (-0.65 to -0.19) n=1		-0.03 (-0.10 to 0.04) tau <sup>2</sup> =0.00; n=9		-0.24 (-0.45 to -0.02) <sup>c</sup> tau <sup>2</sup> =0.06; n=7		-0.12 (-0.26 to 0.01) tau <sup>2</sup> =0.00; n=6		-0.39 (-0.63 to -0.15) tau <sup>2</sup> =0.04; n=4	
								-0.22 (-0.46 to 0.03) tau <sup>2</sup> =0.00; n=2		-0.59 (-1.09 to -0.10) n=1	
Placebo											

a including outlier Baksu 2182, -0.39 (-0.60 to -0.18)  $\tau^2=0.10$ ; n=12  
b including outlier Baksu 2182, -0.27 (-0.51 to -0.02)  $\tau^2=0.11$ ; n=9  
c excluding Baeuug 1920, -0.15 (-0.26 to -0.04)  $\tau^2=0.00$ ; n=6  
d excluding Lipovac 2172, -0.28 (-0.49 to -0.07)  $\tau^2=0.04$ ; n=7  
e excluding Lipovac 2172, -0.39 (-0.64 to -0.15)  $\tau^2=0.06$ ; n=6

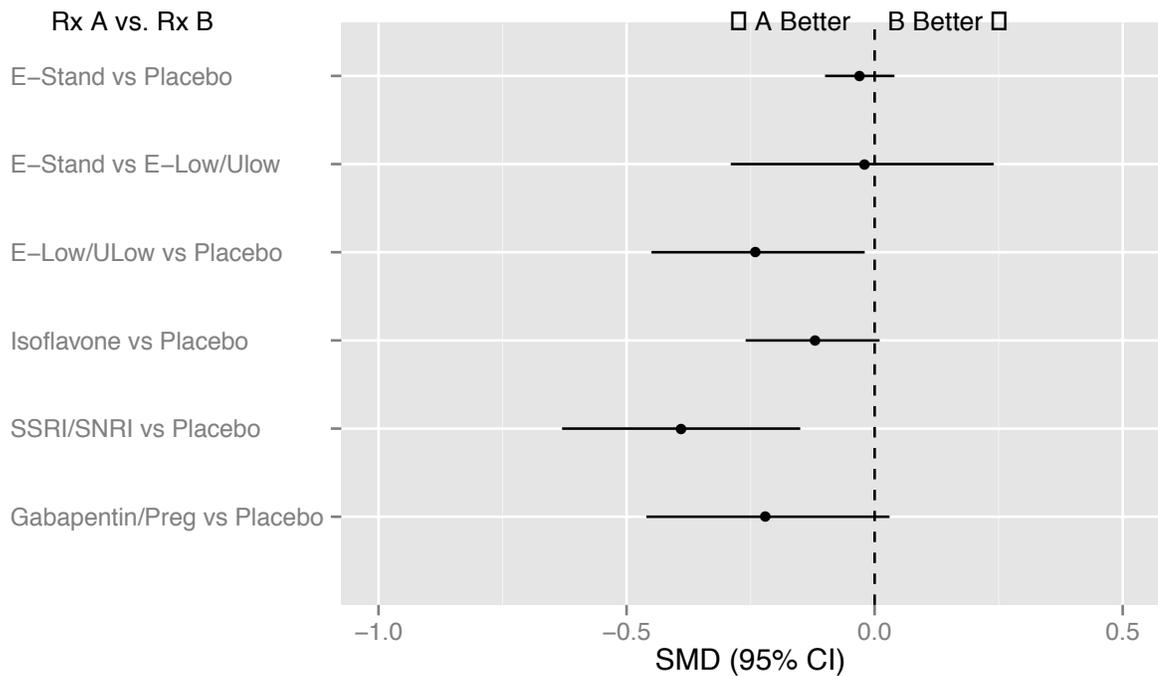
**Figure 7. Caterpillar plot displaying all depression comparisons and 95 percent confidence intervals**



**Figure 8. Caterpillar plot displaying all anxiety comparisons and 95 percent confidence intervals**



**Figure 9. Caterpillar plot displaying all global mental health comparisons and 95 percent confidence intervals**



## **Trials Not Pooled**

### **Different Modes of Estrogen Administration**

Three trials (Table 32) compared similar doses of estrogen administered through different modes. (See Appendix D for dose categorization by mode of administration). One trial compared oral estrogen with estrogen patch,<sup>111</sup> one compared estrogen patches administered with sequentially or combined progestin,<sup>157</sup> and a three-arm trial compared oral estrogen with estrogen skin gel and with estrogen patch.<sup>144</sup> The three-arm trial was rated as poor quality and reported that both the skin gel and the patch improved global mental health scores (SMD: -0.77; 95 percent CI: -1.30 to -0.25 and SMD: -1.07; 95 percent CI: -1.59 to -0.54, respectively). The other two trials, also of poor quality, found no difference in psychological outcomes between the different modes.

One trial compared a low-dose oral estrogen (n=75) with a high-dose estrogen vaginal ring (n=84) and reported Greene anxiety score as an outcome.<sup>213</sup> The trial was rated as fair quality and Buckler et al. found that both oral and vaginal ring treatments significantly improved anxiety. There was no significant difference between the two modes of administration (SMD: 0.32; 95 percent CI: 0.00 to 0.63).

Given the different treatments and outcomes, the strength of evidence was not rated.

**Table 32. Trials comparing different modes of estrogen administration reporting psychological outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	Outcome	SMD (95% CI)
Lubbert, 1997 <sup>157</sup>	Estradiol + progestin	0.05 E + P	1232	patch	12	poor	D	—
	Estradiol + progestin	0.05 E + P	1227	cont. patch cycl.				-0.01 (-0.09 to 0.06)
Buckler, 2003 <sup>213</sup>	Estradiol + NETA	1 E + 1 P	75	oral	24	fair	A	—
	Estradiol + NETA	0.05 1 P	84	vaginal ring			G	0.32 (0.00 to 0.63)
Akhila, 2006 <sup>144</sup>	CEE + progestin	0.625 E + 2.5 P	35	oral	52	poor	G	—
	Estradiol + progestin	1.5 E + 2.5 P	25	skin gel				-0.77 (-1.30 to -0.25)
	Estradiol + progestin	0.05 E + 2.5 P	28	patch				-1.07 (-1.59 to -0.54)
Serrano, 2006 <sup>111</sup>	CEE + MPA	0.625 E + 10 P	55	oral	52	poor	G	—
	Estradiol + MPA	0.05 E + 10 P	59	patch				-0.09 (-0.46 to 0.27)

SMD: standardized mean difference; CI: confidence interval; E: estrogen; P: progestin; NETA: norethisterone acetate; CEE: conjugated equine estrogen; MPA: medroxyprogesterone acetate; cont.: continuous; cycl: cycling; D: depression; A: anxiety; G: global mental health; FU: followup; Wks: weeks.

### Estrogen Compared With Estrogen Plus Testosterone

One trial (Table 33) compared an estrogen/progestin skin gel (n=53) with an estrogen/progestin plus testosterone skin gel (n=53) and reported depression, anxiety, and global mental health outcomes using the Psychological General Well-Being scale.<sup>214</sup> The trial was rated as poor quality and reported no difference between groups in depression scores or total scores. A significant improvement was reported in anxiety scores in the testosterone group (SMD: -0.65; 95 percent CI: -1.04 to -0.26).

**Table 33. Trials comparing estrogen with estrogen plus testosterone reporting psychological outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	Outcome	SMD (95% CI)
Nathorst-Boos, 2006 <sup>214</sup>	Estrogen + progestin Estrogen + progestin + testosterone	NR 10 T	53 53	skin gel skin gel	26	poor	D	—
							A	-0.17 (-0.55 to 0.21)
							G	-0.65 (-1.04 to -0.26)
							G	-0.45 (-0.83 to -0.07)

SMD: standardized mean difference; CI: confidence interval; NR: not reported; T: testosterone; D: depression; A: anxiety; G: global mental health; FU: followup; Wks: weeks.

### Progestin Alone Compared With Placebo

One trial (Table 34) compared four different progestin skin cream doses (5 mg, 20 mg, 40 mg, and 60 mg) with placebo skin cream and reported Greene psychological scores.<sup>105</sup> The trial was rated as fair quality and found no significant difference between any of the doses of progestin skin cream compared with placebo.

**Table 34. Trials comparing progestin alone with placebo reporting psychological outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	Outcome	SMD (95% CI)
Benster, 2009 <sup>105</sup>	Placebo	—	43	skin cream				—
	Progestin	5	46	skin cream				0.04 (-0.38 to 0.45)
	Progestin	20	44	skin cream	24	fair	G	0.11 (-0.31 to 0.53)
	Progestin	40	43	skin cream				-0.07 (-0.49 to 0.35)
	Progestin	60	45	skin cream				0.07 (-0.34 to 0.49)

SMD: standardized mean difference; CI: confidence interval; G: global mental health; FU: followup; Wks: weeks.

### Prescription Compared With Placebo

One randomized double-blind trial (Table 35) compared eszopiclone, a treatment used for insomnia (n=30), with placebo (n=29) and reported the Beck anxiety score as an outcome.<sup>108</sup> The trial was rated poor quality and found a significant improvement in anxiety among the treatment group with a wide confidence interval (SMD: -0.57; 95 percent CI: -1.09 to -0.06).

**Table 35. Trials comparing prescription treatments with placebo reporting psychological outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	Outcome	SMD (95% CI)
Joffe, 2010 <sup>108</sup>	Placebo	—	29	oral	4	poor	A	—
	Eszopiclone	3	30	oral				-0.57 (-1.09 to -0.06)

SMD: standardized mean difference; CI: confidence interval; G: global mental health; FU: followup; Wks: weeks.

### Nonprescription Agents Compared with Placebo

Seventeen trials (Table 36) compared various nonprescription agents with placebo and reported 30 psychological outcomes (depression [n=9], anxiety [n=10], and global mental health [n=11]). Four trials studied mixed herbal extracts,<sup>102, 118, 130, 131</sup> three used DHEA,<sup>136, 155, 156</sup> and one trial studied a combination of vitamins and minerals.<sup>138</sup> One three arm trial compared isoflavones, flaxseed, and placebo.<sup>160</sup> The remaining trials used various plant extracts: green tea,<sup>215</sup> pine extract,<sup>121</sup> rheum rhaponticum,<sup>123</sup> ginseng,<sup>101</sup> dioscorea alata,<sup>128</sup> and St. John's wort.<sup>134, 135</sup> Eleven trials were rated of poor quality, one as fair quality, and five as good quality.

Most of the trials did not detect significant improvements in psychological symptoms. Trials reporting significant improvements were: Chang et al.—improvements in depression and anxiety with herbal extracts (SMD: -0.91; 95 percent CI: -1.82 to -0.40 and SMD: -1.04; 95 percent CI: -1.56 to -0.52)<sup>130</sup>; Hsu et al—improvements in global mental health with dioscorea alata (SMD: -0.75; 95 percent CI: -1.32 to -0.19)<sup>128</sup>; Labrie et al—inconsistent improvements in global mental health with different doses of vaginal DHEA.<sup>156</sup>; Yang et al—improvements in depression and anxiety with maritime pine extract (SMD: -0.41; 95 percent CI: -0.73 to -0.09 and SMD: -0.81; 95 percent CI: -1.13 to -0.48)<sup>121</sup>; Mucci—improvements in depression and anxiety with a combination of isoflavones and magnolia bark (SMD: -1.00; 95 percent CI: -1.44 to -0.57 and SMD: -0.96; 95 percent CI: -1.39 to -0.52)<sup>122</sup>; and Heger et al—improvements in anxiety and global mental health with rheum rhaponticum (SMD: -2.30; 95 percent CI: -2.78 to -1.82 and SMD: -0.50; 95 percent CI: -0.88 to -0.12).<sup>123</sup>

**Table 36. Trials comparing nonprescription agents with placebo reporting psychological outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	Outcome	SMD (95% CI)	
Xia, 2012 <sup>131</sup>	Placebo	—	36	Oral	12	Good	G	—	
	Jiawei Qing'e Fang	3500	36	Oral			-0.41 (-0.87 to 0.05)		
Chang, 2011 <sup>130</sup>	Placebo	—	33	Oral	12	Fair	D	-0.91 (-1.82 to -0.40)	
	EstroG-100® <sup>a</sup>	257	31	Oral			A	-1.04 (-1.56 to -0.52)	
Andrikoula, 2011 <sup>138</sup>	Placebo	—	34	Oral	12	Poor	A	—	
	Nutritional supplement <sup>b</sup>	*	36	Oral			D	0.17 (-0.30 to 0.63)	
Hsu, 2011 <sup>128</sup>	Placebo	—	25	Oral	52	Poor	A	—	
	Dioscorea alata	24	25	Oral			G	0.03 (-0.52 to 0.58)	
Shen, 2010 <sup>215</sup>	Placebo	—	44	Oral	24	Poor	G	—	
	Green tea polyphenols	500	47	Oral			-0.14 (-0.55 to 0.26)		
Labrie, 2009 <sup>156</sup>	Placebo	—	53	Ovule	12	Poor	G	—	
	DHEA	3.25	53	Ovule				-0.78 (-1.17 to -0.39)	
	DHEA	6.5	56	Ovule				-0.36 (-0.74 to 0.01)	
Panjari, 2009 <sup>155</sup>	Placebo	—	43	Oral	26	Good	G	—	
	DHEA	50	46	Oral				0.17 (-0.24 to 0.58)	
van Die, 2009 <sup>135</sup>	Placebo	—	50	Oral	16	Good	D	—	
	St John's wort	900	50	Oral			A	0.16 (-0.23 to 0.55)	
				G			0.22 (-0.17 to 0.61)		
Haines, 2008 <sup>118</sup>	Placebo	—	39	Oral	26	Poor	G	—	
Herbal extract	3000	45	Oral	0.39 (-0.04 to 0.82)					
Yang, 2007 <sup>121</sup>	Placebo	—	75	Oral	24	Poor	D	—	
	Maritime pine extract	200	80	Oral			A	-0.41 (-0.73 to -0.09)	
Mucci, 2006 <sup>122</sup>	Placebo	—	45	Oral	24	Poor	D	—	
	Isoflavones + magnolia bark	60	44	Oral			A	-1.00 (-1.44 to -0.57)	
Heger, 2006 <sup>123</sup>	Placebo	—	55	Oral	12	Poor	A	—	
	Rheum raphonticum	4	54	Oral			G	-0.77 (-1.16 to -0.38)	
Lewis, 2006 <sup>160</sup>	Placebo	—	33	Oral	16	Good	G	—	
	Isoflavones	42	33	Oral				-0.24 (-0.72 to 0.23)	
Uebelhack, 2006 <sup>134</sup>	Placebo	—	150	Oral	16	Good	D	—	
	Black cohosh + St John's wort	3.75+70	151	Oral			G	-1.32 (-1.57 to -1.07)	
Hartley, 2004 <sup>102</sup>	Placebo	—	27	Oral	12	Poor	D	—	
	Gingko biloba + ginseng	320	30	Oral			A	-0.15 (-0.67 to 0.36)	
								A	-0.23 (-0.74 to 0.29)

								G	0.11 (-0.40 to 0.62)
Wiklund, 1999 <sup>101</sup>	Placebo Ginseng	— 200	191 193	Oral Oral	16	Poor	D	-0.12 (-0.32 to 0.08)	
							A	-0.18 (-0.38 to 0.02)	
Barnhart, 1999 <sup>136</sup>	Placebo DHEA	— 50	30 30	Oral Oral	12	Poor	D	-0.23 (-0.73 to 0.27)	
							A	0.16 (-0.34 to 0.66)	

SMD: standardized mean difference; CI: confidence interval; DHEA: dehydroepiandrosterone; D: depression; A: anxiety; G: global mental health; FU: followup; Wks: weeks.

<sup>a</sup> combination of cynanchum wilfordii, phlomis umbrosa, angelica gigas

<sup>b</sup> combination of 21 vitamins and minerals

## Nonprescription Compared With Nonprescription

One trial (Table 37) compared isoflavones with isoflavones plus magnolia bark and reported depression and anxiety outcomes.<sup>141</sup> The trial was rated poor quality and found no difference in depression scores between the two groups (SMD: -0.09; 95 percent CI: -0.24 to 0.07) with some slight improvement in the group with isoflavones and magnolia bark compared with the isoflavones only group in anxiety scores (SMD: -0.16; 95 percent CI: -0.32 to -0.01).

**Table 37. Trials comparing nonprescription agents with nonprescription agents reporting psychological outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	Outcome	SMD (95% CI)
Agosta, 2011 <sup>141</sup>	Isoflavones	60	301	Oral	12	Poor	D	-0.09 (-0.24 to 0.07)
	Isoflavones + magnolia bark	60	335	Oral			A	-0.16 (-0.32 to -0.01)

SMD: standardized mean difference; CI: confidence interval; D: depression; A: anxiety; FU: followup; Wks: weeks.

## Antidepressant Compared With Antidepressant

One randomized double-blind trial (Table 38) compared flexible dose desvenlafaxine (100-200 mg/d) with flexible dose escitalopram (10-20 mg/d) and reported Hamilton depression and anxiety scores.<sup>216</sup> The trial was rated good quality. Both antidepressants improved both depression and anxiety scores significantly. The antidepressants were equally effective in reducing both depression and anxiety scores (SMD: -0.10; 95 percent CI: -0.29 to 0.08, and SMD: -0.05; 95 percent CI: -0.24 to 0.13, respectively).

**Table 38. Trials comparing antidepressants with antidepressants, reporting psychological outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	Outcome	SMD (95% CI)
Soares, 2010 <sup>216</sup>	Desvenlafaxine	100-200	224	Oral	8	Poor	D	-0.10 (-0.29 to 0.08)
	Escitalopram	10-20	237	Oral			A	-0.05 (-0.24 to 0.13)

SMD: standardized mean difference; CI: confidence interval; D: depression; A: anxiety; FU: followup; Wks: weeks.

## Antidepressant Compared With Nonprescription

One trial (Table 39) compared black cohosh with fluoxetine, reporting global mental health outcomes.<sup>100</sup> After 12 weeks of followup, Oktem et al. report that both treatments significantly improved the SF-36 global mental health score. There was no difference between the treatment groups (SMD: 0.09; 95 percent CI: -0.34 to 0.53). The trial was rated poor quality.

**Table 39. Trials comparing antidepressants with nonprescription agents, reporting psychological outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	Outcome	SMD (95% CI)
Oktem, 2007 <sup>100</sup>	Black cohosh	40	40	oral	26	poor	G	—
	Fluoxetine	20	40	oral				

SMD: standardized mean difference; CI: confidence interval; G: global mental health; FU: followup; Wks: weeks.

## Trials without Quantifiable or Poolable Data

Following is a description of three trials that did not allow determination of standardized effect size estimates because of reporting or reporting metric. Results of these trials would not have affected the overall outcomes presented above.

Pitkin et al. conducted a 52-week randomized, double-blind trial with three arms: 1 mg E<sub>2</sub>V + 2.5 mg MPA (n=152), 1 mg E<sub>2</sub>V + 5 mg MPA (n=153), and 2 mg E<sub>2</sub>V + 5 mg MPA (n=154).<sup>217</sup> WHQ depression and WHQ anxiety scores were the reported outcomes but could not be pooled. Between group differences were reported only as not significant and there were not enough other studies that compared high- with standard-dose estrogens and reported depression and anxiety. Authors reported that all three treatments improved depression and anxiety equally.

Odabasi et al. conducted a 12-week randomized trial comparing a standard-dose estrogen spray with a standard-dose estrogen patch.<sup>110</sup> Both treatments significantly improved global mental health scores. The authors reported no difference between the treatment groups, but no quantifiable data were provided. Only one other trial had similar comparators, but this trial also did not provide quantifiable data for pooling.

Davis et al. conducted a randomized crossover trial that also compared a standard-dose estrogen spray with a standard-dose estrogen patch.<sup>112</sup> Both treatments significantly improved global mental health scores. No significant difference between the two treatments was found. No quantifiable data between the groups were provided.

**Table 40. Strength of evidence ratings domains for psychological symptoms**

Domain	Comparisons	Comparators <sup>a</sup>	Risk of Bias	Consistency	Directness	Precision	Reporting Bias	SOE	Downgrading Rationale
Depression	4	<b>Estrogen (high dose)</b> vs. Placebo	H	C	D	P	U	Mod	3 poor-quality trials
Anxiety	2	<b>Estrogen (high dose)</b> vs. Placebo	H	U	D	P	U	Low	1 poor-quality trial; consistency unknown with only 2 trials
Depression	1	<b>Estrogen (standard)</b> vs. Placebo	M	U	D	P	U	Mod	11 trials; unknown consistency with 4 trial SMDs between -0.06 and 0.12
Anxiety	8	<b>Estrogen (standard)</b> vs. Placebo	H	I	D	I	U	Insuff	7 poor-quality trials; effect direction differed in 3 trials; CI overlaps 0
Global	9	<b>Estrogen (standard)</b> vs. Placebo	H	I	D	I	U	Insuff	7 poor-quality trials; effect direction differed among trials; CI overlaps 0
Depression	3	<b>Estrogen (low/ultralow)</b> vs. Placebo	H	I	D	I	U	Insuff	3 poor-quality trials; effect direction differed among trials; CI overlaps 0
Anxiety	3	<b>Estrogen (low/ultralow)</b> vs. Placebo	H	U	D	I	U	Insuff	3 poor-quality trials; all CIs and pooled CI overlaps 0
Global	7	<b>Estrogen (low/ultralow)</b> vs. Placebo	M	C	D	P	U	High	3 poor-quality trials
Depression	8	<b>Isoflavones</b> vs. Placebo	H	C	D	P	S	Low	7 poor-quality trials; potential reporting bias
Anxiety	7	<b>Isoflavones</b> vs. Placebo	H	C	D	I	U	Low	6 poor-quality trials; CI overlaps 0
Global	7	<b>Isoflavones</b> vs. Placebo	M	I	D	I	U	Low	2 poor-quality trials; CI overlaps 0
Depression	3	<b>SSRI/SNRI</b> vs. Placebo	H	C	D	P	U	Mod	3 poor-quality trials
Anxiety	2	<b>SSRI/SNRI</b> vs. Placebo	H	U	D	P	U	Low	2 poor-quality trials; consistency unknown with only 2 trials
Global	4	<b>SSRI/SNRI</b> vs. Placebo	H	C	D	P	U	Mod	4 poor-quality trials
Global	2	<b>Gabapentin/Pr egabalin</b> vs. Placebo	H	U	D	I	U	Insuff	2 poor-quality trials; consistency unknown with only 2 trials; CI overlaps 0

Risk of Bias: High (H), Medium (M), Low (L); Consistency: Inconsistent (I), Unknown (U), Consistent (C); Directness: Indirect (I), Direct (D); Precision: Imprecise (I), Unknown (U), Precise (P); Reporting Bias: Suspected (S), Undetected (U); SOE: strength of evidence; Mod: moderate; Insuff: insufficient; SSRI: selective serotonin reuptake inhibitor; SNRI: serotonin and norepinephrine reuptake inhibitor.

<sup>a</sup> Bold font of comparator indicates the more effective treatment; if both comparators are bold, the treatments are equivalently effective

## Key Points

- 90 trials including 48,894 women reported some measure of psychological outcome (depression, anxiety, and/or global mental health) in women treated with prescription (estrogen, testosterone, SSRIs, SNRIs) and nonprescription agents (isoflavones, black cohosh, ginseng, DHEA, herbal extracts, and others).
- Study quality was generally rated poor (76%). Thirty-five trials were funded by industry, 15 by public, 9 by industry and public, and 31 did not report funding source.
- Psychological outcomes were reported using a variety of scales and addressed three domains: global, anxiety and depression.
- Strength of evidence of relative effectiveness of agents in treating psychological symptoms.
  - For high-dose estrogen, there is **moderate** evidence of improvement in depression (N= 648 and SMD -0.64, 95 percent CI -0.94 to -0.33) and **low** strength evidence of improvement in anxiety (N= 469 and SMD -0.35, 95 percent CI: -0.58 to -0.13).
  - For standard-dose estrogen, there is **insufficient** strength of evidence of improvement in global mental health (N= 28,201 and SMD -0.03, 95 percent CI: -0.10 to 0.04) or anxiety (N=3193 and SMD -0.16, 95 percent CI -0.34 to 0.03); and **moderate** strength of evidence of improvement in depression (N=19,956 and SMD -0.19, 95 percent CI -0.31 to -0.07).
  - For low/ultralow-dose estrogen, there is **insufficient** strength of evidence of improvement in depression (N=162 and SMD -0.04, 95 percent CI -0.40 to 0.31) and anxiety (N=339 and SMD -0.19, 95 per cent CI: -0.41 to 0.02); and **high** strength of evidence of improvement in global mental health (N=1574 and SMD -0.24, 95 percent CI -0.45 to -0.02).
  - There is **low** strength of evidence that isoflavones compared with placebo improve all three domains: depression (N=1,005 and SMD -0.41, 95 percent CI: -0.69 to -0.13); anxiety (N= 873 and SMD -0.53, 95 percent CI: -0.87 to -0.20); global mental health (N=863 and SMD -0.12, 95 percent CI -0.26 to 0.01).
  - There is **moderate** strength of evidence that antidepressants compared with placebo improve depression (N= 2639 and SMD -0.40, 95 percent CI: -0.59 to -0.22 ) and global mental health (N=2535, SMD -0.39, 95 percent CI:-0.63 to -0.15); and **low** strength of evidence for improvement of anxiety (N=2,490 and SMD -0.31, 95 percent CI -0.53 to -0.08)
  - There is **insufficient** evidence that gabapentin/pregabalin compared with placebo improves global mental health (N= 256, SMD -0.22, 95 per cent CI: -0.46 to 0.03).
  - There was insufficient evidence on the effectiveness of other agents and comparators on psychological outcomes.

## Sexual Function

### Included Trials

Of the 254 trials included in this CER, 76 trials (29.9 percent) trials reported sexual function outcomes (36 trials specified sexual function as a primary outcome). Fifty-five trials examined hormone treatment effects on sexual function, with the following comparators: placebo (29 trials), other hormones (25 trials), and nonprescription treatments (one trial). Eighteen trials examined the effects of nonprescription treatments compared with placebo. The nonprescription treatments included isoflavones, DHEA, herbal extracts, and ginseng. Three trials compared antidepressants with placebo and reported sexual function outcomes.

Trials were performed in 21 different individual countries and 19 trials were multinational. Countries included the United States (n=12), Australia (n=8), the United Kingdom (n=5), Canada (n=4), Taiwan (n=4), Denmark (n=3), and Italy (n=3); fewer were performed in Germany, Turkey, Hong Kong, and Singapore. The trials were conducted at 2,011 sites. Length of followup ranged from 8 to 260 weeks.

Sexual function was reported using a variety of measures and scales. The domains of sexual activity assessed fell into four broad categories: global (i.e., assessed two or more domains), pain (dyspareunia), interest, or activity frequency. We included a single result from each trial according to commonness of reporting—global, pain, interest, and activity. Few trials reported more than one outcome. Thirty-one trials reported a global measure (MENQOL, WHQ, MRS, and McCoy scales were most common, though others were also used); 22 reported pain during intercourse, 15 interest in sexual activity, and eight reported frequency of sexual activity. Specific items in the different scales include:

- Greene Climacteric Scale rated a single question, “loss of interest in sex,” scaled from zero (none) to three (severe)—15 trials.
- Menopause-specific Quality of Life (MENQOL) assessed sexual function in three questions scaled from zero (not bothered) to eight (extremely bothered)—14 trials.
- Women’s Health Questionnaire queried sexual function through three questions on interest, pain, and activity, rated in a four point scale—eight trials.
- Self-reported dyspareunia—nine trials.
- Satisfying sexual episodes—eight trials.
- The remaining trials used other sexual function scales.

Study quality was generally poor (79 percent), with four trials judged to be fair and 12 trials judged to be good. Length of followup ranged from 8 weeks to 260 weeks. Industry funding was indicated in 41 trials, public funding 11 trials, and three trials reported both industry and public funding. Table 41 describes additional trial and patient characteristics.

**Table 41. Characteristics of trials assessing efficacy for sexual function**

	Characteristic	Number (%)
	Number of trials	76
	Total number of patients	23,923
		2,077
Trial Characteristics	Total sites from trials that specified site # (n=64)	1 to 502 (mean: 32; median: 9)
	Trials described only as multicenter	6 (7.9)
	Multicenter trials	51 (67.1)

	Two-arm trials	62 (81.6)
	Multi-arm trials	14 (18.4)
	Patients per trial	50 to 2,459 (mean: 315; median: 158)
	Range of followup (weeks)	8 to 260 (mean: 27.6; median:16)
Funding	Industry only	41 (53.9)
	Public only	11 (14.5)
	Industry and public	3 (3.9)
	Not stated	21 (27.6)
Comparator Category	Placebo vs. hormone	29 (38.2)
	Antidepressant vs. placebo or other antidepressant	3 (3.9)
	Placebo vs. other prescription	0 (0.0)
	Placebo vs. nonprescription	18 (23.7)
	Placebo vs. hormone vs. nonprescription	0 (0.0)
	Hormone vs. hormone	25 (32.9)
	Hormone vs. nonprescription	1 (1.3)
	Nonprescription vs. nonprescription	0 (0.0)
Study Quality	Good	12 (15.8)
	Fair	4 (5.3)
	Poor	60 (78.9)
Patient Demographics	Mean age (years)	46.8 to 66.7 (NR: 7)
	Age range (years)	26.0 to 86.0 (NR: 58)
	Years since menopause	5.1 (0.7 to 16.5) (NR: 45)
	Current smokers (%)	0.0 to 44.0 (NR: 57)
	Mean BMI (kg/m <sup>2</sup> )	17.3 to 28.4 (NR: 28)
	White (%)	0.0 to 98.6
	Black (%)	0.0 to 12.0
	Hispanic (%)	0.0 to 10.5
	Asian (%)	0.0 to 100.0
	Other (%)	0.0 to 26.6
Uterus Status	All intact	24 (31.6)
	All absent	6 (7.9)
	Mixed	31 (40.8)
	Range, percentage intact among trials with mixed	25.0 to 94.3
	Not reported	15 (19.7)

Note: Demographics were not reported in all studies.  
NR: not reported.

## Evidence Synthesis for Sexual Function

Standard mean differences were calculated to allow comparisons of outcomes from different sexual function scoring systems. Analyses were conducted by domain (pain, global, activity and interest), by mode of administration (oral or vaginal), and by uterine status (all intact, all absent,

or mixed) when possible. Pooling was considered possible for pairwise comparisons where evidence included at least three trials. Pooling of the following comparators and conditions was performed (Table 42):

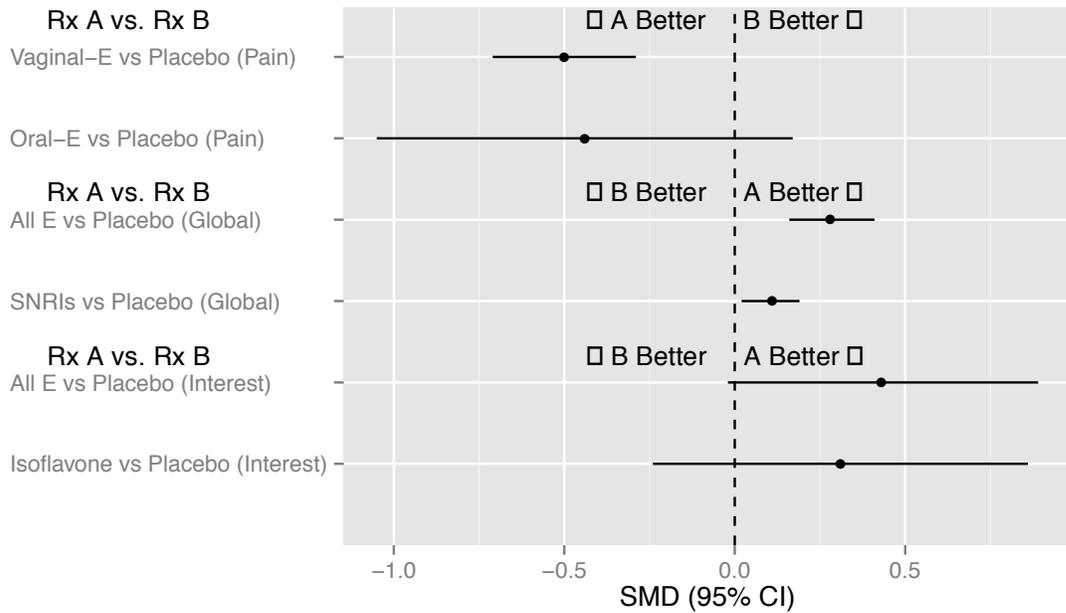
- Pain: placebo versus vaginally applied estrogens (n=10); placebo versus oral estrogens (n=3); placebo versus all estrogens (either vaginally applied or oral) (n=13)
- Global: placebo versus all estrogens (either vaginally applied or oral) (n=10)
- Activity: placebo versus testosterone in trials with women with/without uteri mixed or trials with women with intact uteri (n=4); placebo versus testosterone in trials with all women without intact uteri (n=4); placebo versus testosterone all trials combined (n=8)
- Interest: placebo versus all estrogens (n=3); placebo versus isoflavones (n=3)

**Table 42. Pooled effect sizes from single trials for improvement in sexual function**

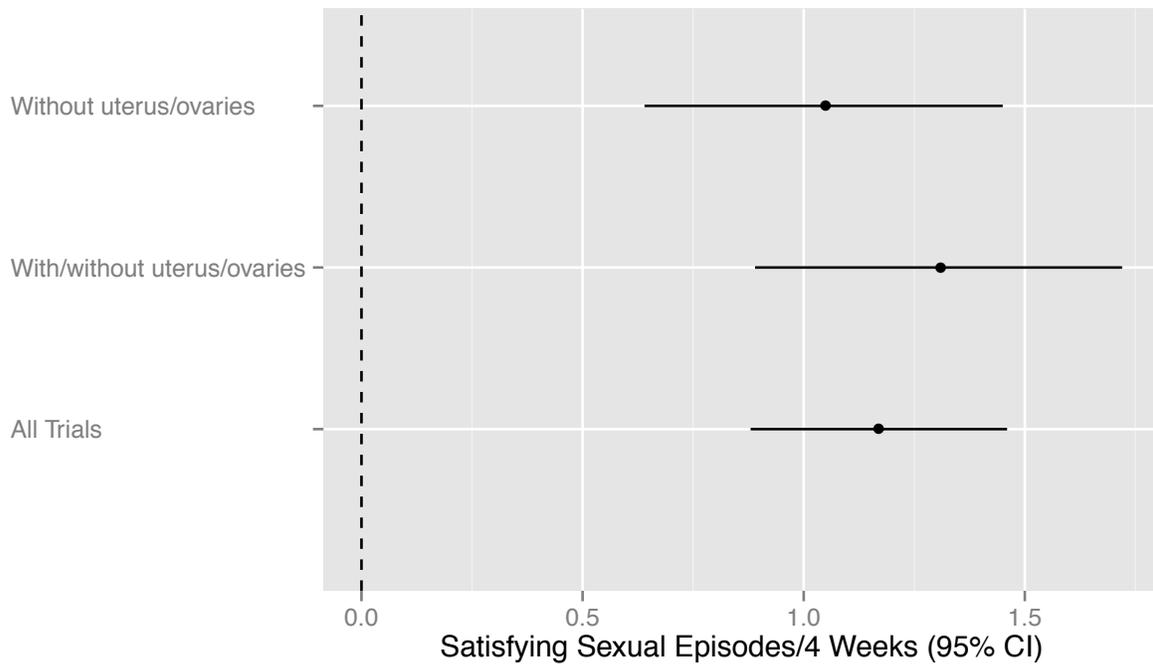
<b>Sexual Function Domain</b>	<b>Comparators vs. Placebo</b>	<b>No. of Trials</b>	<b>SMD (95% CI)</b>
Pain (lower is better)	Vaginally applied estrogens	10	-0.50 (-0.71 to -0.29); tau <sup>2</sup> =0.09
	Oral estrogens	3	-0.44 (-1.05 to 0.17); tau <sup>2</sup> =0.28
	All estrogens	13	-0.49 (-0.69 to -0.29); tau <sup>2</sup> =0.11
Global (higher is better)	All estrogens	10	0.28 (0.16 to 0.41); tau <sup>2</sup> =0.02
Interest (higher is better)	All estrogens	3	0.43 (-0.02 to 0.89); tau <sup>2</sup> =0.12
	Isoflavones	3	0.31 (-0.24 to 0.86); tau <sup>2</sup> =0.20
<b>Mean Difference SSE/4 Weeks</b>			
Activity (higher is better)	Testosterone, no women with intact uteri/ovaries	4	1.05 (0.64 to 1.45); tau <sup>2</sup> =0.00
	Testosterone, women with/without uteri/ovaries	4	1.31 (0.89 to 1.72); tau <sup>2</sup> =0.00
	Testosterone, all trials	8	1.17 (0.88 to 1.46); tau <sup>2</sup> =0.00

SMD: standardized mean difference; CI: confidence interval; SSE: satisfying sexual episodes.

**Figure 10. Caterpillar plot for sexual function: pain, global, and interest—standardized effects and 95 percent confidence intervals**



**Figure 11. Caterpillar plot sexual function: satisfying sexual episodes with testosterone compared with placebo—mean difference and 95 percent confidence intervals**



### **Estrogen Compared With Placebo (Pain)**

Thirteen trials compared estrogens with placebo and reported pain outcomes. Ten trials compared vaginal estrogens with placebo<sup>174, 218-225</sup> and three trials compared oral estrogens with placebo.<sup>173, 226, 227</sup> Pooling outcomes from the estrogen trials revealed that any estrogen significantly improved reported pain during sex compared with placebo—standardized effect size of -0.49 (95 percent CI: -0.69 to -0.29;  $\tau^2=0.11$ ). Analyses by mode of administration (vaginal or oral) showed that vaginally administered estrogens significantly diminished pain compared with placebo (standardized effect size -0.50; 95 percent CI: -0.71 to -0.29), while oral estrogens did not (standardized effect size -0.44; 95 percent CI: -1.05 to 0.17). Only three trials compared oral estrogens with placebo, which may explain those findings. The evidence comparing estrogens with placebo consists of 13 trials. Eleven trials were judged as poor quality and one trial judged as fair quality. The strength of evidence that vaginal estrogens compared with placebo improve reported pain during sex among menopausal women is rated as high. The strength of evidence that oral estrogens compared with placebo improve reported pain during sex among menopausal women is rated as **low**.

### **Estrogen Compared With Placebo (Global)**

Ten trials compared estrogens with placebo and reported a global outcome for sexual function. Because of the variety of modes of administration among these trials, analysis according to mode of administration was not feasible. Five trials administered oral estrogens,<sup>26, 87, 190, 228, 229</sup> two trials used patches,<sup>176, 230</sup> and one trial each used spray,<sup>188</sup> gel,<sup>194</sup> and cream<sup>231</sup> estrogens. Pooling results from the ten trials showed that estrogens significantly improve overall sexual function compared with placebo, with a standardized effect size of 0.28 (95 percent CI: 0.16 to 0.41;  $\tau^2=0.02$ ). Six trials were judged as poor quality, one trial judged as fair quality, and three trials judged as good quality. The strength of evidence that estrogens compared with placebo improve a global assessment of sexual function is rated as high.

### **Testosterone Compared With Placebo (Activity)**

Eight trials compared testosterone with placebo and assessed satisfying sexual episodes. The outcome was the number of episodes over a 4-week period. Four of the trials included only women without intact uteri and ovaries,<sup>232-235</sup> two trials included only women with intact uteri and ovaries,<sup>236, 237</sup> and two trials included women with and without intact uteri and ovaries.<sup>238, 239</sup> Combining the eight trials showed that testosterone significantly improved sexual activity compared with placebo by 1.17 episode/4 weeks (95 percent CI: 0.88 to 1.46;  $\tau^2=0.00$ ). Analyses limited to the four trials including only women without intact uteri and ovaries also showed significant improvements in episodes compared with placebo (1.05; 95 percent CI: 0.64 to 1.45).

The evidence comparing testosterone with placebo to increase the number of satisfying sexual episodes consists of eight trials. Seven trials were judged as poor quality and one trial was fair quality. Compared with placebo, the strength of evidence that testosterone increases the number of satisfying sexual episodes is rated as moderate.

### **Estrogens Compared With Placebo (Interest)**

Three trials compared estrogens with placebo and assessed sexual interest. Two trials administered oral estrogens<sup>181, 192</sup> and one included two estrogen arms—oral and patch.<sup>183</sup> Results from pooling the three trials did not show a significant increase in reported sexual

interest, with a standardized effect size of 0.43 (95 percent CI: -0.02 to 0.89; tau<sup>2</sup>=0.12). All three trials were judged to be of poor quality. The strength of evidence that estrogens improve sexual interest compared with placebo is rated as insufficient.

### Isoflavones Compared With Placebo (Interest)

Three trials compared isoflavones with placebo and assessed sexual interest.<sup>150, 196, 201</sup> A higher score indicates an improvement in interest in sexual activity. The treatment groups in the three trials were given 30 mg to 80 mg daily doses of isoflavones. Results from pooling the three trials did not show a significant improvement in sexual interest, with a standardized effect size of 0.31 (95 percent CI: -0.24 to 0.86; tau<sup>2</sup>=0.20). The small number of trials and the small collective sample size (441 total participants) may explain the lack of effect. The strength of evidence that estrogens improve sexual interest compared with placebo is rated as insufficient.

### SSRI/SNRI Compared With Placebo

Two trials compared antidepressants with placebo and reported sexual function outcomes in the global domain (both trials rated poor quality).<sup>207, 208</sup> A higher score indicates improvement in this sexual function domain. The large trial sponsored by Pfizer (n=2118), administering 100 mg desvenlafaxine to the treatment group, reported an improvement in a global metric but it was not significant (p=0.08).<sup>207</sup> The smaller trial, with a total study population of 80 and administering 75 mg venlafaxine to the treatment group, reported no difference in overall sexual function between the study groups.<sup>208</sup> The pooled SMD was 0.11 (95 percent CI: 0.02 to 0.19) tau<sup>2</sup>=0.00. The strength of evidence that SNRIs compared with placebo improve a global assessment of sexual function is rated as insufficient.

### Trials Not Pooled

#### Estrogen Compared With Estrogens or Other Hormones

Five trials (Table 43) compared different doses of estrogen. Four of the five trials compared standard with low doses<sup>240-243</sup> and one trial compared standard with a high dose.<sup>217</sup> Outcomes of two trials were global sexual function, two measured interest in sexual activity, and one measured pain during sexual activity. In four of the five trials, standard doses were found to be similar in effectiveness compared with the lower and higher doses. In one trial, the larger estrogen doses of 0.45 mg and 0.625 mg were not found to significantly improve overall sexual function compared with 0.3 mg.<sup>243</sup> Due to the variety in outcome measures, synthesizing these data was not possible; because of treatment heterogeneity, a strength of evidence was not rated.

**Table 43. Trials comparing different estrogen doses, reporting a global assessment of sexual function**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	Sexual Function Domain	SMD (95% CI)
Pitkin, 2007 <sup>217</sup>	E2V + MPA	1 E + 2.5 P	152	Oral	52	Poor	Global	—
	E2V + MPA	1 E + 5 P	153	Oral				0.00 (-0.22 to 0.22)
	E2V + MPA	2 E + 5 P	154	Oral				-0.23 (-0.45 to 0.00)
Cieraad, 2006 <sup>240</sup>	Estradiol + progestin CEE + progestin	1 E + 10 P	98	Oral	24	Poor	Interest	—
		0.625 E + 0.15 P	91	Oral				0.10 (-0.19 to 0.38)
Utian,	Estradiol +	0.9 E	79	Oral	12	Good	Pain	—

2005 <sup>241</sup>	progestin	0.625 E	85	Oral					-0.14 (-0.45 to 0.17)
	CEE + progestin	1 E	84	Oral					-0.29 (-0.60 to 0.02)
	Estradiol + progestin								
Loh, 2002 <sup>242</sup>	Estradiol + NETA	1 E + 0.5 P	48	Oral	26	Poor	Interest		—
	Estradiol + NETA	2 E + 1 P	48	Oral					0.08 (-0.32 to 0.48)
Limpaphayom, 2006 <sup>243</sup>	CEE + MPA	0.3 E + 1.5 P	342	Oral					—
	CEE + MPA	0.45 E + 1.5 P	342	Oral	24	Poor	Global		NS
	CEE + MPA	0.625 E + 2.5 P	344	Oral					NS

SMD: standardized mean difference; CI: confidence interval; E: estrogen; E2V: estradiol valerate; P: progestin; CEE: conjugated equine estrogen; MPA: medroxyprogesterone acetate; NETA: norethisterone acetate; NS: not significant; FU: followup; Wks: weeks.

Ten trials (Table 44) compared similar estrogen doses using different modes of administration. Two trials used a vaginal ring in one treatment group and vaginal cream in another<sup>244, 245</sup>; two trials used oral estrogens in one arm and estrogen patches in another<sup>111, 246</sup>; one trial used patches, either adding progestin combined or sequential<sup>157</sup>; and one trial each used the following pairs of modes of administration: patch/spray,<sup>112</sup> oral/ring,<sup>213</sup> ring/tablet,<sup>247</sup> oral/cream,<sup>248</sup> and ring/pessary.<sup>249</sup> Five trials reported a global sexual function outcome, four reported pain, and one reported interest in sexual activity. No trial found a significant difference in outcomes between modes of administration. These results on mode of administration combined with the findings from the analysis on vaginal and oral estrogens compared with placebo in diminishing pain during sex, suggest global and pain outcomes also do not differ according to mode of administration (strength of evidence moderate Table 48).

**Table 44. Trials comparing different estrogen modes of administration, reporting pain, interest, and global sexual function outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	Sexual Function Domain	SMD (95% CI)
Serrano, 2006 <sup>111</sup>	CEE + MPA	0.625 E + 10 P	55	Oral	5	Poor	Global	—
	Estradiol + MPA	0.05 E + 10 P	59	Patch	2			0.36 (-0.01 to 0.73)
Davis, 2005 <sup>112</sup>	Estradiol	0.05	60	Patch	1	Poor	Global	—
	Estradiol	0.30	60	Spray	6			-0.10 (-0.46 to 0.26)
Buckler, 2003 <sup>213</sup>	Estradiol + progestin	1 E + 1 P	75	Oral Ring	2	Fair	Interest	—
	Estradiol + progestin	0.05 E + 1 P	84		4			0.00 (-0.31 to 0.31)
Weisberg, 2005 <sup>247</sup>	Estradiol	0.008	126	Ring	4	Poor	Global	—
	Estradiol	0.025	59	Tablet	8			0.01 (-0.29 to 0.32)
Lubbert, 1997 <sup>157</sup>	Estradiol, combined	0.05	1232	Patch	1	Poor	Global	—
	Estradiol, sequential	0.05	1227	Patch	2			0.02 (-0.06 to 0.10)
Barentsen, 1997 <sup>244</sup>	Estradiol	0.0075	83	Ring	1	Poor	Pain	—
	Estriol	0.5	82	Cream	2			-0.06 (-0.37 to 0.25)
Ayton, 1996 <sup>245</sup>	Estradiol	0.0075	131	Ring	1	Poor	Pain	—
	CEE	0.625	63	Cream	2			-0.16 (-0.46 to 0.14)
Henriksson, 1994 <sup>249</sup>	Estradiol	0.0095	106	Ring	1	Poor	Pain	—
	Estriol	0.5	51	Pessary	2			-0.04 (-0.37 to 0.29)
Long, 2006 <sup>248</sup>	CEE	0.625	37	Oral	1	Poor	Pain	—
	CEE	0.625	36	Cream	2			-0.36 (-0.82 to 0.11)
Hilditch, 1996 <sup>246</sup>	CEE + MPA	0.635 E + 10 P	35	Oral	1	Poor	Global	—
	Estradiol + MPA	0.014 E + 10 P	39	Patch	4			NS

SMD: standardized mean difference; CI: confidence interval; CEE: conjugated equine estrogen; MPA: medroxyprogesterone acetate; E: estrogen; P: progestin; NS: not significant; FU: followup; Wks: weeks.

One trial (Table 45) randomized patients to either 0.625 mg esterified estrogens or 0.625 mg esterified estrogens plus 1.25 mg methyltestosterone.<sup>250</sup> The outcome was a global measure of sexual function. After 16 weeks of followup, the group receiving testosterone with estrogen improved significantly compared with the estrogen alone group, with a standardized mean difference of 0.39 (95 percent CI: 0.12 to 0.66).

**Table 45. Trials comparing estrogens with testosterone, reporting sexual function outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	Sexual Function Domain	SMD (95% CI)
Lobo, 2003 <sup>250</sup>	Estrogen	0.625	111	oral	16	Poor	Global	—
	Estrogen + testosterone	0.625	107	oral				

SMD: standardized mean difference; CI: confidence interval FU: followup; Wks: weeks.

### Estrogen Compared With Nonprescription Agents

A single trial (Table 46) compared estrogen/progestin therapy with pueraria mirifica for the treatment of pain relating to sexual function.<sup>139</sup> Pueraria mirifica is an herb considered highly estrogenic, found in Thailand. This small study with a sample size of 60 women, did not find a significant difference between groups.

**Table 46. Trials comparing estrogens with nonprescription agents reporting sexual function outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	Sexual Function Domain	SMD (95% CI)
Chandeying, 2007 <sup>139</sup>	CEE + MPA	0.625 E + 2.5 P	30	oral	24	poor	pain	—
	Pueraria mirifica	50	30	oral				

SMD: standardized mean difference; CI: confidence interval; CEE: conjugated equine estrogen; MPA: medroxyprogesterone acetate; E: estrogen; P: progestin; FU: followup; Wks: weeks.

### Nonprescription Agents Compared With Placebo

Fourteen trials (Table 47) compared nonprescription agents with placebo and reported sexual function outcomes. Seven trials compared various herbal or plant extracts with placebo in the treatment of sexual function. Four of the trials reported global sexual function outcomes<sup>118, 121, 123, 131</sup> and three reported on interest in sexual activity.<sup>120, 128, 135</sup>

Some of the treatments consisted of combinations of herbs, while others consisted of single plants such as pine extract, rheum raphonticum, and dioscorea alata. None of these trials reported a significant improvement in sexual function except for the trial using maritime pine extract, reporting an improvement in overall sexual function (standardized mean difference 0.50; 95 percent CI: 0.18 to 0.82).<sup>121</sup>

Three trials compared isoflavones with placebo.<sup>160, 195, 204</sup> One of the trials had a third arm, which administered flaxseed as a treatment. Two of the trials reported global sexual function outcomes and one reported pain. None of the three trials found a significant difference between the treatment groups and sexual function.

Two trials compared ginseng with placebo, with one trial reporting a global sexual function outcome<sup>101</sup> and one reporting on interest in sexual activity.<sup>102</sup> Neither trial reported significant improvements.

Two of the 14 trials compared DHEA with placebo and reported global sexual function outcomes.<sup>155, 156</sup> One was a four-arm trial with increasing doses of DHEA which was administered through an ovule and the other was a two-arm trial administering DHEA orally. The ovule trial showed significant improvements in global sexual function, while the oral trial did not show a difference. Due to the variety of dosages and treatments, pooling was not appropriate.

**Table 47. Trials comparing nonprescription agents with placebo, reporting sexual function outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	Sexual Function Domain	SMD (95% CI)
Haines, 2008 <sup>118</sup>	Placebo	—	39	Oral	26	Poor	Global	—
	Dang Gui Buxue Tang	3000	45	Oral				0.10 (-0.33 to 0.53)
Van der Sluijs, 2009 <sup>120</sup>	Placebo	—	42	Oral	16	Good	Interest	—
	Plant extracts <sup>a</sup>	3820	41	Oral				0.19 (-0.24 to 0.62)
Xia, 2012 <sup>131</sup>	Placebo	—	36	Oral	12	Good	Global	—
	Jiawei Qing'e Fang	3500	36	Oral				0.19 (-0.28 to 0.65)
Van Die, 2009 <sup>135</sup>	Placebo	—	50	Oral	16	Good	Interest	—
	St John's wort	900	50	Oral				-0.21 (-0.61 to 0.18)
Yang, 2007 <sup>121</sup>	Placebo	—	75	Oral	24	Poor	Global	—
	Maritime pine extract	200	80	Oral				0.50 (0.18 to 0.82)
Heger, 2006 <sup>123</sup>	Placebo	—	55	Oral	12	Poor	Global	—
	Rheum raphonticum	4	54	Oral				0.38 (0.00 to 0.76)
Hsu, 2011 <sup>128</sup>	Placebo	—	25	Oral	52	Poor	Interest	—
	Dioscorea alata	24	25	Oral				NS
Wiklund, 1999 <sup>101</sup>	Placebo	—	191	Oral	16	Poor	Global	—
	Ginseng	200	193	Oral				NS
Hartley, 2004 <sup>102</sup>	Placebo	—	27	Oral	12	Poor	Interest	—
	Ginseng	120	30	Oral				0.50 (-0.03 to 1.03)
Labrie, 2009 <sup>156</sup>	Placebo	—	53	Ovule	12	Poor	Global	—
	DHEA	3.25	53	Ovule				0.81 (0.41 to 1.21)
	DHEA	6.5	56	Ovule				0.39 (0.01 to 0.77)
	DHEA	13	54	Ovule				0.74 (0.35 to 1.13)
Panjari, 2009 <sup>155</sup>	Placebo	—	43	Oral	26	Good	Global	—
	DHEA	50	46	Oral				0.24 (-0.18 to 0.65)
Kotsopoulos, 2000 <sup>195</sup>	Placebo	—	50	Oral	13	Poor	Pain	—
	Isoflavone	118	44	Oral				0.26 (-0.15 to 0.67)
Basaria, 2009 <sup>204</sup>	Placebo	—	46	Oral	12	Good	Global	—
	Isoflavone	160	38	Oral				0.15 (-0.28 to 0.58)
Lewis, 2006 <sup>160</sup>	Placebo	—	33	Oral	16	Good	Global	—
	Isoflavone	42	33	Oral				-0.18 (-0.66 to 0.30)
	Flaxseed	50	33	Oral				0.14 (-0.35 to 0.62)

SMD: standardized mean difference; CI: confidence interval; DHEA: dehydroepiandrosterone; FU: followup; Wks: weeks.

<sup>a</sup> combination of black cohosh, er xian tang, zhi bai di huang wan

## Trials With No Quantifiable Data

Following is a description of five trials that did not have data that could be analyzed by the standardized effect size methods. Results of these trials would not have affected the overall outcomes presented above.

In a double-blind trial, women were randomized to either oral estrogen/progestin (n=75) or estrogen/progestin administered through a vaginal ring (n=84).<sup>213</sup> While this study was not designed to evaluate statistically significant changes in sexual function, the authors report a comparable decrease in the intensity of selected local symptoms (vaginal dryness, urinary incontinence, and pain during intercourse) observed in both treatment groups.

In a double-blind trial, women were randomized to either a progestin skin cream (n=38) or a placebo skin cream (n=42), and were followed for 12 weeks.<sup>106</sup> Sexual function outcomes were measured by the Greene sexual function subscore and reported as baseline median and post-treatment median. Similar improvements were seen in both study groups.

In a trial comparing a mixture of 12 Chinese herbs (n=28) with placebo (n=27), the sexual function subscore for the MENQOL was reported.<sup>125</sup> Followup was 12 weeks. Baseline measures were provided for both the placebo and treatment groups, but followup measures were provided for only the group treated with the Chinese herbs. The authors report that there was no statistical difference in sexual function between the two groups.

A four-arm randomized trial was conducted, with two arms receiving two different dosages of raloxifene, one arm receiving 0.625 mg/day CEE, and one arm receiving placebo.<sup>185</sup> Raloxifene is not a treatment of interest for this CER, so information from only the estrogen and placebo arms were abstracted. The study lasted for 52 weeks and sexual function was measured using the WHQ sexual subscale score. Within-group mean change scores were small (0.01 for the placebo group and 0.12 for the treatment group), but the estrogen group showed a larger change. The small mean changes and the lack of p-values made imputing standard deviations and pooling problematic.

A 26-week, double-blind, crossover trial of 53 women added a testosterone skin gel and placebo gel to postmenopausal women's already existing hormone treatments.<sup>214</sup> One of the outcomes measured was on pain during intercourse measured by the McCoy scale. Outcomes were reported as a series of p-values, so standardized effect sizes could not be calculated.

## Strength of Evidence Ratings—Sexual Function

**Table 48. Strength of evidence ratings domains for sexual function**

Comparisons	Comparators <sup>a</sup>		Risk of Bias	Consistency	Directness	Precision	Reporting Bias	SOE	Downgrading Rationale
<b>Pain</b>									
10	<b>Vaginal estrogen</b>	vs. Placebo	M	C	D	P	U	High	
3	<b>Oral estrogen</b>	vs. Placebo	H	C	D	I	U	Low	Poor trial quality; confidence interval overlapping 0
<b>Global</b>									
10	<b>All estrogens</b>	vs. Placebo	M	C	D	P	U	High	

2	<b>SSRI</b>	vs. Placebo	H	U	D	I	U	Insuff	2 poor-quality trials; CIs for SMDs overlap 0; consistency not evaluable
<b>Activity</b>									
8	<b>Testosterone</b>	vs. Placebo	H	C	D	P	U	Mod	7 poor-quality trials
<b>Interest</b>									
3	<b>All estrogens</b>	vs. Placebo	H	I	D	I	U	Insuff	3 poor-quality trials; positive and negative effects; imprecise estimate
3	<b>Isoflavone</b>	vs. Placebo	H	I	D	I	U	Insuff	1 good and 2 poor quality trials; positive and negative effects; imprecise estimate
<b>Pain, interest, global</b>									
10	<b>Estrogen mode a</b>	vs. <b>Estrogen mode b</b>	M	C	D	U	U	Mod	Precision unknown with 3 domains assessed

Risk of Bias: High (H), Medium (M), Low (L); Consistency: Inconsistent (I), Unknown (U), Consistent (C); Directness: Indirect (I) Direct (D); Precision: Imprecise (I), Unknown (U), Precise (P); Reporting Bias: Suspected (S) Undetected (U); SOE: strength of evidence; Mod: moderate; Insuff: insufficient; SMD: standardized mean difference; CI: confidence interval; SSRI: selective serotonin reuptake inhibitor.

<sup>a</sup> Bold font of comparator indicates the more effective treatment; if both comparators are bold, the treatments are equivalently effective

## Key Points

- Seventy-six trials, including almost 24,000 women, reported on sexual function outcomes of treatment with hormones, antidepressants or nonprescription agents such as isoflavones, DHEA and herbal extracts.
- Study quality was generally rated poor (79 per cent). Forty-one trials reported industry funding, 11 reported public funding and 3 trials reported both industry and public funding.
- Sexual function outcomes were reported using a variety of scales, representing four domains of sexual function: global, pain, interest or activity frequency.
- Strength of evidence of relative effectiveness of agents in ameliorating symptoms of sexual function:
  - Thirteen trials, with a total of 4,608 women participating, compared estrogens with placebo. Combined results showed that estrogen reduced pain during sex compared with placebo (SMD: -0.49; 95 percent CI: -0.69 to -0.29). There is **high** strength of evidence (N= 3,218) that vaginal estrogen reduced pain compared with placebo (SMD -0.50, 95 percent CI -0.71 to -0.29). There is **insufficient** strength of evidence (N=1,390) that oral estrogen reduces pain (SMD -0.44, 95 per cent CI -1.05 to 0.17).
  - There is **high** strength of evidence that estrogen improves global measures of sexual function compared with placebo. Combined results of 10 trials with a total of 3,936 women showed that estrogen improves global measures of sexual function (SMD 0.28, 95 percent CI: 0.16 to 0.41). There is **insufficient** strength of evidence that SNRIs improve global measures of sexual function.
  - There is **moderate** strength of evidence that testosterone compared with placebo improves measures of sexual activity. Combined results of 8 trials (N= 2,820) yielded an SMD of 1.17 (95 percent CI: 0.88 to 1.50).
  - There is **insufficient** evidence to determine whether, compared to placebo, estrogens (N=417) or isoflavones (N=441) improves measures of interest in sex.

## Urogenital Atrophy

### Included Trials

Of the 254 total included trials in this CER, 63 (24.8 percent) reported urogenital atrophy outcomes (35 trials specified urogenital atrophy as a primary outcome). Forty-seven trials examined effects of hormones including the following comparators: placebo (29 trials), other hormones (16 trials), and nonprescription treatments (two trials). Sixteen trials examined the effects of nonprescription treatments such as isoflavones, black cohosh, and herbal extracts.

The 63 trials were performed in 23 different countries including the United States (n=14), Italy (n=6), and Germany (n=5); nine were multinational. The trials were conducted at 1,711 sites with followup ranging from 12 to 260 weeks.

Urogenital atrophy outcomes were reported using a variety of measures and scales. The most common metrics included

- Vaginal dryness on a dichotomous scale.
- Vaginal dryness severity score, ranging from 0 (none) to 3 (severe).
- The Menopause Rating Scale (MRS) with a single item rating vaginal dryness on a five-point scale from 0 (none) to 4 (extremely severe).
- Several researchers devised their own outcome measurement for urogenital symptoms, either patient or physician assessed. Different researchers used different combinations of the following symptoms, assigning scores, resulting in an overall urogenital score: vaginal discomfort, loss of libido, dyspareunia, vaginal dryness, vaginal itching, and incontinence.
- Dryness improvement.
- The Modified Greene Climacteric Scale including a single item assessing vaginal dryness on a scale from 0 (none) to 3 (most severe).
- Visual analog scale
- The Kupperman Menopausal Index vaginal dryness on a scale from 0 (none) to 3 (most severe).

Study quality was generally rated poor (84 percent), with four fair- and six high-quality trials. Industry funding was indicated in 30 trials and public funding was reported in 11 trials. Table 49 describes other trial and patient characteristics.

**Table 49. Characteristics of trials assessing efficacy for urogenital atrophy**

	Characteristic	Number (%)
	Number of trials	63
	Total number of patients	18,339
	Number of sites from trials that specified (n=145)	1,711 (mean: 35; median: 9)
Trial Characteristics	Trials described only as multicenter	8 (12.7)
	Multicenter trials	43 (68.3)
	Two-arm trials	49 (77.8)
	Multi-arm trials	14 (22.2)
	Patients per trial	52 to 2,459 (mean: 291; median: 159)
	Range of followup (weeks)	12 to 260 (mean: 26.9; median: 13)

Funding	Industry only	27 (42.9)
	Public only	8 (12.7)
	Industry and public	3 (4.8)
	Not stated	25 (39.7)
Comparator Category	Placebo vs. hormone	29 (46.0)
	Antidepressant vs. placebo or other antidepressant	0 (0.0)
	Placebo vs. other prescription	0 (0.0)
	Placebo vs. nonprescription	13 (20.6)
	Placebo vs. hormone vs. nonprescription	0 (0.0)
	Hormone vs. hormone	16 (25.4)
	Hormone vs. nonprescription	2 (3.2)
	Nonprescription vs. nonprescription	3 (4.8)
	Nonprescription vs. antidepressant	0 (0.0)
Study Quality	Good	6 (9.5)
	Fair	4 (6.3)
	Poor	53 (84.1)
Patient Demographics	Mean age (years)	43.8 to 66.7 (NR: 7)
	Age range (years)	29 to 86 (NR: 45)
	Years since menopause	5.3 (0.6 to 16.5 (NR: 41)
	Current smokers (%)	0.0 to 44.0 (NR: 54)
	Mean BMI (kg/m <sup>2</sup> )	22.1 to 29.3 (NR: 30)
	White (%)	0 to 100
	Black (%)	0 to 15.5
	Hispanic (%)	0 to 10.5
	Asian (%)	0 to 100
	Other (%)	0 to 26.6
Uterus Status	All intact	19 (32.2)
	All absent	4 (6.3)
	Mixed	27 (42.9)
	Range, percentage intact among trials with mixed	30.6 to 87.2
	Not reported	13 (20.6)

Note: Demographics were not reported in all studies.

NR: not reported.

## Evidence Synthesis for Urogenital Atrophy

Standard mean differences were calculated to allow comparing outcomes across different scales. Pooling was performed for pairwise comparisons where evidence included more than two trials. Analyses of estrogen treatments were performed separately for vaginal and nonvaginal estrogens. Pooling of the following comparators was performed: placebo versus vaginal estrogens according to dose; placebo versus nonvaginal estrogens according to dose; placebo versus isoflavones; and placebo versus black cohosh. Forest plots are displayed in Appendix J.

**Table 50. Pooled standardized effect sizes from trials for improvement in urogenital atrophy among vaginal estrogen doses. The estimate in each cell represents comparison of the treatment intersecting the diagonal above it to that on the right (n is number of trials pooled). For example, the pooled standardized effect size for standard dose estrogen versus placebo is -0.41.**

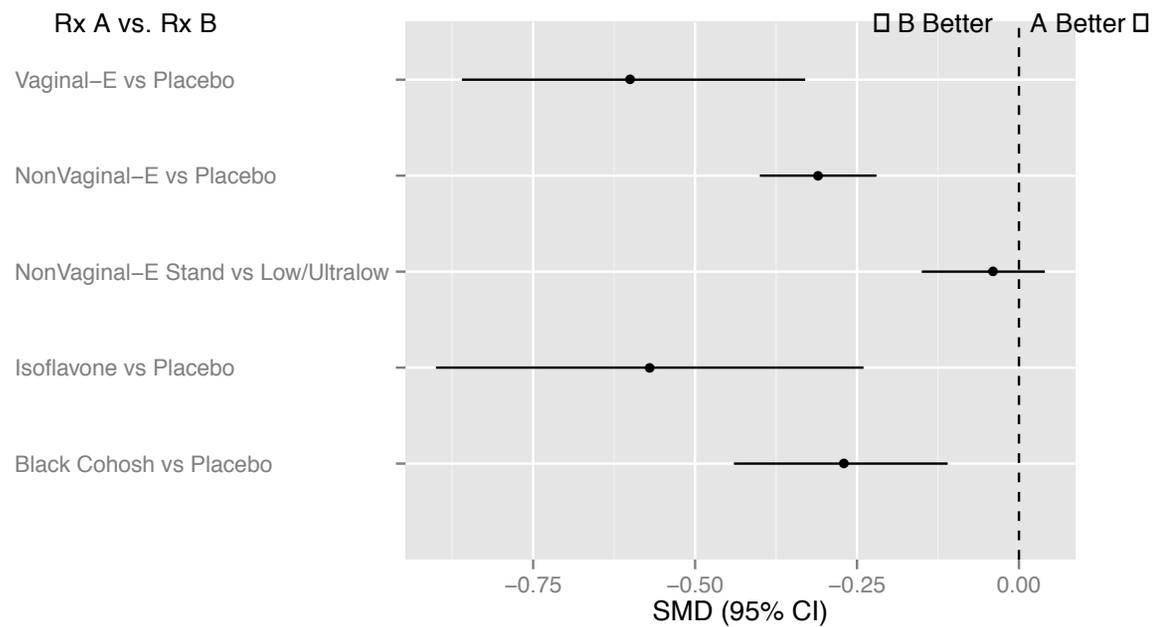
Standardized Effect Sizes				
Any Estrogen	E-High			
		E-Standard		
		0.11 (-0.07 to 0.28) tau <sup>2</sup> <0.0001; n=3	E-Low/Ultralow	
-0.60 (-0.86 to -0.33) tau <sup>2</sup> =0.17; n=11	-0.14 (-0.37 to 0.09) n=1	-0.41 (-0.60 to -0.22) tau <sup>2</sup> <0.0001; n=3	-0.73 (-1.07 to -0.38) tau <sup>2</sup> =0.18; n=7	Placebo

**Table 51. Pooled standardized effect sizes from single trials for improvement in urogenital atrophy among nonvaginal agents. The estimate in each cell represents comparison of the treatment intersecting the diagonal above it to that on the right (n is number of trials pooled). For example, the pooled standardized effect size for standard-dose estrogen versus placebo is -0.31.**

Other Routes of Administration						
Standardized Effect Sizes						
Any Estrogen	E-High					
	0.11 (-0.09 to 0.30) n=1	E-Standard				
		-0.04 (-0.15 to 0.07) tau <sup>2</sup> =0.08; n=4	E-Low/Ultralow			
			Isoflavone			
			-0.99 (-1.51 to -0.47) n=1	Black Cohosh		
-0.31 (-.40 to -0.22) <sup>a</sup> tau <sup>2</sup> =0.02; n=16	-0.34 (-0.68 to 0.0) n=1	-0.31 (-0.46 to -0.16) tau <sup>2</sup> =0.02; n=6	-0.34 (-0.48 to -0.21) tau <sup>2</sup> =0.03; n=10	-0.57 (-0.90 to -0.24) tau <sup>2</sup> =0.10; n=5	-0.27 (-0.44 to -0.11) tau <sup>2</sup> <0.0001; n=2	Placebo

<sup>a</sup>Includes 16 comparison because one trial<sup>186</sup> included two arms with different estrogen doses

**Figure 12. Urogenital atrophy life caterpillar plot displaying pooled comparisons and 95 percent confidence intervals**



## Estrogen Compared With Placebo

### Vaginal Estrogens

Eleven trials compared vaginal estrogens with placebo (Table 50). The modes of administration among these trials included creams, rings, ovules, and pessaries. One trial compared high-dose estrogens with placebo,<sup>174</sup> three trials compared standard-dose estrogens with placebo,<sup>219, 223, 231</sup> and seven trials compared low- or ultralow-dose estrogens with placebo.<sup>218, 221, 222, 225, 251, 252</sup> Pooled trial results found any vaginal estrogen significantly improved reported urogenital atrophy symptoms compared with placebo (SMD -0.60; 95 percent CI: -0.86 to -0.33;  $\tau^2=0.17$ ; 11 comparisons). One potential outlier<sup>223</sup> was apparent (Appendix J); excluding it diminished the estimated effect size and resulted in less heterogeneity (SMD -0.48; 95 percent CI: -0.71 to -0.57;  $\tau^2=0.13$ ). Analyses by estrogen dose category (high, standard, and low or ultralow) showed improvement in urogenital atrophy symptoms compared with placebo for standard and low doses. There was a single high-estrogen dose trial (two estrogen arms versus placebo ring),<sup>174 75</sup> in one arm there was a significant effect (SMD -2.42; 95 percent CI: -2.65 to -2.16) but not the other (result for both arms combined is shown in Table 50). Standardized effect sizes for standard- and low- or ultralow-dose vaginal estrogens compared with placebo were -0.41 (95 percent CI: -0.60 to -0.22;  $\tau^2<0.0001$ ; three comparisons), and -0.73 (95 percent CI: -1.07 to -0.38;  $\tau^2=0.18$ ; seven comparisons), respectively (Table 50). The strength of evidence that vaginal estrogens improve urogenital atrophy compared with placebo symptoms is rated as high.

### Nonvaginal Estrogens

Sixteen trials compared nonvaginal estrogens with placebo (Table 51). The modes of administration included oral, transdermal patch, and skin gel. One trial compared high dose estrogens with placebo,<sup>173</sup> six trials compared standard dose estrogens with placebo,<sup>26, 179, 186, 226, 253, 254</sup> and 10 trials compared low/ultralow dose estrogens with placebo.<sup>145, 186, 194, 218, 227, 229, 230, 255-257</sup> One of the 16 trials had three arms, comparing placebo with both a standard and low estrogen dose.<sup>186</sup> Pooled trial results showed that any estrogen improved urogenital atrophy symptoms compared with placebo (SMD -0.31 (95 percent CI: -0.68 to -0.22;  $\tau^2=0.02$ ). Analyses according to estrogen dose category (high, standard, and low/ultralow) showed improvement in all alleviating urogenital atrophy symptoms. The standardized effect sizes for high dose, standard dose, and low/ultralow dose estrogens were: -0.34 (95 percent CI: -0.68 to 0.00); one comparison), -0.31 (95 percent CI: -0.46 to -0.16;  $\tau^2=0.02$ ; six comparisons), and -0.34 (95 percent CI: -0.48 to -0.21;  $\tau^2=0.03$ ; 10 comparisons), respectively (Table 51). The strength of evidence that nonvaginal estrogens improve urogenital atrophy symptoms compared with placebo is rated high.

### Isoflavones Compared With Placebo

Five trials compared isoflavones with placebo.<sup>122, 150, 195, 258, 259</sup> Isoflavone dose ranged from 60 mg per day to 350 mg per day. Treatment arms enrollment ranged from 44 to 60 women. The trials found improved urogenital atrophy symptoms among those taking isoflavones compared with placebo (SMD -0.57 (95 percent CI: -1.07 to -0.38;  $\tau^2=0.18$ ). The strength of evidence that isoflavones compared with placebo improve urogenital atrophy symptoms is rated as low.

## Nonprescription Agents Compared With Placebo

Two trials compared black cohosh with placebo. The trials showed a significant improvement in urogenital atrophy symptoms among those taking black cohosh (SMD -0.27 (95 percent CI: -0.44 to -0.11;  $\tau^2 = 0.00$ ). The strength of evidence that black cohosh compared with placebo improves urogenital atrophy symptoms is rated as low.

## Trials Not Pooled

### Estrogen Compared with Other Hormones

One trial (Table 52) compared estrogen/progestin versus estrogen/progestin plus testosterone.<sup>237</sup> Estrogen/progestin doses were identical in both groups, with the experimental group receiving 2 mg testosterone. Both groups reported significant improvements in vaginal dryness. There was no difference in the magnitude of improvement between the groups.

**Table 52. Hormone therapies compared, reporting urogenital atrophy outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	SMD (95% CI)
Penteado, 2008 <sup>237</sup>	CEE + MPA	0.625 E + 2.5 P	27	Oral	52	Poor	—
	CEE + MPA + testosterone	0.625 E + 2.5 P + 2 T	29	Oral			

SMD: standardized mean difference; CI: confidence interval; CEE: conjugated equine estrogen; E: estrogen; MPA: medroxyprogesterone acetate; P: progestin; T: testosterone; FU: followup; Wks: weeks.

### Different Modes of Estrogen Administration

Seven trials compared similar estrogen doses administered by different modes (Table 53)<sup>110, 157, 244, 247-249, 260</sup> (see Appendix G for dose categorization by route). One trial showed a significant improvement administering estrogen via pessary compared with tablet;<sup>260</sup> one trial showed a patch was better than spray ( $p < 0.02$ ).<sup>110</sup> All other trials reported no difference between the modes of administration. The strength of evidence that mode of administration modifies estrogen efficacy on urogenital symptoms is rated as low.

**Table 53. Trials comparing different modes of estrogen administration, reporting urogenital atrophy symptoms**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	SMD (95% CI)
Odabasi, 2007 <sup>110</sup>	Estradiol + progestin	0.3 E + 90 P	32	Spray	12	Poor	—
	Estradiol + progestin	0.05 E + 90 P	29	Patch			
Weisberg, 2005 <sup>247</sup>	Estradiol	0.008	126	Ring	48	Poor	—
	Estradiol	0.025	59	Tablet			
Dugal, 2000 <sup>260</sup>	Estradiol	0.025	48	Tablet	24	Poor	—
	Estriol	0.5	48	Pessary			
Lubbert, 1997 <sup>157</sup>	Estradiol + progestin	0.05 E + P <sup>a</sup>	1232	Patch – continuous	12	Poor	—
	Estradiol + progestin	0.05 E + P	1227	Patch – cyclical			
Barentsen, 1997 <sup>244</sup>	Estradiol	0.0075	83	Ring	12	Poor	—
	Estriol	0.5	82	Cream			
Henriksson, 1994 <sup>249</sup>	Estradiol	0.0095	106	Ring	12	Poor	—
	Estriol	0.5	51	Pessary			
Long,	CEE	0.625	37	Oral	12	Poor	—

2006 <sup>248</sup>	CEE	0.625	36	Cream				-0.32(-0.78 to 0.14)
---------------------	-----	-------	----	-------	--	--	--	----------------------

<sup>a</sup> Recommended 5 mg/day dose but various agents and doses used.

SMD: standardized mean difference; CI: confidence interval; CEE: conjugated equine estrogen; E: estrogen; P: progesterin; FU: followup; Wks: weeks.

## Nonprescription Agents Compared With Placebo

Five trials (Table 54) compared nonprescription agents with placebo. Two examined dehydroepiandrosterone (DHEA),<sup>136, 156</sup> and three trials compared plant extracts,<sup>123, 126, 130</sup> Findings among the two DHEA trials were inconsistent. Due to the variety of dosages and treatments, pooling was not appropriate.

**Table 54. Trials comparing nonprescription agents with placebo, reporting urogenital atrophy symptoms**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	SMD (95% CI)
Heger, 2006 <sup>123</sup>	Placebo	—	55	Oral	12	Poor	—
	Rheum raphonticum	4	54	Oral			-1.33 (-1.74 to -0.91)
Hirata, 1997 <sup>126</sup>	Placebo	—	36	Oral	24	Poor	—
	Dong quai	4,500	35	Oral			-0.42 (-0.89 to 0.05)
Labrie, 2009 <sup>156</sup>	Placebo	—	53	Ovule	12	Poor	—
	DHEA	3.25	53	Ovule			-0.79 (-1.18 to -0.39)
	DHEA	6.5	56	Ovule			-0.52 (-0.90 to -0.14)
	DHEA	13	54	Ovule			-0.71 (-1.10 to -0.32)
Barnhart, 1999 <sup>136</sup>	Placebo	—	30	Oral	12	Poor	—
	DHEA	50	30	Oral			-0.06 (-0.57 to 0.45)
Chang, 2011 <sup>130</sup>	Placebo	—	33	Oral	12	Fair	—
	EstroG-100® <sup>a</sup>	257	31	Oral			-0.39 (-0.88 to 0.11)

SMD: standardized mean difference; CI: confidence interval; DHEA: dehydroepiandrosterone; FU: followup; Wks: weeks.

<sup>a</sup> combination of cynanchum wilfordii, phlomis umbrosa, angelica gigas

## Nonprescription Agents Compared With Nonprescription Agents

One trial (Table 55) compared isoflavones versus isoflavones combined with pine bark extract<sup>141</sup> and one trial compared different dosages of pueraria mirifica.<sup>140</sup> The isoflavone trial reported a minimal improvement with the addition of pine bark extract and the pueraria mirifica trial reported no difference between dosages.

**Table 55. Trials comparing nonprescription agents, reporting urogenital atrophy outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	SMD (95% CI)
Agosta, 2011 <sup>141</sup>	Isoflavones	60	301	Oral	12	Poor	—
	Isoflavones/pine bark extract	60	335	Oral			-0.16 (-0.31 to 0.00)
Virojchaiwong, 2011 <sup>140</sup>	Pueraria mirifica	25	26	Oral	26	Poor	—
	Pueraria mirifica	50	26	Oral			0.42 (-0.13 to 0.97)

SMD: standardized mean difference; CI: confidence interval; FU: followup; Wks: weeks.

## Trials With No Quantifiable Data

Publications from six trials lacked sufficient data to estimate effect sizes (standardized or otherwise). Results of these trials would not have affected the overall outcomes presented above. One trial was rated as fair quality<sup>261</sup> and the remainder rated as poor quality.

Schulman et al.<sup>261</sup> compared placebo with a low dose estrogen patch given with two different progestin doses. After 12 weeks, vaginal dryness was reported less frequently in both estrogen arms compared with placebo (p=0.013 and p=0.016).

Le Donne et al. conducted a 3-month, randomized, double-blind trial comparing 5 mg hyaluronic acid (n=31) with 97 µg genistein (n=31), administered through vaginal suppository.<sup>262</sup> Outcomes were reported as median genital score and both treatments provided significant relief of symptoms.

A randomized, double-blind trial compared the effect of pomegranate seed oil (n=43) with placebo (n=38).<sup>129</sup> Outcomes were reported as pre- and post-median scores in the urogenital domain of the Menopause Rating Scale. Women in the treatment and placebo arms experienced the same improvement in scores.

A trial comparing hormone therapy (n=30) with pueraria mirifica (n=30) reported that neither treatment affected vaginal dryness significantly.<sup>139</sup> The outcome was measured by the modified Greene Climacteric Scale.

Al-Azzawi et al. conducted a randomized, double-blind trial comparing oral estrogen/progestin (n=75) with estrogen/progestin administered through a vaginal ring (n=84).<sup>263</sup> Outcomes included a vaginal dryness symptom intensity score, which decreased in both treatment groups, but no variance estimates or p-values were reported.

A randomized, double-blind trial compared the effectiveness of a vaginal ring administering placebo (n=34) or estrogen (n=33) on urogenital symptoms.<sup>224</sup> Outcomes reported included physician-assessed vaginal dryness, which decreased in both arms, with no difference between the groups.

## Strength of Evidence Ratings—Urogenital Atrophy

**Table 56. Strength of evidence ratings domains for urogenital atrophy**

Comparisons	Comparators <sup>a</sup>		Risk of Bias	Consistency	Directness	Precision	Reporting Bias	SOE	Downgrading Rationale
11	Estrogen vaginal	vs. Placebo	M	C	D	P	U	High	
17	Estrogen oral	vs. Placebo	M	C	D	P	U	High	
5	Isoflavone	vs. Placebo	H	C	D	I	U	Low	Poor-quality trials; wide confidence interval for pooled effect
2	Black Cohosh	vs. Placebo	M	U	D	I	U	Low	One poor-quality trial; too few trials to assess consistency
7	Estrogen mode a	vs. Estrogen mode b	H	I	D	P	U	Low	All poor-quality trials; 2 trials showed differences; wide confidence intervals for trials

Risk of Bias: High (H), Medium (M) Low (L); Consistency: Inconsistent (I) Unknown (U) Consistent (C); Directness: Indirect (I) Direct (D); Precision: Imprecise (I) Unknown (U) Precise (P); Reporting Bias: Suspected (S) Undetected (U); SOE: strength of evidence.

<sup>a</sup>Bold font of comparator indicates the more effective treatment; if both comparators are bold, the treatments are equivalently effective

## Key Points—Treatment of Urogenital Atrophy

- Sixty-three trials including more than 18,000 women, reported on urogenital atrophy outcomes of treatment with estrogen or nonprescription agents such as isoflavones, black cohosh and herbal extracts.
- Study quality was generally rated as poor (84 percent). Thirty trials reported industry funding and 11 reported public funding.
- Results were reported using a variety of scales to assess symptoms of urogenital atrophy. The most common outcome was change in vaginal dryness.
- Strength of evidence of relative effectiveness of agents in ameliorating symptoms of vaginal atrophy:
  - There is **high** strength of evidence that vaginal and nonvaginal estrogens improve symptoms of urogenital atrophy.
  - 11 trials including 3,368 women compared vaginal estrogens with placebo. Overall, vaginal estrogens improved urogenital atrophy symptoms compared with placebo (SMD: -0.60, 95 percent CI: -0.86 to -0.33).
  - 16 trials including 5,921 women compared nonvaginal estrogens with placebo. Nonvaginal estrogens improved urogenital atrophy symptoms compared with placebo (SMD: -0.31, 95 percent CI: -0.40 to -0.22). Modes of administration included oral, transdermal patch and skin gel; and dosages included high, standard and low/ultralow.
  - There is **low** strength of evidence that isoflavones improve symptoms of urogenital atrophy. Five trials including 496 women compared isoflavones with placebo. Sample sizes were small, but the trials showed improvement in urogenital atrophy symptoms (SMD: -0.57, 95 percent CI: -0.90 to -0.24).
  - There is **insufficient** evidence to determine whether any other nonprescription agent improve symptoms of vaginal atrophy.

# Sleep

## Included Trials

Of the 254 included trials, 46 (18.1 percent) reported sleep outcomes (23 trials specified sleep as a primary outcome). In addition to placebo, the most common comparator included hormones (n=25), isoflavones (n=6), and nonprescription agents including ginseng and other various herbal extracts (Table 57).

The 46 trials originated from 22 different countries in Europe (n=18) and Asia (n=8), and in the United States (n=8), Australia (n=3), and South America (n=2); seven trials were multinational. Of the multinational studies, six involved European countries and one Asian. The trials were conducted at 1,434 sites with followup ranging from 4 weeks in the gabapentin trial to 260 weeks.

Sleep dysfunction outcomes were reported using a variety of measures and scales. The most common outcome reported was percent insomnia (13 trials). Other measurements included subscales of the Women's Health Questionnaire (WHQ) (eight trials), Kupperman Menopausal Index (eight trials), Greene Climacteric Scale (six trials), WHI Insomnia Rating Scale (two trials), and Menopausal Rating Scale (MRS) (three trials). Other trials reported sleep using graphic rating scales. Following are brief descriptions of the most commonly used scales:

- WHQ consists of nine domains, with three questions comprising the sleep domain: waking early, sleeping badly for the rest of the night, and difficulty in falling asleep. A 4-point scale is used to answer the questions, the answers are converted to binary scores, then the total score is divided by number of questions per domain. WHQ domain scores range from 0 to 1.<sup>264</sup>
- Kupperman Index assesses 11 menopausal symptoms, including insomnia. Each symptom is scored from 0 (no symptoms) to 3 (most severe).<sup>172</sup>
- Greene Climacteric Scale has a single question about difficulty in sleeping, which is scored on a 4-point scale, from 0 (none) to 3 (severe).
- WHI Insomnia Rating Scale consists of four questions: trouble falling asleep, waking several times at night, waking up earlier than planned, and trouble falling back asleep. A 5-point scale is used to answer the questions and is coded so that the higher score indicates more severe insomnia.<sup>93</sup>
- MRS includes one question encompassing difficulty in falling asleep, difficulty in sleeping through the night, and waking up early, scaled from 0 (none) to 4 (extremely severe).

Study quality was generally rated as poor (41 of the 46 trials). Funding sources were unreported in 18 trials, while industry funding was indicated in 20 trials, and solely public funding was cited in 10 trials. Table 57 provides a summary of additional trial and patient characteristics.

**Table 57. Characteristics of trials assessing efficacy for sleep dysfunction**

	Characteristic	Number (%)
Trial Characteristics	Number of Trials	46
	Total number of patients	43,710
	Number of sites from trials specifying (n=36)	1,434
		1 to 502 (mean 42; median 3)

	Trials described only as multicenter	6 (13.0)
	Multicenter trials	27 (58.7)
	Two-arm	36 (78.3)
	Multi-arm	10 (21.7)
	Patients per trial	50 to 16,608 (mean 950; median 142)
	Range of followup (weeks)	4 to 260 (mean 38.8; median 24)
Funding	Industry only	13 (28.3)
	Public only	8 (17.4)
	Industry and public	7 (15.2)
	Not stated	18 (39.1)
Comparator Category	Placebo vs. hormone	19 (41.3)
	Antidepressant vs. placebo or other antidepressant	1 (2.2)
	Placebo vs. other prescription	2 (4.3)
	Placebo vs. nonprescription	15 (32.6)
	Placebo vs. hormone vs. nonprescription	0 (0.0)
	Hormone vs. hormone	5 (10.9)
	Hormone vs. nonprescription	1 (2.2)
	Nonprescription vs. antidepressant	0 (0.0)
Study Quality	Good	4 (8.7)
	Fair	1 (2.2)
	Poor	41 (89.1)
Patient Demographics	Mean age (years)	43.8 to 66.7 (NR 7)
	Age range (years)	40.0 to 81.0 (NR 39)
	Years since menopause	2.8 (0.6 to 16.5) (NR 33)
	Current Smokers (percent)	0.0 to 44.0 (NR 34)
	Mean BMI (kg/m <sup>2</sup> )	23.4 to 30.1 (NR 17)
	White (%)	0.0 to 100.0
	Black (%)	0.0 to 58.8
	Hispanic (%)	0.0 to 6.1
	Asian (%)	0.0 to 100.0
	Other (%)	0.0 to 2.0
Uterus Status	All intact	16 (34.8)
	All absent	4 (8.7)
	Mixed	12 (26.1)
	Range, percentage intact among trials with mixed	47.7 to 94.3
	Not reported	14 (30.4)

Note: Demographics were not reported in all studies.

NR: not reported.

## Evidence Synthesis for Sleep Dysfunction

As for other outcomes, standardized mean differences were calculated to allow comparing outcomes across different sleep scales. Pooling was performed for pairwise comparisons where evidence included more than a single comparison. Forest plots are displayed in Appendix K.

## Estrogen Compared With Placebo

Estrogen-placebo comparisons were included in 24 trials (25 comparisons). One trial compared high-dose estrogens with placebo,<sup>173</sup> 12 trials compared standard-dose estrogens with placebo,<sup>26, 28, 149, 153, 161, 176, 179, 180, 185, 188, 226, 253</sup> and seven trials compared low-dose estrogens with placebo.<sup>145, 188, 190, 255-257, 265</sup> Analyses according to estrogen dose (high, standard, and low) showed improvements in sleep compared with placebo in each category, though the confidence intervals were wide for the high- and low-dose estrogens due to the small number of trials in those categories. The standardized effect sizes for high-dose, standard-dose, and low-dose estrogens were: 0.57 (95 percent CI: 0.23 to 0.92; one comparison), 0.33 (95 percent CI: 0.22 to 0.44;  $\tau^2=0.02$ ; 12 comparisons), and 0.42 (95 percent CI: 0.09 to 0.75;  $\tau^2=0.15$  [including results from Gambacciani 2003<sup>190</sup> an outlier yielded 0.60]), respectively (Table 58). Compared with placebo, the strength of evidence that estrogens improve self-reported sleep metrics among menopausal women is rated as high.

## Estrogen Compared With Estrogen

Three trials compared standard-dose with low- or ultralow-dose estrogen (all rated as poor quality).<sup>188, 243, 266</sup> No difference was apparent in effect on sleep metrics (SMD -0.24; 95 percent CI: -0.67 to 0.18,  $\tau^2=0.13$ ) (strength of evidence is rated as low).

## Isoflavones Compared With Placebo

Six trials compared isoflavones with placebo (five rated as poor quality).<sup>122, 150, 151, 195, 200, 258</sup> Three trials reported nonsignificant findings<sup>195, 200, 258</sup> and three trials reported improvements with isoflavones compared with placebo.<sup>122, 150, 151</sup> There were a total of 531 participants in the six trials with considerable heterogeneity ( $\tau^2=0.95$  exceeding the pooled SMD of 0.63 when all trials were included). Pooled trial results found no significant difference in reported sleep metrics for all six trials or excluding the one outlier (Table 58).<sup>200 228</sup> The strength of evidence that isoflavones improve self-reported sleep metrics among menopausal women was rated insufficient.

## Gabapentin Compared With Placebo

A single trial<sup>73</sup> compared gabapentin treatment with placebo on sleep dysfunction. In this double-blind placebo-controlled trial, the authors reported no significant difference in sleep dysfunction between the treatment groups. The standardized effect size was 0.44 (95 percent CI: -0.07 to 0.44).

## SSRIs Compared With Placebo

A single trial<sup>205</sup> compared antidepressant treatment with placebo on sleep dysfunction. The double-blind placebo-controlled trial randomized women to placebo (n=50), fluoxetine (n=50), and citalopram (n=50). The authors reported a significant improvement in sleep dysfunction in the citalopram group compared with placebo, but no difference in sleep dysfunction in the fluoxetine group compared with placebo. The standardized effect size with either SSRI was 0.19 (95 percent CI: -0.15 to 0.52).

**Table 58. Pooled standardized effect sizes from trials for improvement in sleep dysfunction. The estimate in each cell represents comparison of the treatment intersecting the diagonal above it to that on the right (n=number of trials pooled). For example, the pooled standardized effect size for standard dose estrogen versus placebo is 0.25.**

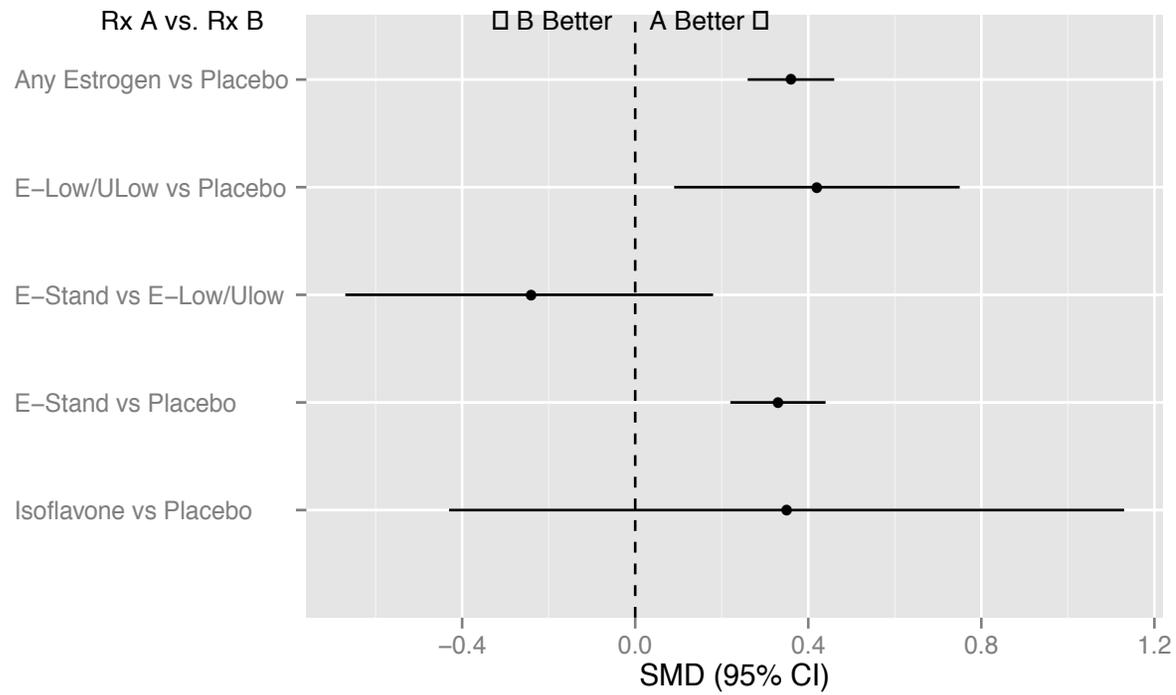
E-High						
0.50 (0.30 to 0.70) n=1	E-Standard					
	-0.24 (-0.67 to 0.18) tau <sup>2</sup> =0.13; n=3	E-Low/Ultralow				
Isoflavone						
Gabap/Preg						
SSRI/SNRI						
0.57 (0.23 to 0.92) n=1	0.33 (0.22 to 0.44) tau <sup>2</sup> =0.02; n=12	0.42 (0.09 to 0.75) <sup>a</sup> tau <sup>2</sup> =0.15; n=6	0.35 (-0.43 to 1.13) <sup>b</sup> tau <sup>2</sup> =0.74; n=5	0.44 (-0.70 to 0.44) n=1	0.19 (-0.15 to 0.52) n=1	Placebo

E: estrogen; Gabap/Preg: gabapentin and pregabalin.

<sup>a</sup> Including Gambacciani <sup>190</sup> 0.60 (0.23 to 0.97) tau<sup>2</sup>=0.22; n=7

<sup>b</sup> Including Jassi 2010 <sup>200 228</sup> 0.63 (-0.18 to 1.43) tau<sup>2</sup>=0.95; n=6

**Figure 13. Caterpillar plot displaying all pooled sleep comparisons and 95% confidence intervals.**



## Trials Not Pooled

### Estrogens

Two trials compared estrogens in similar doses and reported sleep outcomes (Table 59). One trial compared a ring and tablet delivering a standard estrogen dose and found no difference in reported sleep outcomes.<sup>247</sup> Another trial compared estrogen patches administered with sequentially or combined progestin, and also reported no differences.<sup>157</sup>

**Table 59. Trials comparing similar estrogen doses, and reporting sleep outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	SMD (95% CI)
Weisberg, 2005 <sup>247</sup>	Estradiol	0.008	126	Ring	48	Poor	-0.17 (-0.48 to 0.14)
	Estradiol	0.025	59	Tablet			
Lubbert, 1997 <sup>157</sup>	Estradiol, combined	0.05	1232	Patch	12	Poor	NS
	Estradiol, sequential	0.05	1227	Patch			

SMD: standardized mean difference; CI: confidence interval; NS: not significant; FU: followup; Wks: weeks.

### Estrogen Compared With Nonprescription Agents

One trial (Table 60) compared standard-dose estrogen with an herbal extract for the treatment of sleep dysfunction.<sup>139</sup> In this unblinded trial, 30 women received a standard dose of conjugated equine estrogen with or without medroxyprogesterone acetate depending on uterine status, and 30 women received pueraria mirifica, a medicinal herb containing phytoestrogens. There was no significant difference in reported outcomes between treatment groups.

**Table 60. Trial comparing hormone with a nonprescription agent, reporting sleep outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	SMD (95% CI)
Chandeying, 2007 <sup>139</sup>	CEE + MPA	0.625 E + 2.5 P	Oral	30	24	Poor	NS
	Pueraria mirifica	NA	Oral	30			

SMD: standardized mean difference; CI: confidence interval; CEE: conjugated equine estrogen; MPA: medroxyprogesterone; P: progestin; NA not applicable; NS: not significant; FU: followup; Wks: weeks..

### Prescription Agents Compared With Placebo

One randomized, double-blind trial (Table 61) compared eszopiclone, a treatment used for insomnia (n=30), with placebo (n=29) and reported Insomnia Severity Index scores.<sup>108</sup> The trial was rated as poor quality with a substantial effect (SMD: 1.51; 95 percent CI: 0.94 to 2.08).

**Table 61. Trials comparing prescription treatments with placebo reporting sleep outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	SMD (95% CI)
Joffe, 2010 <sup>108</sup>	Placebo	—	29	Oral	4	Poor	1.51 (0.94 to 2.08)
	Eszopiclone	3	30	Oral			

SMD: standardized mean difference; CI: confidence interval; NS: not significant; FU: followup; Wks: weeks.

## Other Nonprescription Agents Compared With Placebo

Nine trials compared different nonprescription agents with placebo (Table 62): St. John's wort,<sup>135</sup> pine bark extract,<sup>121</sup> rheum raphanicum,<sup>123</sup> ginseng (two trials),<sup>101, 102</sup> diascorea alata,<sup>128</sup> DHEA,<sup>136</sup> pomegranate seed oil,<sup>129</sup> and herbal extract.<sup>130</sup> Neither ginseng trial (both rated as poor quality) showed an effect on reported sleep (strength of evidence is rated as insufficient).

**Table 62. Trials comparing nonprescription agents with placebo and reporting sleep outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	SMD (95% CI)
Van Die, 2009 <sup>135</sup>	Placebo	—	50	Oral	16	Good	—
	St. John's wort/Chaste tree	900	50	Oral			
Yang, 2007 <sup>121</sup>	Placebo	—	75	Oral	24	Poor	—
	Pine bark extract	200	80	Oral			
Heger, 2006 <sup>123</sup>	Placebo	—	55	Oral	12	Poor	—
	Rheum raphanicum	4	54	Oral			
Wiklund, 1999 <sup>101</sup>	Placebo	—	191	Oral	16	Poor	—
	Ginseng	200	193	Oral			
Hartley, 2004 <sup>102</sup>	Placebo	—	27	Oral	12	Poor	—
	Ginseng	200	30	Oral			
Hsu, 2011 <sup>128</sup>	Placebo	—	25	Oral	52	Poor	—
	Diascorea alata (yam)	24	25	Oral			
Barnhart, 1999 <sup>136</sup>	Placebo	—	30	Oral	12	Poor	—
	DHEA	50	30	Oral			
Auerbach, 2012 <sup>129</sup>	Placebo	—	38	Oral	12	Poor	—
Pomegranate seed oil	0.254	43	Oral				
Chang, 2011 <sup>130</sup>	Placebo	—	33	Oral	12	Fair	—
EstroG-100® <sup>a</sup>	251	31	Oral				

SMD: standardized mean difference; CI: confidence interval; NS: not significant; FU: followup; Wks: weeks.

<sup>a</sup> combination of cynanchum wilfordii, phlomis umbrosa, angelica gigas

## Nonprescription Agents Compared

Three trials (Table 63) compared nonprescription treatments with other nonprescription treatments. One trial compared isoflavones with isoflavones plus magnolia bark, and reported that the treatment group with magnolia bark experienced significant improvements in sleep dysfunction.<sup>141</sup> Another trial compared two different dosages of isoflavones combined with vitamin E, and reported significant improvements in sleep dysfunction with both groups.<sup>146</sup> One trial compared isoflavones with vitamin E.<sup>142</sup> Pooling of the results was not appropriate.

**Table 63. Trials comparing nonprescription agents, and reporting sleep outcomes**

Trial	Treatment	Dose (mg)	N	Mode	FU Wks	Study Quality	SMD (95% CI)
Agosta, 2011 <sup>141</sup>	Isoflavones	60	301	Oral	12	Poor	—
	Isoflavones + magnolia bark	60	335	Oral			
Hidalgo, 2006 <sup>146</sup>	Isoflavones + vitamin E	60	478	Oral	26	Poor	—
	Isoflavones + vitamin E	120	447	Oral			
Zervoudis, 2008 <sup>142</sup>	Isoflavones	NR	31	Oral	52	Poor	—
	Vitamin E	500 IU	31	Oral			

SMD: standardized mean difference; CI: confidence interval; NS: not significant; FU: followup; Wks: weeks; NR: not reported

## Strength of Evidence Ratings—Sleep Outcomes

Table 64. Strength of evidence ratings domains for sleep outcomes

Comparisons	Comparators <sup>a</sup>		Risk of Bias	Consistency	Directness	Precision	Reporting Bias	SOE	Downgrading Rationale
25	<b>Estrogen</b>	vs. Placebo	M	C	D	P	U	High	—
3	<b>Estrogen</b>	vs. <b>Estrogen (different dose)</b>	H	C	D	I	U	Insuff	3 poor-quality trials; wide CI; consistently no difference
6	<b>Isoflavone</b>	vs. Placebo	H	I	D	I	U	Insuff	5 poor-quality trials; magnitude and direction of effects heterogeneous; confidence interval overlapping 0
2	<b>Ginseng</b>	vs. Placebo	H	I	D	I	U	Insuff	2 poor-quality trials; CIs for SMD overlap 0 in one trial, no significant effect reported in the other

Risk of Bias: High (H), Medium (M), Low (L); Consistency: Inconsistent (I), Unknown (U), Consistent (C); Directness: Indirect (I), Direct (D); Precision: Imprecise (I), Unknown (U), Precise (P); Reporting Bias: Suspected (S), Undetected (U); SOE: strength of evidence; Insuff: insufficient; SMD: standardized mean difference; CI: confidence interval.

<sup>a</sup> Bold font of comparator indicates the more effective treatment; if both comparators are bold, the treatments are equivalently effective

### Key Points

- Forty-six trials including a total of 43,710 women reported on sleep dysfunction in women treated with prescription agents (estrogen, antidepressant, gabapentin) and nonprescription agents (isoflavones, St. John’s Wort, pine bark extract, rheum rhaponticum, ginseng, dioscorea alata, DHEA, pomegranate seed oil, and herbal extract)
- Forty-one of 46 trials were rated as poor quality. Twenty trials reported industry funding, 10 were publicly funded, and funding sources were not reported in 20 trials.
- Results were reported from a variety of scales. The most common outcome reported was insomnia (13 trials)
- Strength of evidence of relative effectiveness of agents on ameliorating measures of sleep dysfunction:
  - There is **high** strength of evidence from 21 trials including more than 38,000 women that estrogen is the most effective agent in improving measures of sleep dysfunction. Combined results showed SMD 0.35 (95 percent CI: 0.25 to 0.44) compared with placebo. There is **low** strength of evidence from placebo-controlled trials and direct comparisons with no significant difference between standard- and low/ultralow-dose estrogen.
  - There is **insufficient** evidence to determine whether any other agent, prescription or nonprescription, is effective in ameliorating measures of sleep dysfunction.

## Compounded Hormone Therapies

There is insufficient evidence regarding the safety and efficacy of compounded “bioidentical” hormone therapy for treatment of menopausal symptoms. We were unable to identify any clinical trials comparing compounded hormone therapy for menopausal symptoms that met our criteria for inclusion. One randomized trial compared the pharmacokinetics of estrogen containing compounded “bioidentical” cream and a conventional “bioidentical” patch, but the outcome did not include a discussion of vasomotor symptoms or other harms/benefits, the study length was less than the 12-week duration for hormone trials and the number of participants was too low for inclusion in this review (NCT00864214).<sup>267</sup> Four evidence-based position statements from professional societies and special committee reports were reviewed and included in the report to illustrate the general consensus indicating that evidence-based research on compounded hormone therapy is lacking.<sup>53, 65, 66, 69-71</sup> Due to growing interest and an increase in prescriptions of compounded hormones, the limitations in the evidence base regarding the safety and efficacy of these therapies emphasizes the priority that should be given to future research. Many claims regarding the safety, efficacy, and superiority of compounded hormones have not been supported and FDA has voiced concern over pharmacies misleading patients and practitioners by unsupported claims of safety and greater efficacy than FDA-approved menopausal hormone therapies.

## Key Question 2. Long-Term Effects of Hormone Therapy Preparations

This key question addresses the long-term effects of hormone therapies on breast cancer; gall bladder disease; colorectal cancer; coronary heart disease, stroke, and thromboembolism; endometrial cancer; osteoporotic fractures; and ovarian cancer among women taking hormone therapies for menopausal symptom relief. Systematic reviews (SR) and meta-analyses (MA) provided the evidence base for this question.

As detailed in the Methods, selection was based on AHRQ guidance on incorporating existing SRs in comprehensive effectiveness reviews<sup>268</sup> and on a modified version of the AMSTAR tool.<sup>77</sup> First, SRs and MAs identified from the literature search were screened for relevance. Next, the following AMSTAR criteria were added as inclusion criteria to enable the assessment of potential bias: 1) at least two electronic sources were searched and key words and/or MeSH® terms were stated, 2) trial inclusion/exclusion criteria were adequately described, and 3) trial quality (risk for bias) of included studies was assessed and documented. Thirty SRs and one MA met the criteria. The MA examined ovarian cancer and menopausal hormone therapy.<sup>269</sup> Out of the 30 SRs, the SR with the most current literature search was the 2012 review conducted by Nelson et al. for the United States Preventive Services Task Force (USPSTF) comparing hormone therapy with placebo for the prevention of chronic conditions.<sup>37</sup> This report was comprehensive, addressing most outcomes included in this CER. Accordingly, this report along with the MA, were adopted as the primary sources for Key Question 2.

The Nelson et al. SR included 51 publications from nine RCTs collectively enrolling over 36,000 participants: the Women’s Health Initiative (WHI) combination estrogen plus progestin trial (referred to hereafter as “estrogen/progestin”),<sup>49, 270-273</sup> WHI estrogen only trial,<sup>271, 274, 275</sup> WHI Memory Study (WHIMS),<sup>276</sup> WHI Study of Cognitive Aging (WHISCA),<sup>277</sup> Heart and Estrogen/Progestin Replacement Study (HERS and HERS-II),<sup>278, 279</sup> Women’s International Study of Long Duration Oestrogen After Menopause (WISDOM),<sup>280</sup> Oestrogen in the Prevention

of Reinfarction Trial (ESPRIT),<sup>281</sup> Estrogen Memory Study (EMS),<sup>282</sup> and Ultra-Low-Dose Transdermal Estrogen Assessment (ULTRA).<sup>283</sup> The report also included a WHI followup published subsequent to the literature search.<sup>284</sup>

Among the trials identified by Nelson et al, four trials met our inclusion criteria for this key question: WHI estrogen/progestin, WHI estrogen only, HERS/HERS-II, and ESPRIT. WHIMS and WHISCA were excluded because the outcomes were not outcomes included in this key question. ULTRA was excluded due to a sample size of less than 250 per arm, and WISDOM and EMS were excluded because of short followup periods. Hazard ratios and 95% confidence intervals were abstracted from nine articles from the four trials (Table 65). Nelson et al. rated the overall quality of the body of evidence as fair, based on the number, quality, and size of studies; consistency of results between studies; and directness of effect.<sup>14</sup> Details of the study quality ratings from the nine articles included can be found in Appendix L, quality assessments.

Women enrolled in the trials were on average older than the target population of this CER. While there is overlap in the age groups, women seeking symptom relief are in general younger than the populations of WHI (mean age of 63 years) and HERS (mean age of 67 years). We identified observational studies from the original literature search enrolling peri- and recently menopausal women in order to inform the discussion on applicability. The clinical content expert was also queried regarding relevant publications. Consistency between trials with older populations and observational studies with younger populations was addressed in the strength of evidence discussion. These steps were added to those outlined in AHRQ guidance (which notes “the exact process needs to be flexible and will likely evolve”).

**Table 65. Evidence base for long-term effects of hormone therapies**

Condition	Estrogen/Progestin	Estrogen Alone
Breast cancer	Chlebowski 2010 <sup>270</sup> /WHI Hulley 2002 <sup>279</sup> /HERS/HERS-II	LaCroix 2011 <sup>275</sup> /WHI
Gall bladder disease	Cirillo 2005 <sup>271</sup> /WHI	Cirillo 2005 <sup>271</sup> /WHI
Colorectal cancer	Heiss 2008 <sup>272</sup> /WHI Hulley 2002 <sup>279</sup> /HERS/HERS-II	LaCroix 2011 <sup>275</sup> /WHI
Coronary heart disease, stroke, thromboembolism	Heiss 2008 <sup>272</sup> /WHI	LaCroix 2011 <sup>275</sup> /WHI Cherry 2002 <sup>281</sup> /ESPRIT
Endometrial cancer	Heiss 2008 <sup>272</sup> /WHI Hulley 2002 <sup>279</sup> /HERS/HERS-II	Previously established causal association
Osteoporotic fractures	Rossouw 2002 <sup>49</sup> /WHI Hulley 2002 <sup>279</sup> /HERS/HERS-II	Anderson 2004 <sup>274</sup> /WHI
Ovarian cancer	Anderson 2003 <sup>273</sup> /WHI	Greiser 2007/MA

ESPRIT: Estrogen in the Prevention of Reinfarction Trial; HERS/HERS II: Heart and Estrogen/Progestin Replacement Study; MA: meta-analyses; WHI: Women’s Health Initiative.

## Breast Cancer

### Summary

Three trials reported breast cancer incidence: WHI estrogen/progestin,<sup>270</sup> WHI estrogen-only,<sup>275</sup> and HERS-II.<sup>279</sup> All three trials administered oral conjugated equine estrogens (CEE) with the addition of medroxyprogesterone acetate in the estrogen/progestin trials. Mean followup ranged from 5.2 years in the WHI estrogen/progestin trial to 6.8 years in the HERS-II trial.

In the WHI trial, estrogen/progestin increased breast cancer risk compared with placebo whereas estrogen alone reduced the risk (Table 66 and Table 67). HERS-II found no significant increase in breast cancer risk in women using estrogen/progestin (Table 66).

Using only WHI data, the review by Nelson et al estimated that the use of estrogen/progestin increased invasive breast cancer incidence by eight additional events per 10,000 woman-years (95 percent CI: 3 to 14). However, the use of estrogen-only reduced invasive breast cancer incidence by eight fewer events per 10,000 woman-years (95 percent CI: 1 to 14).<sup>37</sup> A 2012 update to the WHI report presents consistent results for both estrogen/progestin and estrogen-only therapies.<sup>284</sup> The authors of this update caution that despite the risk reduction found in the estrogen-only trial, the use of estrogen for breast cancer risk reduction remains unsupported, particularly among the subgroup of women at increased breast cancer risk.

**Table 66. Overall breast cancer incidence among women treated with estrogen/progestin**

Trial	Treatment	N	Average Followup	Results
WHI – CEE + MPA <sup>270</sup>	0.625mg CEE + 2.5mg MPA	16,608	5.2 years	Overall: HR: 1.25; 95% CI: 1.07 to 1.46; p=0.004
HERS/HERS-II <sup>279</sup>	0.625mg CEE + 2.5mg MPA	2,321	6.8 years	Overall: HR: 1.08; 95% CI: 0.52 to 2.24; p=0.83

CEE: conjugated equine estrogen; HERS: Heart and Estrogen/Progestin Replacement Trial; HR: hazard ratio; MPA: medroxyprogesterone acetate; WHI: Women’s Health Initiative

**Table 67. Overall breast cancer incidence among women treated with estrogen alone**

Trial	Treatment	N	Average Followup	Results
WHI – CEE alone <sup>275</sup>	0.625mg CEE	10,739	6.8 years	Intervention: HR: 0.79; 95% CI: 0.61 to 1.02 Postintervention: HR: 0.75; 95% CI: 0.51 to 1.09 Overall: HR: 0.77; 95% CI: 0.62 to 0.95

CEE: conjugated equine estrogen; HR: hazard ratio; WHI: Women’s Health Initiative.

## Applicability

Evidence informing breast cancer risk in younger populations can be found in secondary analyses of the WHI trial<sup>285</sup> and in the Million Women Study, a large observational study.<sup>286</sup> In addition to focusing on younger women, these studies also explore potential treatment factors modifying breast cancer risk, including hormone treatment duration and time from menopause onset to hormone initiation (referred to as “gap time”).

In an analysis combining the WHI estrogen/progestin trial and the WHI observational study, women using estrogen/progestin therapy with a gap time of less than five years were at greater risk of breast cancer compared to women initiating therapy later.<sup>285</sup> However, there was no evidence in the WHI estrogen-only trial that women starting therapy soon after menopause were at increased breast cancer risk.<sup>287</sup>

The Million Women Study conducted in the United Kingdom also examined gap time and breast cancer risk, but report some findings that are inconsistent with WHI. Women taking estrogen/progestin experienced increased risk of breast cancer, whether gap time was less than five years (RR: 2.04; 95 percent CI: 1.97 to 2.12) or greater than five years (RR: 1.53; 95 percent CI: 1.38 to 1.69).<sup>286</sup> Women taking estrogen alone, with a gap time less than five years, experienced increased risk of breast cancer (RR: 1.43; 95 percent CI: 1.36 to 1.49), but did not experience an increased risk if gap time was greater than five years (RR: 1.05; 95 percent CI: 0.89 to 1.23).<sup>286</sup>

When assessing treatment duration, the WHI combined trial and observational study report that longer use plus a short gap time was associated with increased breast cancer risk. Among women who initiated estrogen/progestin therapy soon after menopause and had 10 years of use, the estimated HR was 2.19 (95% CI: 1.56 to 3.08).<sup>285</sup>

The Million Women Study reported that women using estrogen/progestin longer than five years, regardless of gap time, were at increased risk of breast cancer. However, the study also found that women using estrogen alone for longer than five years were at increased breast cancer risk only if gap time was less than five years.<sup>286</sup>

Trends in breast cancer incidence in relation to trends in hormone use should be noted. The WHI published a report in July 2002 explaining that the trial was stopped early because the number of invasive breast cancer events indicated that risks of hormone therapy were exceeding benefits.<sup>49</sup> Subsequently, the number of prescriptions for estrogen/progestin dropped 66 percent and for estrogen dropped 33 percent in January to June 2003 compared to the previous year.<sup>288</sup> In 2003, invasive breast cancer incidence decreased 10.6% in women 60 to 64 and 14.3% in women 65 to 69.<sup>289</sup>

## Conclusions

Two large RCTs, WHI<sup>270</sup> and HERS-II,<sup>279</sup> examined breast cancer risk accompanying estrogen/progestin treatment. Both trials were rated as fair quality. The hazard ratios are consistent and show an increased risk of breast cancer, though statistical significance was demonstrated only in the WHI trial. The measures were direct and precise. The strength of evidence is rated high that estrogen/progestin therapy increases breast cancer risk.

One large RCT, the WHI estrogen only trial,<sup>275</sup> examined breast cancer risk associated with estrogen alone treatment. Overall risk was significantly reduced. Trial quality was rated fair quality. The findings are inconsistent when intervention and postintervention phases are considered separately. An update to the WHI study cautions that results may not apply to subgroups of women, such as those at increased risk of breast cancer. The findings are also inconsistent with the results of the observational Million Women Study. The strength of evidence is rated low that estrogen alone reduces breast cancer risk.

## Gall Bladder Disease

### Summary

Two trials reported gall bladder disease incidence: WHI estrogen/progestin<sup>271</sup> and WHI estrogen-only.<sup>271</sup> Oral conjugated estrogens (CEE) were administered in both trials with the addition of medroxyprogesterone acetate in the estrogen/progestin trial. Women with prior gallbladder disease or cholecystectomy were excluded. Both trials found an increased incidence of gall bladder disease with estrogen/progestin and estrogen alone compared to placebo (Table 68 and Table 69).

Using WHI data, Nelson et al. calculated additional gall bladder disease events—defined as cholecystitis and cholelithiasis—attributable to hormone therapy. Estrogen/progestin use was associated with an additional 20 gall bladder disease events per 10,000 women-years (95 percent CI: 11 to 29); and estrogen-only therapy with an additional 33 events per 10,000 women-years (95 percent CI: 20 to 45).<sup>37</sup>

**Table 68. Gall bladder disease incidence among women treated with estrogen/progestin**

Trial	Treatment	N	Average Followup	Results
WHI – CEE + MPA <sup>271</sup>	0.625mg CEE + 2.5mg MPA	14,203	5.2 years	HR: 1.54; 95% CI: 1.22 to 1.94; p<0.001

CEE: conjugated equine estrogen; HR: hazard ratio; MPA: medroxyprogesterone acetate; WHI: Women’s Health Initiative

**Table 69. Gall bladder disease incidence among women treated with estrogen alone**

Trial	Treatment	N	Average Followup	Results
WHI – CEE alone <sup>271</sup>	0.625mg CEE	8,376	6.8 years	HR: 1.80; 95% CI: 1.42 to 2.28; p<0.001

CEE: conjugated equine estrogen; HR: hazard ratio; WHI: Women’s Health Initiative.

## Applicability

Though the WHI trials enrolled an older population, the increased risk of gall bladder disease among women using hormone therapy is supported by results from large observational cohort studies of younger populations. The Nurses’ Health Study found a relative risk for gall bladder disease of 2.1 (95 percent CI: 1.9 to 2.4)<sup>290</sup> and the Million Women Study 1.64 (95 percent CI: 1.58 to 1.69) for all current hormone therapy users.<sup>291</sup> In the Atherosclerosis Risk in Communities Study, compared to women who never used hormone therapy, former users had an age-adjusted relative risk for gall bladder disease of 1.84 (95 percent CI: 1.3 to 2.6) and current users had a risk of 1.76 (95 percent CI: 1.3 to 2.4).<sup>292</sup>

## Conclusions

The evidence for estrogen/progestin treatment and gall bladder disease risk consists of one large RCT, the WHI trial.<sup>271</sup> Trial quality was rated as fair. Consistency is unknown, but results from the trial are supported by the results of several large observational studies. The measures are direct and precise. The strength of evidence was rated moderate that estrogen/progestin increases gall bladder disease risk.

The evidence for treatment with estrogen alone and gall bladder disease risk consists of one large RCT, the WHI trial.<sup>271</sup> Trial quality was rated fair. Consistency is unknown, but the results of the trial are supported by the results of several large observational studies. The measures are direct and precise. The strength of evidence is rated moderate that estrogen alone increases gall bladder disease risk.

## Colorectal Cancer

### Summary

Three trials reported colorectal cancer incidence: WHI estrogen/progestin,<sup>272</sup> WHI estrogen-only,<sup>275</sup> and HERS-II.<sup>279</sup> Oral conjugated equine estrogen (CEE) was used in all three trials with the addition of medroxyprogesterone acetate in the estrogen/progestin trials. None of the trials reported an effect of hormone therapy on colorectal cancer incidence (Table 70 and Table 71).

**Table 70. Overall colorectal cancer incidence among women treated with estrogen/progestin**

Trial	Treatment	N	Average Followup	Results
WHI – CEE + MPA <sup>272</sup>	0.625mg CEE + 2.5mg MPA	16,608	5.2 years	Intervention: HR: 0.62; 95% CI: 0.43 to 0.89 Postintervention: HR: 1.08; 95% CI: 0.66 to 1.77 Overall: HR: 0.75.; 95% CI: 0.57 to 1.00
HERS/HERS-II <sup>279</sup>	0.625mg CEE + 2.5mg MPA	2,321	6.8 years	HR: 0.81; 95% CI: 0.46 to 1.45; p=0.48

CEE: conjugated equine estrogen; HERS: Heart and Estrogen/Progestin Replacement Trial; HR: hazard ratio; MPA: medroxyprogesterone acetate; WHI: Women’s Health Initiative.

**Table 71. Overall colorectal cancer incidence among women treated with estrogen alone**

Trial	Treatment	N	Average Followup	Results
WHI – CEE alone <sup>275</sup>	0.625mg CEE	10,739	6.8 years	Intervention: HR: 1.15; 95% CI: 0.81 to 1.64 Postintervention: HR: 1.01; 95% CI: 0.58 to 1.79 Overall: HR: 1.11.; 95% CI: 0.82. to 1.50

CEE: conjugated equine estrogen; HR: hazard ratio; WHI: Women’s Health Initiative.

## Applicability

Several large observational studies following younger populations also examined hormone therapy and colorectal cancer risk: the Breast Cancer Detection Demonstration Project (BCDDP),<sup>293</sup> the Nurses’ Health Study,<sup>294</sup> and the Molecular Epidemiology of Colon Cancer Study.<sup>295</sup>

The BCDDP reported that women treated with estrogen/progestin for 2 to 5 years had a relative risk for colorectal cancer of 0.52 (95 percent CI: 0.32 to 0.87), but results for women treated less than 2 years and women treated greater than 5 years were nonsignificant.<sup>293</sup> Women treated with estrogen alone for greater than 10 years had a relative risk of 0.69 (95% CI: 0.56 to 0.96), but women treated less than 10 years did not show a significant relationship (e.g., 5 to 9 years of use RR 0.74 [95 percent CI: 0.53 to 1.02]).<sup>293</sup> Current hormone users (75 percent of person-time was estrogen alone and 25 percent estrogen/progestin) in the Nurses’ Health Study had a colorectal cancer relative risk of 0.65 (95 percent CI: 0.50 to 0.83). This same relationship was not found in past users.<sup>294</sup> The Molecular Epidemiology of Colon Cancer Study reported an odds ratio for colon cancer among hormone users of 0.37 (95 percent CI: 0.22 to 0.62), adjusting for age, sex, aspirin use, statin use, sports activities, family history of colon cancer, ethnic group, and vegetable consumption level.<sup>295</sup>

Overall, the observational studies show either no effect, or a protective effect for subgroups of hormone users, which support the findings of the randomized trials. Two of the large studies combined estrogen/progestin and estrogen alone users into one broad category of hormone users in the analyses.

## Conclusions

The evidence for estrogen/progestin therapy and colorectal cancer risk consists of two large RCTs, the WHI trial<sup>272</sup> and HERS-II.<sup>279</sup> The quality of both trials was rated as fair. Results are inconsistent, with WHI reporting a protective effect during the treatment phase and no effect

during the post treatment phase. The overall analysis has a confidence interval touching 1.00. HERS-II reports no effect. The evidence is direct. The estimates are imprecise (WHI overall CI including 1.00 and HERS-II with wide CI). The strength of evidence was rated low that estrogen/progestin therapy does not affect the risk of colorectal cancer.

The evidence informing estrogen therapy and colorectal cancer risk consists of one large RCT, the WHI trial.<sup>275</sup> Trial quality was rated as fair. The results do not show a significant relationship between estrogen therapy and colorectal cancer risk. Consistency is unknown with only one trial, though intervention, postintervention, and overall measures, all show no effect. The measures are direct and precise. The strength of evidence is rated as moderate that estrogen therapy has no affect the risk of colorectal cancer.

## Coronary Heart Disease, Stroke, and Venous Thromboembolic Events

### Summary

Three trials examined the incidence of coronary heart disease, stroke or venous thromboembolic events: WHI estrogen/progestin,<sup>272</sup> WHI estrogen-only,<sup>275</sup> and ESPRIT.<sup>281</sup> Oral conjugated estrogen (CEE) was administered in the WHI trials and estradiol valerate (E2V) was administered in the ESPRIT trial.

The WHI trial found that neither hormone therapies increased mortality due to coronary heart disease or myocardial infarction. However, both therapies were associated with an increased incidence of stroke (Table 72 and Table 73). Using WHI data, Nelson et al. calculated that estrogen/progestin therapy resulted in nine more strokes per 10,000 woman-years (95 percent CI: 2 to 15), and estrogen-only therapy resulted in 11 more strokes per 10,000 woman-years (95 percent CI: 2 to 20). Deep vein thromboembolic (DVT) events were also increased with both estrogen/progestin and estrogen-only therapies. Estrogen/progestin resulted in 12 more DVT events per 10,000 woman-years (95 percent CI: 6 to 17) and estrogen-only therapy results in seven more DVT events per 10,000 woman-years (95 percent CI: 1 to 14).

**Table 72. Coronary heart disease, stroke, and venous thromboembolic events incidence among women treated with estrogen/progestin**

Trial	Overall CHD	All CVD Events	Total MI	Stroke	PE	DVT	CHD Death
WHI – CEE + MPA <sup>272</sup>	HR: 1.22 95% CI:0.99 to 1.51	HR: 1.13 95% CI:1.02 to 1.25	HR: 1.26 95% CI:1.00 to 1.59	HR: 1.34 95% CI:1.05 to 1.71	HR: 1.98 95% CI:1.36 to 2.87	HR: 1.88 95% CI:1.38 to 2.55	HR: 1.04 95% CI:0.67 to 1.64

CEE: conjugated equine estrogen; CHD: coronary heart disease; CVD: cardiovascular disease; DVT deep venous thrombosis; HR: hazard ratio; MPA: medroxyprogesterone acetate; PE: pulmonary embolism; WHI: Women’s Health Initiative.

**Table 73. Coronary heart disease, stroke, and venous thromboembolic events incidence among women treated with estrogen alone**

Trial	Overall CHD	All CVD Events	Total MI	Stroke	PE	DVT	CHD Death
WHI – CEE alone <sup>275</sup>	HR: 0.95 95% CI: 0.78 to 1.15	HR: 1.11 95% CI: 1.01 to 1.23	HR: 0.98 95% CI: 0.79 to 1.21	HR: 1.36 95% CI: 1.08 to 1.71	HR: 1.37 95% CI: 0.90 to 2.07	HR: 1.47 95% CI: 1.06 to 2.05	HR: 0.98 95% CI: 0.70 to 1.39
ESPRIT <sup>281</sup>				RR: 1.64 95% CI:	RR: 0.98 95% CI:	RR: 1.96 95%CI:	RR: 0.68 95% CI:

0.60 to 4.47; p=0.45	0.20 to 4.84 p=1.00	0.18 to 21.6 p=1.00	0.39 to 1.19; p=0.17
----------------------------	---------------------------	---------------------------	----------------------------

CEE: conjugated equine estrogen; CHD: coronary heart disease; CVD: cardiovascular disease; DVT: deep venous thrombosis; HR: hazard ratio; PE: pulmonary embolism; RR: rate ratio; WHI: Women's Health Initiative.

## Applicability

The WHI trial enrolled an older population and the ESPRIT trial enrolled women surviving a first myocardial infarction. A younger population was followed in the Danish Osteoporosis Prevention Study (DOPS),<sup>296</sup> published subsequent to the USPSTF literature search. DOPS is an open label randomized trial of recently menopausal healthy women aged 45 to 58 (mean age 50 years).

After ten years followup, women receiving estrogen/progestin had a hazard ratio of 0.57 (95 percent CI: 0.28 to 1.16) and women receiving estrogen had a hazard ratio of 0.32 (95 percent CI: 0.10 to 1.00) for mortality, heart failure or myocardial infarction. The total population had a hazard ratio of 2.01 (95 percent CI: 0.18 to 22.16) for DVT and a hazard ratio of 0.77 (95 percent CI: 0.35 to 1.70) for stroke after 10 years of followup. These nonsignificant findings contradict the significant negative effects of hormones on cardiovascular events found in WHI and ESPRIT, underscoring the need for further research in this area.

## Conclusions

The evidence for estrogen/progestin therapy and coronary heart disease consists of one large RCT, the WHI trial.<sup>272</sup> The trial did not find a significant relationship between treatment and overall coronary heart disease, myocardial infarctions, or death from coronary heart disease. Trial quality was rated as fair. With one trial, consistency is unknown, though the findings are supported by a trial published subsequent to the USPSTF literature search. The measures are imprecise (CI for total MI touches 1.00 and CHD death has a wide CI). The strength of evidence is rated low that estrogen/progestin has no effect on coronary heart disease.

The evidence for estrogen/progestin therapy and venous thromboembolic events consists of one large RCT, the WHI trial.<sup>272</sup> There were significant increases in all three measures: stroke, pulmonary embolism, and DVT. Trial quality was rated as fair. With one trial, consistency is unknown, though all three measures show increased risk. The strength of evidence is rated moderate that estrogen/progestin therapy increases the risk of stroke, pulmonary embolisms, and DVT.

The evidence concerning estrogen therapy and coronary heart disease consists of one large RCT, the WHI trial<sup>275</sup> and one small RCT, the ESPRIT trial.<sup>281</sup> The WHI trial reported total MI, CHD death, and overall CHD. The ESPRIT trial reported only CHD death. All four measures show no effect of estrogen therapy. Both trials were rated fair quality. Consistency is unknown for total MI and overall CHD because only one trial reported those measures. CHD death was consistent between the two trials. The strength of evidence is rated moderate that estrogen has no effect on coronary heart disease.

The evidence for estrogen therapy and venous thromboembolic events consists of one large RCT, the WHI trial<sup>275</sup> and one small RCT, the ESPRIT trial.<sup>281</sup> The WHI trial found significant increases in stroke and DVT. ESPRIT also found increases in stroke and DVT events, though the increases were not significant, possibly due to the small sample size. Both trials were rated fair quality. The strength of evidence is rated as high that estrogen therapy increases the risk of venous thromboembolic events.

## Endometrial Cancer

### Summary

Two trials (Table 74) reported the incidence of endometrial cancer: WHI estrogen/progestin<sup>272</sup> and HERS/HERS-II.<sup>279</sup> Both trials administered oral conjugated equine estrogen (CEE). Followup ranged from 5.2 years in WHI to 6.8 years in HERS/HERS-II. No significant differences in endometrial cancer incidence were observed in the trials of estrogen/progestin therapies. The increased risk of endometrial cancer when using estrogen-only therapies has already been established.<sup>76</sup>

**Table 74. Overall endometrial cancer incidence among women treated with estrogen/progestin**

Trial	Treatment	N	Average Followup	Results
WHI – CEE + MPA <sup>272</sup>	0.625mg CEE + 2.5mg MPA	15,730	5.2 years	Postintervention: HR: 0.75; 95% CI: 0.40 to 1.43 Overall: HR: 0.78; 95% CI: 0.52 to 1.16
HERS/HERS-II <sup>279</sup>	0.625mg CEE + 2.5mg MPA	2,485	6.8 years	HR: 0.25; 95% CI: 0.05 to 1.18; p=0.08

CEE: conjugated equine estrogen; HERS: Heart and Estrogen/Progestin Replacement Trial; HR: hazard ratio; MPA: medroxyprogesterone acetate.

### Applicability

Two large observational studies with younger populations, the Nurses' Health Study and the European Prospective Investigation into Cancer and Nutrition report an increased endometrial cancer risk among women treated with estrogen/progestin. The European study shows an overall increased risk of endometrial cancer, with a HR of 1.41 (95 percent CI: 1.08 to 1.83)<sup>297</sup> and the Nurses' Health Study reports a RR of 1.33 (95 percent CI: 1.01 to 1.75).<sup>298</sup> Additional research in this area is necessary.

### Conclusions

The evidence concerning estrogen/progestin therapy and endometrial cancer included two large RCTs, the WHI trial<sup>272</sup> and HERS/HERS-II.<sup>279</sup> Neither showed a significant relationship between the therapy and endometrial cancer. Both trials were rated as fair quality. While the results are consistent between these trials, they are inconsistent with results from two large observational studies. The measures are also imprecise, with large confidence intervals. The strength of evidence is rated as low that estrogen/progestin therapy has no effect on endometrial cancer incidence.

## Osteoporotic Fractures

### Summary

Three trials reported the incidence of osteoporotic fractures: WHI estrogen/progestin,<sup>49</sup> WHI estrogen-only,<sup>274</sup> and HERS/HERS-II.<sup>279</sup> Oral conjugated estrogen (CEE) was administered in all three trials. Followup ranged from 5.2 years in the WHI trial to 6.8 years in the HERS/HERS-II trial.

The HERS/HERS-II trial did not detect an effect on fracture incidence with estrogen/progestin therapy. In the WHI trials, both estrogen/progestin and estrogen alone were

associated with a decreased osteoporotic fracture incidence (Table 75 and Table 76). Based on the WHI estimates, estrogen/progestin therapy resulted in 46 fewer fractures per 10,000 woman-years (95 percent CI: 29 to 63), and estrogen-only therapy resulted in 56 fewer fractures per 10,000 woman-years (95 percent CI: 37 to 75). Decreased incidences of hip and vertebral fractures were observed for both therapies as well. Estrogen/progestin therapy resulted in 6 fewer hip fractures (95 percent CI: 1 to 10) and six fewer vertebral fractures (95 percent CI: 1 to 11). Estrogen-only therapy resulted in seven fewer hip fractures (95 percent CI: 1 to 12) and six fewer vertebral fractures (95 percent CI: 1 to 12).<sup>37</sup>

**Table 75. Osteoporotic fracture incidence among women treated with estrogen/progestin**

Trial	Total	Hip	Vertebral	Wrist	Other
WHI – CEE + MPA <sup>49</sup>	HR: 0.76 95% CI:0.69 to 0.85	HR: 0.66 95% CI:0.45 to 0.98	HR: 0.66 95% CI:0.44 to 0.98		HR: 0.77 95% CI:0.69 to 0.86
HERS/HERS-II <sup>279</sup>	HR: 1.04 95% CI:0.87 to 1.25	HR: 1.61 95% CI:0.98 to 2.66	HR: 0.87 95% CI:0.52 to 1.48	HR: 0.98 95% CI:0.64 to 1.50	HR: 0.94 95% CI:0.75 to 1.18

CEE: conjugated equine estrogen; HERS: Heart and Estrogen/Progestin Replacement Trial; HR: hazard ratio; MPA: medroxyprogesterone acetate; WHI: Women’s Health Initiative.

**Table 76. Osteoporotic fracture incidence among women treated with estrogen alone**

Trial	Total	Hip	Vertebral	Wrist	Other
WHI – CEE alone <sup>274</sup>	HR: 0.70 95% CI:0.63 to 0.79	HR: 0.61 95% CI:0.41 to 0.91	HR: 0.62 95% CI:0.42 to 0.93		

CEE: conjugated equine estrogen; HERS: Heart and Estrogen/Progestin Replacement Trial; HR: hazard ratio; WHI: Women’s Health Initiative.

## Applicability

The WHI and HERS trials have older but overlapping populations compared to the target population of this CER. Additional evidence for younger populations was not identified.

## Conclusions

The evidence concerning estrogen/progestin therapy and osteoporotic fractures consists of two large RCTs, the WHI trial<sup>49</sup> and HERS/HERS-II.<sup>279</sup> The WHI trial found significant decreases in hip, vertebral, other, and total fractures. The HERS trial did not find significant relationships, possibly due to a small sample size, as seen with the wide confidence intervals in the estimates. Both trials were rated as fair quality. While results were inconsistent, the measures were direct, and the WHI estimates were precise. Strength of evidence is rated as moderate that estrogen/progestin therapy decreases the incidence of osteoporotic fractures.

The evidence for estrogen therapy and osteoporotic fractures consists of the WHI trial.<sup>274</sup> The trial reported significant reductions in hip, vertebral, and total osteoporotic fractures. Trial quality was rated as fair. Consistency is unknown with one trial. The measures are direct and precise. Strength of evidence is rated as moderate that estrogen therapy reduces the risk of osteoporotic fractures.

## Ovarian Cancer

### Summary

One trial reported the incidence of ovarian cancer: WHI estrogen/progestin.<sup>273</sup> This trial administered oral conjugated estrogen (CEE) with the addition of medroxyprogesterone acetate. The hazard ratio shows an increased risk for ovarian cancer, though the wide confidence interval includes 1.00 (Table 77).

No RCTs in the Nelson report provided evidence for an association between estrogen alone and ovarian cancer.

**Table 77. Ovarian cancer incidence among women treated with estrogen/progestin**

Trial	Treatment	N	Average Followup	Results
WHI – CEE + MPA <sup>273</sup>	0.625mg CEE + 2.5mg MPA	16,608	5.6 years	HR: 1.58; 95% CI: 0.77 to 3.24

CEE: conjugated equine estrogen; HERS: Heart and Estrogen/Progestin Replacement Trial; HR: hazard ratio; MPA: medroxyprogesterone acetate; WHI: Women’s Health Initiative.

### Applicability

Two large observational studies with younger populations have reported on risks of ovarian cancer among women treated with estrogen/progestin: the European Prospective Investigation into Cancer and Nutrition and the Cancer Prevention Study II (CPS-II). Both studies report a nonsignificant relationship between estrogen/progestin use and ovarian cancer incidence. The European study reports an adjusted HR of 1.20 (95 percent CI: 0.89 to 1.62)<sup>299</sup> and CPS-II reports an adjusted RR for former estrogen/progestin users of 1.40 (95 percent CI: 0.86 to 2.28) and for current estrogen/progestin users of 1.18 (95 percent CI: 0.79 to 1.76).<sup>300</sup> Accordingly, the evidence reviewed was judged consistent with the WHI results.

### Conclusions

The evidence concerning estrogen/progestin therapy and ovarian cancer consists of one large RCT, the WHI trial.<sup>273</sup> The trial reports an increased risk of ovarian cancer, though the findings are not significant. Trial quality was rated as fair. Consistency is unknown with one trial, but results from two large observational studies also show increased, but nonsignificant findings. Measures were direct. Evidence is imprecise (wide CI) due to a low number of events. Strength of evidence is rated as low that estrogen/progestin therapy increases ovarian cancer risk.

## Strength of Evidence – Long-Term Effects of Hormone Therapy Preparations

**Table 78. Strength of evidence assessment for long-term effects of hormone therapies<sup>a</sup>**

Outcome	Risk <sup>b</sup>	Treatment vs. Placebo	Risk of Bias	Consistency	Directness	Precision	Reporting Bias	SOE
Breast Cancer	↑	Estrogen/Progestin	M	C	D	P	U	High
	↓	Estrogen	M	I	D	P	U	Low
Gall bladder	↑	Estrogen/Progestin	M	U	D	P	U	Mod

disease	↑	Estrogen	M	U	D	P	U	Mod
VTE	↑	Estrogen/Progestin	M	U	D	P	U	Mod
	↑	Estrogen	M	C	D	P	U	High
Stroke	↑	Estrogen/Progestin	M	U	D	P	U	Mod
	↑	Estrogen	M	C	D	P	U	High
Ovarian Cancer	↑	Estrogen/Progestin	M	U	D	I	U	Low
Colorectal Cancer	—	Estrogen/Progestin	M	I	D	I	U	Low
	—	Estrogen	M	U	D	P	U	Mod
CHD	—	Estrogen/Progestin	M	U	D	I	U	Low
	—	Estrogen	M	U	D	P	U	Mod
Endometrial Cancer	—	Estrogen/Progestin	M	I	D	I	U	Low
Osteoporotic Fractures	↓	Estrogen/Progestin	M	I	D	P	U	Mod
	↓	Estrogen	M	U	D	P	U	Mod

<sup>a</sup> Risk of Bias: High (H), Medium (M), Low (L); Consistency: Inconsistent (I), Unknown (U), Consistent (C); Directness: Indirect (I), Direct (D); Precision: Imprecise (I), Unknown (U), Precise (P); Reporting Bias: Suspected (S), Undetected (U).

<sup>b</sup> Risk: ↑ increased, ↓ decrease, — no change.

SOE: strength of evidence; Mod: moderate; CI: confidence interval; VTE: venous thromboembolic embolic events.

### Key Question 3. Nonhormone Other Benefits/Harms

This key question addresses the long-term effects of nonhormone therapies on the following conditions: breast cancer; gall bladder disease; colorectal cancer; coronary heart disease, stroke, thromboembolism; endometrial cancer; osteoporotic fractures; and ovarian cancer. Agent-specific adverse events were also assessed. Seven randomized controlled trials (Table 80), one cohort study (Table 81) and two case-control (Table 82) studies formed the evidence base (Tables 2 and 3 in Methods detail inclusion criteria). We excluded population-based dietary studies and studies reporting intermediate outcomes.

Evidence examining associations of nonhormone therapies with breast cancer, colorectal cancer, coronary heart disease, stroke, and thromboembolism, osteoporotic fractures, and ovarian cancer was identified<sup>301-311</sup> (Table 79). No evidence was identified evaluating gall bladder disease associations.

**Table 79. Evidence base for long-term effects of nonhormone therapies**

Condition	Antidepressants	Soy (Isoflavones)	Vitamin E
Breast cancer	Chien 2006 <sup>301</sup> Wernli 2009 <sup>303</sup>	No evidence identified	Lonn 2005 <sup>304</sup> Lee 2005 <sup>305</sup> Lin 2009 <sup>306</sup>
Gall bladder disease	No evidence identified	No evidence identified	No evidence identified
Colorectal cancer	No evidence identified	No evidence identified	Lee 2005 <sup>305</sup> Lin 2009 <sup>306</sup>
Coronary heart disease, stroke, thromboembolism	No evidence identified	No evidence identified	Lee 2005 <sup>305</sup> Cook 2007 <sup>311</sup>
Osteoporotic fractures	Spangler 2007 <sup>307</sup>	Maugeri 1994 <sup>308</sup> Passeri 1995 <sup>309</sup> Alexandersen 2001 <sup>310</sup>	No evidence identified
Ovarian cancer	No evidence identified	No evidence identified	Lin 2009 <sup>306</sup>

**Table 80. Study quality assessment for Key Question 3 RCTs**

Study	Comparable groups	Researcher/Subjects Blinded	Adequate Concealment	Comparable Groups Maintained	Differential Loss to Followup	Measures Equal, Reliable	Interventions Clear	Outcomes Defined	Intention to Treat Analysis	Study Quality
Alexandersen 2001 <sup>310</sup>	Y	Y	Y	U	Y	Y	Y	Y	Y	Fair
Cook 2007 <sup>311</sup>	Y	Y	Y	Y	N	Y	Y	Y	Y	Good
Lee 2005 <sup>305</sup>	Y	Y	U	Y	N	Y	Y	Y	Y	Good
Lin 2009 <sup>306</sup>	Y	Y	Y	Y	N	Y	Y	Y	Y	Good
Lonn 2005 <sup>304</sup>	Y	Y	Y	Y	N	Y	Y	Y	Y	Good
Maugeri 1994 <sup>308</sup>	U	Y	U	U	N	Y	Y	Y	U	Poor
Passeri 1995 <sup>309</sup>	U	U	U	U	Y	Y	Y	Y	N	Poor

N: no; U: unknown; Y: yes.

**Table 81. Study quality assessment for Key Question 3 cohort studies**

Study	Comparable Groups	Comparable Groups Maintained	Differential Loss to Followup	Measures Equal, Reliable	Outcomes Defined	Statistical Adjustment for Potential Confounders	Study Quality
Spangler 2008 <sup>307</sup>	Y	Y	U	Y	Y	Y	Fair (observational)

N: no; U: unknown; Y: yes.

**Table 82. Study quality assessment for Key Question 3 case control studies**

Study	Accurate case Ascertainment	Nonbiased Selection of Cases/Controls	Response rate	Diagnostic Tests Equal	Exposure Accurately Measured	Exposure Applied Equally	Statistical Adjustment for Confounders	Study Quality
Chien 2006 <sup>301</sup>	Y	Y	cases 81%; controls 74%	Y	Y <sup>a</sup>	Y	Y	Poor (observational)
Wernli 2009 <sup>303</sup>	Y	Y	cases 74%; controls 67%	Y	Y <sup>a</sup>	Y	Y	Poor (observational)

<sup>a</sup> A-depressant use determined through self-report, in structured interviews

## Breast Cancer

### Summary

Many studies evaluating soy and breast cancer incidence and herbal preparations and breast cancer incidence were identified, but were subsequently excluded because the studies were either population based dietary studies or reported only intermediate outcomes (Appendix B). Three studies on vitamin E and breast cancer incidence<sup>304-306</sup> and two studies on antidepressants and breast cancer incidence<sup>301, 303</sup> were identified and met inclusion criteria.

The Health Outcomes Prevention Evaluation (HOPE) and its extension, HOPE—The Ongoing Outcomes trial (HOPE-TOO), examined vitamin E (400 IU daily) and breast cancer incidence.<sup>304</sup> The trial population enrolled women with vascular disease or diabetes (N=9541 in HOPE, with N=7030 continuing in HOPE-TOO). Followup in HOPE was 6 years, with an additional 4 years in HOPE-TOO. A second RCT,<sup>305</sup> the Women’s Health Study (WHS), enrolled healthy women aged 45 years or older with a 10-year average followup. Participants in the treatment group took 600 IU of vitamin E every other day. A third randomized controlled trial, the Women’s Antioxidant Cardiovascular Study, administered 600 IU of vitamin E every other day to women at high risk for cardiovascular disease. Followup averaged 9.4 years.<sup>306</sup>

A number of studies have investigated a possible antidepressant-breast cancer association. We excluded those enrolling women of all ages because of difficulty assessing modification by age on any exposure-disease association. We also excluded studies that reported results for antidepressants as a whole, without clarifying if the antidepressants were SSRI/SNRIs. Two case-control studies met inclusion criteria. Chien et al<sup>301</sup> enrolled women aged 65 to 79 diagnosed with invasive breast cancer. Information on history of antidepressant use in the 20 years prior to the cancer diagnosis was collected, and results were reported for all antidepressants and for subgroups of antidepressants: tricyclics (TCA), SSRIs, and triazolopyridines. Wernli et al<sup>303</sup> investigated women 20 to 69 years of age, but subgroup analyses for women aged 50 years or older were provided. Results were reported for all antidepressants combined, as well as for specific types of antidepressants (SSRI, TCA, and norepinephrine-dopamine reuptake inhibitors).

**Table 83. Nonhormone therapies and breast cancer**

Condition	Treatment	Source; Evidence Type	Study Description	Comparators	Results
Breast cancer	Vitamin E	Lonn 2005 <sup>304</sup> , RCT	HOPE conducted 1993-1999 (N=9541) <sup>a</sup>	Placebo: Vitamin E:	0.6% (cumulative incidence) 0.5% RR: 0.86; 95% CI: 0.50 to 1.47; p=0.58
			HOPE-TOO conducted 1999-2003 (N=7030) <sup>a</sup>	Placebo: Vitamin E:	0.7% (cumulative incidence) 0.5% RR: 0.73; 95% CI: 0.40 to 1.31; p=0.29
		Lee 2005 <sup>305</sup> , RCT	WHS conducted 1994-2004 (N=39,876)	Placebo: Vitamin E:	3.1% (cumulative incidence) 3.1% RR: 1.00; 95% CI: 0.90 to 1.12; p=0.95
		Lin 2009 <sup>306</sup> , RCT	Women's Antioxidant Cardiovascular Study 1995-2005 (N=8171) <sup>b</sup>	Placebo: Vitamin E:	130 cases 127 cases RR: 0.98; 95% CI: 0.77 to 1.25
Breast cancer	Anti-depressants	Chien 2006 <sup>301</sup> , case-control	Women aged 65 to 79 years	Never used SSRI:	914 cases; 953 controls OR: 1.0
			Cases (n= 975) Controls (n= 1007)	Ever used SSRI:	61 cases; 54 controls OR: 1.2; 95% CI: 0.8 to 1.8
Breast cancer	Anti-depressants	Wernli 2009 <sup>303</sup> , case-control	Women aged 20 to 69, newly diagnosed breast cancer	<50 years (cases=952; controls=900)	12.6% cases ever use SSRI 14.2% controls ever use SSRI OR: 0.79; 95% CI: 0.60 to 1.04
			Cases (n= 2908) Controls (n= 2927)	≥ 50 years (cases=1956; controls=2027)	10.4% cases ever use SSRI 11.3% controls ever use SSRI OR: 0.88; 95% CI: 0.72 to 1.08

CI: confidence interval; HOPE: Health Outcomes Prevention Evaluation trial; HOPE-TOO: Health Outcomes Prevention Evaluation- The Ongoing Outcomes trial; N=number; OR: odds ratio; RCT: randomized controlled trial; RR: relative risk.

<sup>a</sup> Participants were at least 55 years of age with vascular disease or diabetes.

<sup>b</sup> Participants were women 40 years of age and older, at high risk for cardiovascular disease.

## Conclusions

The evidence for vitamin E and breast cancer risk consists of three large randomized controlled trials.<sup>304-306</sup> The population of one trial<sup>305</sup> was healthy women over 45 years of age and the other two trials focused on women with vascular disease or diabetes.<sup>304, 306</sup> Participants received vitamin E supplements or placebo. The trials—with followups of up to 10 years and sample sizes of 7,030,<sup>304</sup> 39,876,<sup>305</sup> and 8,171<sup>306</sup>—found no statistically significant benefit of vitamin E in the prevention of breast cancer. All trials were rated as good quality (Table 80). The results are consistent among all three trials. The measures are direct and the narrow confidence interval around the null in the larger trial<sup>305</sup> indicates precision. The strength of evidence is rated as high that vitamin E has no effect on breast cancer risk in this population of women.

The evidence for antidepressant use and breast cancer risk consists of two case-control studies.<sup>301, 303</sup> The two observational studies are poor quality (Table 82). Results are consistent and direct. One study has a small sample size and imprecise measures. The strength of evidence was rated low that SSRI/SNRIs have no effect on breast cancer risk.

## Gall Bladder Disease

No studies evaluating associations between nonhormone therapies used for menopausal symptom relief and gall bladder disease were identified.

## Colorectal Cancer

### Summary

Two included studies (Table 84) evaluated vitamin E and colorectal cancer. Dietary studies of soy and colorectal cancer incidence and one study reporting results in men and women combined were identified but excluded.

One large RCT,<sup>305</sup> the Women's Health Study (WHS), investigated the use of 600 IU of vitamin E taken every other day. The trial population was healthy women aged 45 years or older and length of followup was an average of 10 years. Colorectal cancer incidence was reported. The second trial, the Women's Antioxidant Cardiovascular Study, also administered 600 IU of vitamin E every other day. The trial enrolled women aged 40 years or older with cardiovascular disease risk factors. Average followup was 9.4 years and colorectal cancer incidence was reported.<sup>306</sup>

**Table 84. Nonhormone therapies and colorectal cancer**

Condition	Treatment	Source; Evidence Type	Trial Description	Comparators	Results
Colorectal cancer	Vitamin E	Lee 2005 <sup>305</sup> , RCT	WHS conducted 1994-2004 (N=39,876)	Placebo: Vitamin E:	0.5% (cumulative incidence) 0.5%
		Lin 2009 <sup>306</sup> , RCT	Women's Antioxidant Cardiovascular Study 1995-2005 <sup>a</sup> (N=8171) <sup>a</sup>	Placebo: Vitamin E:	RR: 1.00; 95% CI: 0.77-1.31; p=0.99 27 cases 17 cases RR: 0.63; 95% CI: 0.34 to 1.15

<sup>a</sup> Participants were women 40 year and older, at high risk for cardiovascular disease.

### Conclusions

Two large RCTs examined the effect of vitamin E on colorectal cancer. One trial, with a sample size of 39,876 and a followup of ten years, found no statistically significant benefit of vitamin E in the prevention of colon cancer (RR=1.00).<sup>305</sup> The second trial with a sample size of 8171 and a followup of 9.4 years, reports a protective effect (RR=0.63), but the estimate was not statistically significant (95 percent CI: 0.34 to 1.15).<sup>306</sup> The trials were rated as good quality (Table 80). The estimates were consistent and direct. The measure for the large study was precise, though the smaller study had a larger confidence interval. The strength of evidence is rated as high that vitamin E has no effect on colon cancer incidence.

# Coronary Heart Disease, Stroke, or Thromboembolism

## Summary

The literature examining the potential effect of soy on the prevention of cardiovascular disease is large, but is limited to population based dietary studies or those reporting intermediate outcomes. Consequently, the studies were excluded. Two studies were identified that met inclusion criteria.

The Women’s Health Study,<sup>305</sup> examined vitamin E supplementation and cardiovascular disease among healthy women, aged 45 years or older. The average length of followup was 10 years. Outcomes included overall cardiovascular events, myocardial infarction, stroke, and cardiovascular death. In the Women’s Antioxidant Cardiovascular Study 600 IU vitamin E was prescribed every other day to women over age 40 at increased risk for cardiovascular disease.<sup>311</sup> Average followup was 9.4 years and outcomes included myocardial infarction, stroke, and cardiovascular death.

**Table 85. Nonhormone therapies and CHD, stroke, or thromboembolism**

Condition	Treatment	Source; Evidence Type	Trial Description	Results
CHD, stroke, or thromboembolism	Vitamin E	Lee 2005 <sup>305</sup> , RCT	WHS conducted 1994-2004 (N=39,876)	CV events: RR: 0.93; 95% CI: 0.82 to 1.05; p = 0.26
				MI: RR: 1.01; 95% CI: 0.82 to 1.23; p = 0.96
				Stroke: RR: 0.98; 95% CI: 0.82 to 1.17; p = 0.82
		CV death: RR: 0.76; 95% CI: 0.59 to 0.98; p = 0.03		
		Cook 2007 <sup>311</sup> , RCT	Women’s Antioxidant Cardiovascular Study 1995-2005 (N=8171) <sup>a</sup>	CV events: RR: 0.94; 95% CI: 0.85 to 1.04; p=0.23
				MI: RR: 0.91; 95% CI: 0.72 to 1.15; p=0.44
Stroke: RR: 0.84; 95% CI: 0.67 to 1.05; p=0.12				
CV death: RR: 0.94; 95% CI: 0.77 to 1.15; p=0.56				

CHD: coronary heart disease; CI: confidence interval; CV: cardiovascular; MI: myocardial infarction; N=number; RCT: randomized controlled trial; RR: relative risk.

<sup>a</sup> Participants were women 40 year and older, at high risk for cardiovascular disease.

## Conclusions

The evidence consists of two trials comparing vitamin E with placebo and the risk for cardiovascular events, myocardial infarction, stroke, and cardiovascular death. The samples were large with mean followups of 9.4 and 10 years. Neither trial found a statistically significant benefit of vitamin E in the prevention of overall cardiovascular events, including MI and stroke.<sup>305, 311</sup> The WHS report found a significant protective effect on cardiovascular death,<sup>305</sup> but the Women’s Antioxidant Cardiovascular Study did not.<sup>311</sup>

Both trials were rated good quality (Table 80). Consistent results were reported for cardiovascular events overall, as well as for myocardial infarction and stroke when analyzed separately. The measures are direct and precise. The strength of evidence is rated as high that vitamin E has no effect on overall cardiovascular events, including myocardial infarction and stroke.

The WHS trial reported a statistically significant benefit of vitamin E in the prevention of cardiovascular death<sup>305</sup> whereas the Women’s Antioxidant Cardiovascular Study did not.<sup>311</sup>

There are uncertainties with the WHS result because it is inconsistent not only with the other trial, but with the WHS results which showed no difference in number of overall cardiovascular events. Additionally, there are well-described inaccuracies in the ascertainment of cardiovascular deaths, as coded in death certificates.<sup>312</sup> Though the trial is of good quality, this outcome may have inaccuracies and may be potentially biased. The strength of evidence that vitamin E is protective against cardiovascular death is rated as low.

## **Endometrial Cancer**

### **Summary**

No studies meeting inclusion criteria evaluating the effect of nonhormone agents on endometrial cancer were identified. However, we briefly note a report from a working group of 22 clinical and research experts in the field of women's health and botanicals convened by the North American Menopause Society.<sup>313</sup> The group evaluated current evidence on health effects of isoflavones in peri- and postmenopausal women, including both menopausal symptom relief and long-term benefits and harms. There was no description provided on how articles were chosen for inclusion in the report. The publication discusses several large population based studies on soy consumption and the risk of endometrial cancer, which are not applicable for this current review.<sup>314-316</sup> The Society paper also reviewed several RCTs on soy treatment and endometrial hyperplasia, which is an intermediate outcome.

### **Conclusions**

There is insufficient evidence that treatment with soy products has an effect on the risk of endometrial cancer in menopausal women.

## **Osteoporotic Fractures**

### **Summary**

We identified three trials evaluating the effect of soy on osteoporotic fractures<sup>308-310</sup> (which were incorporated in a meta-analysis<sup>317</sup>) and one observational study of the association between antidepressants and osteoporotic fractures.<sup>307</sup>

Spangler et al. (2008) analyzed data from participants of the Women's Health Initiative Observational Study, focusing on depressive symptoms, antidepressant use, and bone fractures.<sup>307</sup> After controlling for depressive symptoms, as well as demographic, lifestyle, and reproductive factors, the investigators found SSRI use associated with an increased risk of fractures at any site. Analysis by fracture site found antidepressant users with increased fracture risk in spine and other sites.

Bolaños et al. (2010) performed an indirect treatment comparison, comparing a meta-analysis of three isoflavone versus placebo trials with a meta-analysis of ten hormone replacement therapy versus placebo trials, for the reduction of vertebral fractures. A search through the trials register of Cochrane Osteoporosis Treatment Trial Group, Cochrane Controlled Trials, MEDLINE®, EMBASE®, ProQuest, BIREME, Trip Database, LILACS, and Scielo through September 2009 was conducted. The Jadad scale<sup>318</sup> was used to assess the quality of the RCTs. The three isoflavone trials compared ipriflavone, at a dosage of 600 mg/day plus a calcium supplement versus a calcium supplement alone. The pooled estimate for isoflavones versus placebo in the reduction of vertebral fractures did not show a significant reduction. The authors

concluded that isoflavone therapy was “similar” to hormone therapy for preventing vertebral fracture using a simple calculation of the indirect odds ratios, but did not apply methods necessary to appropriately obtain estimated indirect effects and assess consistency.<sup>319</sup> Because the appropriate statistical methods were not used, the meta-analysis is not included in our evidence table. The three RCTs<sup>308-310</sup> are included in our assessment (Table 86).

**Table 86. Nonhormone therapies and osteoporotic fractures**

Condition	Treatment	Source; Evidence Type	Trial Description	Results
				HR (95% CI): <sup>a</sup>
	Antidepressants	Spangler 2007 <sup>307</sup> , prospective cohort trial	WHI-OS SSRI users (N=7212) vs. nonantidepressant users (N=86,463) average 7.4 year followup	all sites: 1.30 (1.20 to 1.41) hip: 1.33 (0.95 to 1.86) spine: 1.25 (0.96 to 1.63) wrist: 1.29 (1.07 to 1.56) other: 1.32 (1.21 to 1.45)
Osteoporotic fractures		Maugeri 1994 <sup>308</sup> , RCT	N=84 600 mg/day ipriflavone (N=41) or placebo (N=43) ≥ 65 years old 2 year followup	Fracture incidence: Ipriflavone: 2 (4.9%) Placebo: 11 (25.6%)
	Soy (isoflavones, phytoestrogens, lignans)	Passeri 1995 <sup>309</sup> , RCT	N=40 600 mg/day ipriflavone (N=20) or placebo (N=20) 65-79 years of age 2 year followup	Fracture incidence: Ipriflavone: 4 (20.0%) Placebo: 9 (45.0%)
		Alexandersen 2001 <sup>310</sup> , RCT	N=474 600 mg/day ipriflavone (N=234) or placebo (N=240) 45-75 years of age 3 year followup	Fracture incidence: Ipriflavone: 11 (4.7%) Placebo: 11 (4.6%)  RR: 1.07 (95% CI: 0.53 to 2.16)

CI: confidence interval; HR: hazard ratio; HRT: hormone replacement therapy; N: number; RCT: OR: odds ratio; randomized controlled trial; RR: relative risk; WHI-OS: Women’s Health Initiative Observational Trial.

<sup>a</sup> Adjusted for depressive symptoms.

## Conclusions

The evidence for antidepressant use and osteoporotic fractures consists of one large prospective cohort study (n=93,675) with 7212 antidepressant users followed for a mean of 7.4 years.<sup>307</sup> Hazard ratios showed increased risk for fractures in all sites, but the risks were only significant for wrist, other, and all sites. The study is rated fair. Consistency is unknown with only one study. The measures were direct and precise. The strength of evidence that SSRIs increase the incidence of osteoporotic fractures is rated as low.

The evidence for soy supplements and their effect on osteoporotic fractures consists of three trials. Two trials enrolled samples fewer than 100 participants who were followed for two years,<sup>308, 309</sup> and one trial of 474 women had a followup of three years.<sup>310</sup> One trial is rated fair quality and two trials are rated poor quality. The results were inconsistent, with the larger trial reporting no effect and the two smaller trials showing a potential protective effect of isoflavones. The measures were direct, but imprecise due to the small sample sizes. The strength of evidence is rated as insufficient that isoflavones protect against osteoporotic fractures.

## Ovarian Cancer

### Summary

One trial (Table 87) examining the effect of vitamin E on ovarian cancer was identified.<sup>306</sup> The Women's Antioxidant Cardiovascular Study, a double blind placebo-controlled trial, administered 600 IU of vitamin E every other day to women aged 40 years or older and at risk for cardiovascular disease. The study found that vitamin E has no effect on ovarian cancer incidence.

**Table 87. Nonhormone therapies and ovarian cancer**

Condition	Treatment	Source; Evidence Type	Study Description	Results
Ovarian cancer	Vitamin E	Lin 2009 <sup>306</sup> ; RCT	Women's Antioxidant Cardiovascular Study 1995-2005 (N=8171) <sup>a</sup>	Placebo: 14 cases Vitamin E: 8 cases RR: 0.58; 95% CI: 0.24 to 1.37

### Conclusions

The evidence for vitamin E and ovarian cancer consists of one randomized controlled trial. The single trial, with a sample size of 8171, reports a protective, though insignificant, effect.<sup>306</sup> The trial is rated good quality. Consistency is unknown with one trial. The measure is direct, but imprecise due to the small number of cases resulting in a wide confidence interval. The strength of evidence was rated low that vitamin E has no effect on ovarian cancer incidence.

### Strength of Evidence – Nonhormone Other Benefits/Harms

**Table 88. Strength of evidence assessment for long-term effects of nonhormone therapies<sup>a</sup>**

Outcome	Risk <sup>b</sup>	Treatment (vs. Placebo)	Risk of Bias	Consistency	Directness	Precision	Reporting Bias	SOE	Rationale for Downgrading
Breast Cancer	—	Vitamin E	L	C	D	P	U	High	
Breast Cancer	—	SSRI/SNRI	H	C	D	I	U	Low	2 poor quality case control studies; 1 imprecise
Colorectal Cancer	—	Vitamin E	L	C	D	P	U	High	
Cardiovascular Events	—	Vitamin E	L	C	D	P	U	High	
Cardiovascular Death	↓	Vitamin E	L	I	U	I	U	Low	Inconsistent - 2 trials with different results and small magnitude; uncertain directness as no effect on cardiovascular events; imprecise given CIs for effect magnitude
Osteoporotic Fractures	↑	SSRI	H	U	D	P	U	Low	Single observational study

Osteoporotic Fractures	↓	Isoflavones	H	I	D	I	U	Insuff	1 fair and 2 poor quality trials; small sample sizes; directionality of risks differed
Ovarian Cancer	—	Vitamin E	L	U	D	I	U	Low	1 large trial; wide CI
Gallbladder Disease									No evidence identified
Endometrial Cancer									No evidence identified

<sup>a</sup> Risk of Bias: High (H), Medium (M), Low (L); Consistency: Inconsistent (I), Unknown (U), Consistent (C); Directness: Indirect (I), Direct (D); Precision: Imprecise (I), Unknown (U), Precise (P); Reporting Bias: Suspected (S), Undetected (U).

<sup>b</sup> Risk: ↑ increased, ↓ decrease, — no change.

SOE: strength of evidence; Mod: moderate; Insuff: insufficient; CI: confidence interval;

## Adverse Events

### Summary

Twelve trials met inclusion criteria that reported adverse events for nonhormone prescription therapies used to treat menopausal symptoms. Six trials reported adverse events for desvenlafaxine,<sup>159, 207, 210, 320-322</sup> three reported events for gabapentin,<sup>211, 323</sup> two reported events for escitalopram,<sup>159, 324</sup> and one reported events for clonidine.<sup>109</sup> (Appendix Table M-1a and Table M-1b) The most common adverse events reported were in the following categories: nervous system (12 of 12 trials), gastrointestinal (11 of 12 trials), general disorders and administration site conditions (10 of 12 trials), and eye (6 of 12 trials). The highest incidence of reported events was from a trial with desvenlafaxine (47.8% gastrointestinal)<sup>322</sup> and from a trial with clonidine (52.4% nervous system).<sup>109</sup> (Appendix Table M-1a and Table E-1b)

Sixteen trials met inclusion criteria that reported adverse events for nonprescription therapies used to treat menopausal symptoms. Nine trials reported adverse events with the use of soy treatments,<sup>123, 195, 196, 198, 201, 203, 325-327</sup> three with black cohosh,<sup>134, 152, 328</sup> three with plants or multibotanicals,<sup>119, 139, 328</sup> one with St. John's wort,<sup>134</sup> and one with DHEA.<sup>155</sup> (Appendix Table M-2a and Table M-2b) One trial reported adverse events for both a nonhormone prescription therapy (fluoxetine) and a nonprescription therapy (black cohosh) and this trial's results were added to Appendix Tables M-1a and M-1b. The most common adverse events reported were in the following categories: gastrointestinal (15 of 16 trials), nervous system (11 of 16 trials), musculoskeletal (10 of 16 trials), reproductive system/breast (10 of 16 trials), and general disorders and administration site conditions (8 of 16 trials). The highest reported events were from a trial for soy (52.5 percent gastrointestinal)<sup>326</sup> and (25.4% reproductive system/breast).<sup>326</sup>

## Key Question 4. Effectiveness of Treatments for Menopausal Symptoms in Selected Subgroups

This question addresses the effectiveness of therapies for menopausal symptoms among subgroups of patients. The evidence base for this question consisted of the randomized controlled trials from Key Question 1 which also included subgroup analyses. Subgroups of interest included age, BMI, race, severity of menopausal symptoms, time since menopause, uterine status, and comorbidities.

There were a total of 16 trials which reported analyses on subgroups of interest.<sup>125, 145, 152, 175, 185, 192, 196, 206, 238, 242, 324, 329-333</sup> Results of the subgroup analyses are presented by category of outcome: vasomotor symptoms, sexual function, psychological symptoms, quality of life, sleep

dysfunction, and urogenital symptoms. For each outcome category, there is an evidence base table listing the trials by subgroup and type of treatment, study quality assessments, and summaries. Results tables are in Appendix N. Strength of evidence was not assessed due to the variety of treatments, outcome measures, and subgroup definitions.

## Vasomotor symptoms

**Table 89. Evidence base for subgroup analyses reporting vasomotor outcomes**

Subgroup	Hormone Therapies	Nonhormone Prescription Therapies	Nonprescription Therapies
Age	Rigano 2001 <sup>332</sup> Hedrick 2010 <sup>330</sup>		Davis 2001 <sup>125</sup>
BMI			Davis 2001 <sup>125</sup> Tice 2003 <sup>196</sup>
Race		Freeman 2011 <sup>324</sup>	
Severity of symptoms	Mattsson 2007 <sup>334</sup> Diem 2006 <sup>145</sup> Maki 2007 <sup>331</sup>		Aso 2012 <sup>329</sup>
Time since menopause	Baerug 1998 <sup>192</sup> Simon 2001 <sup>333</sup>		Davis 2001 <sup>125</sup> Osmers 2005 <sup>152</sup>
Uterine status	Hedrick 2010 <sup>330</sup>		

**Table 90. Quality assessment for studies reporting vasomotor subgroup analyses**

Study	Comparable Groups	Researcher/Subjects Blinded	Adequate Concealment	Comparable Groups Maintained	Differential Loss To Followup	Measures Equal, Reliable	Interventions Clear	Outcomes Defined	Intention To Treat Analysis	Study Quality
Aso 2012 <sup>329</sup>	Y	Y	U	U	Y	Y	Y	Y	N	Poor
Baerug 1998 <sup>192</sup>	U	Y	U	U	N	Y	Y	Y	U	Poor
Davis 2001 <sup>125</sup>	Y	Y	U	Y	Y	Y	Y	Y	U	Poor
Diem 2006 <sup>145</sup>	U	Y	Y	U	U	Y	Y	Y	Y	Poor
Freeman 2011 <sup>324</sup>	Y	Y	Y	Y	N	Y	Y	Y	Y	Good
Hedrick 2010 <sup>330</sup>	Y	Y	U	Y	N	Y	Y	Y	Y	Good
Maki 2007 <sup>331</sup>	U	Y	U	U	N	Y	Y	Y	Y	Poor
Osmers 2005 <sup>152</sup>	U	Y	Y	Y	N	Y	Y	Y	Y	Poor
Rigano 2001 <sup>332</sup>	N	U	U	U	U	U	U	Y	U	Poor
Simon 2001 <sup>333</sup>	U	Y	N	Y	N	Y	Y	Y	Y	Fair
Tice 2003 <sup>196</sup>	Y	Y	Y	Y	N	Y	Y	Y	Y	Good

## Vasomotor symptoms by Age

Three trials conducted subgroup analyses on vasomotor symptoms by age (Appendix N-Table 1).<sup>125, 330, 332</sup>

Rigano et al. compared a standard dose estrogen patch with placebo. All three age groups (48 to 50, 51 to 53, and 54 to 56) experienced significant improvements in percent without hot flushes at 26 weeks of followup.<sup>332</sup>

In a trial comparing three doses of estrogen skin gel (all low dose) with placebo, Hedrick et al reported number of moderate to severe hot flushes and night sweats among women <50, 50 to

59, and greater than or equal to 60. The authors report significant improvements in outcomes were only seen with the two older age groups, 50-59 and >60.<sup>330</sup>

In a trial comparing Chinese medicinal herbs with placebo, Davis et al. presented vasomotor results for women less than 55 and women greater than or equal to 55 years of age. The only group to see significant improvement in the MENQOL vasomotor score was the less than 55 age group.<sup>125</sup>

### **Vasomotor symptoms by BMI**

Two trials conducted subgroup analyses on vasomotor symptoms by BMI. The interventions were nonprescription, one using isoflavones<sup>196</sup> and one using Chinese medicinal herbs (Appendix N-Table 2).<sup>125</sup>

In the isoflavone trial, total number of hot flushes were reported for women with a BMI less than 25 and a BMI greater than or equal to 25. Both the placebo and the isoflavone groups reported decreased weekly total hot flushes.<sup>196</sup>

The trial using Chinese medicinal herbs reported MENQOL vasomotor scores and reduction in total number of hot flushes and night sweats. No clinically significant differences in vasomotor measures were detected in either BMI subgroup between the Chinese medicinal herbs and placebo group.<sup>125</sup>

### **Vasomotor Symptoms by Race**

One trial conducted a subgroup analysis on vasomotor symptoms by race (African American, white, and other) (Appendix N-Table 3).<sup>324</sup> The trial compared a selective serotonin reuptake inhibitor (escitalopram, 10 to 20 mg) with placebo. While a significant decrease in total hot flushes and night sweats was reported among whites and others in the SSRI group compared with placebo, African American women did not experience a significant decrease.

### **Vasomotor Symptoms by Severity of Symptoms**

Four trials conducted subgroup analyses on vasomotor symptoms by severity of symptoms (Appendix N-Table 4).<sup>145, 329, 331, 334</sup> Two trials compared estrogen and progestin combined treatments;<sup>331, 334</sup> one trial compared estrogen alone treatment;<sup>145</sup> and one trial compared a nonprescription treatment (isoflavones).<sup>329</sup>

In one trial with a standard dose estrogen/progestin therapy, Maki et al present mean change in total hot flushes for women asymptomatic at baseline compared with women symptomatic at baseline. A clinically significant decrease in vasomotor symptoms was observed only in the subgroup that was symptomatic at baseline.<sup>331</sup>

In a three arm trial comparing two standard doses of estrogen/progestin with one high dose of estrogen/progestin, Mattsson et al present weekly moderate to severe hot flushes for women with greater than or equal to 30 at baseline compared to women with less than 30 at baseline. All subgroups showed significant improvements with the estrogen/progestin therapies compared with placebo.<sup>334</sup>

In the Diem et al. (2006) trial,<sup>145</sup> subgroup analysis was conducted on participants with very frequent or somewhat frequent vasomotor symptoms. Treatment with ultra-low-dose estrogen resulted in both subgroups experiencing fewer weekly total hot flushes compared with placebo, though the differences were not statistically significant. Significance tests were not performed between the two subgroups.

The isoflavone trial<sup>329</sup> performed separate analyses on participants that had less than or equal to two hot flushes/day and participants that had greater than or equal to three hot flushes per day. Both subgroups experienced a decrease in hot flushes per day, but the difference was only significant in the subgroup with greater hot flushes per day. Significant tests were not performed between the two groups.

## Vasomotor Symptoms by Time Since Menopause

Four trials conducted subgroup analyses on vasomotor symptoms by time since menopause (Appendix N-Table 5).<sup>125, 152, 192, 333</sup> One trial compared estrogen/progestin with placebo;<sup>192</sup> one trial compared estrogen with placebo;<sup>333</sup> and two trials compared nonprescription treatments (Chinese medicinal herbs and black cohosh) with placebo.<sup>125, 152</sup>

In the Baerug et al. (1998) trial,<sup>192</sup> low dose estrogen/progestin was compared with placebo among late perimenopausal and postmenopausal women. Outcomes were mean weekly hot flushes. Mean weekly hot flushes were lower in both groups following treatment, though symptoms improved more in the late perimenopausal subgroup throughout the trial. The differences between the groups were not statistically significant.

Simon et al. (2001)<sup>333</sup> compared standard dose estrogen with placebo in four patient subgroups (0 to less than or equal to 6 months since last menses; 6 to less than or equal to 12 months since last menses; 12 to less than or equal to 36 months since last menses; and greater than 36 months since last menses). A decrease in mean percent reduction of daily moderate-to-severe hot flushes was observed in all subgroups, though the difference was significant in only two of the groups: 12 to less than or equal to 36 months since last menses and greater than 36 months since last menses.

Davis et al. (2001)<sup>125</sup> compared Chinese medicinal herbs with placebo in two patient subgroups: women experiencing less than 4 years of amenorrhea and women experiencing greater than or equal to 4 years of amenorrhea. MENQOL vasomotor score and total daily hot flushes and night sweats were the outcomes reported. There were no significant differences in vasomotor outcomes between the two subgroups.

Osmers et al. (2005)<sup>152</sup> compared black cohosh with placebo among early and late climacteric women. The difference in changes from placebo on the Menopause Rating Scale for hot flushes was significant in both early ( $p < 0.002$ ) and late ( $p < 0.006$ ) climacteric women.

## Vasomotor Symptoms by Uterus Status

One trial conducted subgroup analyses by uterus status (absent uterus vs. intact uterus) and reported vasomotor outcomes (Appendix N-Table 6). Three dose regimens (all low dose: 0.25 mg, 0.50 mg, and 1.0 mg) of estrogen skin gel were compared with placebo.<sup>330</sup> Among women with absent uteri, number of moderate to severe hot flushes decreased significantly in women treated with 0.25 mg and 1.0 mg estrogen gel, and severity of flushes decreased significantly only in the 1.0 mg estrogen gel group. Among women with intact uteri, number of moderate to severe hot flushes decreased significantly in the 0.50 mg and 1.0 mg treatment groups, and severity of hot flushes decreased significantly in all treatment groups.<sup>330</sup>

## Sexual Function

**Table 91. Evidence base for subgroup analyses reporting sexual function outcomes**

Subgroup	Nonhormone Prescription Therapies	
	Hormone Therapies	Nonprescription Therapies
Age	Rigano 2001 <sup>332</sup>	Davis 2001 <sup>125</sup>
BMI		Davis 2001 <sup>125</sup>
Time since menopause		Davis 2001 <sup>125</sup>
Uterine status	Davis 2008 <sup>238</sup>	

**Table 92. Quality assessment for studies reporting sexual function subgroup analyses**

Study	Comparable Groups	Researcher/Subjects Blinded	Adequate Concealment	Comparable Groups Maintained	Differential Loss To Followup	Measures Equal, Reliable	Interventions Clear	Outcomes Defined	Intention To Treat Analysis	Study Quality
Davis 2001 <sup>125</sup>	Y	Y	U	Y	Y	Y	Y	Y	U	Poor
Rigano 2001 <sup>332</sup>	N	U	U	U	U	U	U	Y	U	Poor

### Sexual Function by Age

Two trials conducted subgroup analyses on sexual function by age (Appendix N-Table 7).<sup>125, 332</sup> One trial compared estrogen treatment with placebo<sup>332</sup> and one trial compared a nonprescription treatment with placebo.<sup>125</sup>

The Rigano et al. (2001)<sup>332</sup> trial compared a standard dose estrogen patch with placebo and measured sexual activity in three age subgroups (48 to 50, 51 to 53, and 54 to 56). Treatment with estrogen resulted in a larger percent of participants reporting decreased sexual activity compared with placebo, and the effect was greatest in the oldest age group.

Davis et al. (2001)<sup>125</sup> compared Chinese herbs with placebo, among women less than 55 years old and greater than or equal to 55 years old. MENQOL sexual score was the outcome reported. Both age groups experienced an improvement in MENQOL sexual score, but the improvement was not statistically significant.

### Sexual Function by BMI

One trial conducted subgroup analyses on sexual function by BMI (Appendix N-Table 8).<sup>125</sup> Davis et al compared Chinese medicinal herbs and placebo for women with BMI less than or equal to 25 and BMI greater than 25. Neither BMI subgroup experienced a significant difference in MENQOL sexual score with treatment compared with placebo.<sup>125</sup>

### Sexual Function by Time Since Menopause

One trial conducted subgroup analyses on sexual function by time since menopause (Appendix N-Table 9).<sup>125</sup> Davis et al compared Chinese medicinal herbs with placebo in women with less than four years of amenorrhea and women with greater than or equal to four years of amenorrhea. The difference in change from placebo in MENQOL sexual score was slightly lower in participants with greater than or equal to four years of amenorrhea, but statistical significance was not reported.<sup>125</sup>

## Sexual Function by Uterus Status

One trial conducted subgroup analyses on sexual function by uterus status (Appendix N-Table 10).<sup>238</sup> The frequency of satisfying sexual episodes per week was compared between two doses of testosterone (0.15 mg and 0.30 mg) with placebo in participants with natural or surgical menopause. In women with natural menopause, a significant difference from placebo in frequency of satisfying sexual episodes per week was observed in the 0.15 mg testosterone group ( $p=0.02$ ) and in the 0.30 mg testosterone group ( $p<0.001$ ). In participants with surgical menopause, no significant differences from placebo were observed in the frequency of satisfying sexual episodes per week.<sup>238</sup>

## Psychological Symptoms

**Table 93. Evidence base for subgroup analyses reporting psychological outcomes**

Subgroup	Nonhormone Prescription Therapies		Nonprescription Therapies
	Hormone Therapies		
Age			Davis 2001 <sup>125</sup>
BMI			Davis 2001 <sup>125</sup>
Time since menopause	Strickler 2000 <sup>185</sup>	Kornstein 2010	Davis 2001 <sup>125</sup> Osmer 2005 <sup>152</sup>
Comorbidities	Rudolph 2004 <sup>175</sup> Strickler 2000 <sup>185</sup>		

**Table 94. Quality assessment for studies reporting psychological symptoms subgroup analyses**

Study	Comparable Groups	Researcher/Subjects Blinded	Adequate Concealment	Comparable Groups Maintained	Differential Loss To Followup	Measures Equal, Reliable	Interventions Clear	Outcomes Defined	Intention To Treat Analysis	Study Quality
Davis 2001 <sup>125</sup>	Y	Y	U	Y	Y	Y	Y	Y	U	Poor
Kornstein 2010 <sup>206</sup>	Y	Y	U	Y	N	Y	Y	Y	N	Poor
Osmer 2005 <sup>152</sup>	U	Y	Y	Y	N	Y	Y	Y	Y	Poor
Rudolph 2004 <sup>175</sup>	Y	Y	U	Y	Y	Y	Y	Y	Y	Poor
Strickler 2000 <sup>185</sup>	Y	Y	Y	U	U	Y	Y	Y	Y	Poor

## Psychological Symptoms by Age

One trial conducted subgroup analyses on psychological symptoms by age (less than 55 and greater than or equal to 55 years) (Appendix N-Table 11),<sup>125</sup> comparing Chinese medicinal herbs with placebo. Neither age group showed a statistically significant difference in MENQOL psychological score between the treatment and placebo groups.

## Psychological Symptoms by BMI

One trial conducted subgroup analyses on psychological symptoms by BMI (Appendix N-Table 12).<sup>125</sup> Davis et al. (2001) compared Chinese medicinal herbs with placebo in women with a BMI less than or equal to 25 and women with a BMI greater than 25. Neither BMI group experienced significant differences in MENQOL psychological score between the treatment and placebo groups.

## Psychological Symptoms by Time Since Menopause

Four trials conducted subgroup analyses on psychological symptoms by time since menopause (Appendix N-Table 13).<sup>125, 152, 185</sup> Two trials compared nonprescription treatments (Chinese medicinal herbs<sup>125</sup> and black cohosh<sup>152</sup>) with placebo, one trial compared estrogen with placebo,<sup>185</sup> and one trial compared antidepressants with placebo.<sup>206</sup>

Davis et al. compared Chinese herbs with placebo among women experiencing amenorrhea less than four years and women experiencing amenorrhea greater than or equal to four years. The MENQOL psychological scores did not change significantly in either of the subgroups.<sup>125</sup>

Osmers et al. compared black cohosh (40 mg) with placebo and reported Menopausal Rating Scale psychological subscale scores for early climacteric women and late climacteric women. The authors report a significant improvement in psychological subscale scores among the early climacteric group ( $p=0.05$ ), but no significant change among the late climacteric women ( $p=0.08$ ).<sup>152</sup>

Strickler et al. compared a standard dose of conjugated equine estrogen with placebo and reported WHQ anxiety scores for women less than four years postmenopausal and women greater than or equal to four years postmenopausal. The WHQ anxiety scores did not change significantly in either subgroup of the treatment groups compared with placebo.<sup>185</sup>

Kornstein et al. compared an antidepressant (10 mg desvenlafaxine) with placebo and reported Hamilton depression scores for perimenopausal and postmenopausal women. Both subgroups experienced significant improvements in depression scores following antidepressant treatment compared with placebo.<sup>206</sup>

## Psychological Symptoms by Comorbidities

Two trials conducted subgroup analyses on psychological symptoms by comorbidities. (Appendix N-Table 14).<sup>175, 185</sup>

Rudolph et al. (2004)<sup>175</sup> compared the effect of high dose estrogen/progestin with placebo on Hamilton depression scores. Subgroup analyses were conducted for women with or without premenstrual syndrome or postnatal depression. Both groups experienced improvements in Hamilton depression scores, and there was not a significant difference between the groups ( $p=0.09$ ).

Strickler et al. (2000)<sup>185</sup> compared the effect of standard dose estrogen with placebo on WHQ anxiety scores. Subgroup analyses were conducted on women with a baseline anxiety score of less than 3.5 and women with a baseline anxiety score greater than or equal to 3.5. A significant reduction in WHQ anxiety scores was observed only in the subgroup with higher baseline anxiety scores.

## Quality of Life

**Table 95. Evidence base for subgroup analyses reporting quality of life outcomes**

Subgroup	Nonhormone Prescription		
	Hormone Therapies	Therapies	Nonprescription Therapies
Severity of symptoms	Maki 2007 <sup>331</sup>		
Time since menopause	Loh 2002 <sup>242</sup>		Osmers 2005 <sup>152</sup>

**Table 96. Quality assessment for studies reporting quality of life subgroup analyses**

Study	Comparable Groups	Researcher/Subjects Blinded	Adequate Concealment	Comparable Groups Maintained	Differential Loss To Followup	Measures Equal, Reliable	Interventions Clear	Outcomes Defined	Intention To Treat Analysis	Study Quality
Loh 2002 <sup>242</sup>	U	Y	U	U	N	Y	Y	Y	U	Poor
Maki 2007 <sup>331</sup>	U	Y	U	U	N	Y	Y	Y	Y	Poor
Osmer 2005 <sup>152</sup>	U	Y	Y	Y	N	Y	Y	Y	Y	Poor

## Quality of Life by Severity of Symptoms

One trial conducted subgroup analyses on quality of life by severity of symptoms (Appendix N-Table 15).<sup>331</sup> Maki et al compared standard dose estrogen/progestin with placebo and performed subgroup analyses on symptomatic women (hot flush severity score of greater than or equal to 1.2 at baseline ) and asymptomatic women (hot flush severity score less than 1.2 at baseline). Two different quality of life scales were used as outcomes: total Greene Climacteric Scale (GCS), a condition-specific quality of life scale designed for menopausal women, in which a lower score indicates a better quality of life and Utian Quality of Life (QOL) Scale, a general health quality of life scale, in which a higher score indicates a better quality of life. A statistically significant improvement in quality of life was reported with the Utian QOL among symptomatic women in the treatment group compared to placebo. There was no difference using the GCS scale between the subgroups.

## Quality of Life by Time Since Menopause

Two trials conducted subgroup analyses on quality of life by time since menopause (Appendix N-Table 16).<sup>152, 242</sup> One trial intervention was estrogen/progestin therapy<sup>242</sup> and the other trial was nonprescription (black cohosh).<sup>152</sup>

Loh et al. (2002)<sup>242</sup> compared low dose estrogen/progestin with standard dose estrogen/progestin and used Kupperman Index as an outcome. Subgroup analyses were performed on women whose time since menopause was less than three years and women whose time since menopause was greater than or equal to three years. No difference in total Kupperman Index was observed between the subgroups.

In the Osmer et al. (2005) trial,<sup>152</sup> black cohosh was compared with placebo and total Menopause Rating Scale (MRS) score was an outcome. Subgroup analyses were performed on early climacteric women and late climacteric women. For both early and late climacteric women, significant improvements in the black cohosh groups compared with the placebo groups was observed. No difference between subgroups was reported.

## Sleep Dysfunction

**Table 97. Evidence base for subgroup analyses reporting sleep dysfunction outcomes**

Subgroup	Nonhormone Prescription		
	Hormone Therapies	Therapies	Nonprescription Therapies
Age	Rigano 2001 <sup>332</sup>		
Severity of symptoms	Diem 2006 <sup>145</sup>		

**Table 98. Quality assessment for studies reporting sleep dysfunction subgroup analyses**

Study	Comparable groups	Researcher/ subjects blinded	Adequate concealment	Comparable groups maintained	Differential loss to followup	Measures equal, reliable	Interventions clear	Outcomes defined	Intention to treat analysis	Study Quality
Diem 2006 <sup>145</sup>	U	Y	Y	U	U	Y	Y	Y	Y	Poor
Rigano 2001 <sup>332</sup>	N	U	U	U	U	U	U	Y	U	Poor

## Sleep Dysfunction by Age

One trial conducted subgroup analyses on sleep dysfunction by age categories (Appendix KQ4-Table 17).<sup>332</sup> The intervention was a standard dose estrogen administered through a transdermal patch. All age groups receiving hormone therapy reported less insomnia compared to the placebo groups. Differences between subgroups were not reported.

## Sleep Dysfunction by Severity of Symptoms

One trial conducted subgroup analyses on sleep dysfunction by severity of symptoms categories (Appendix N-Table 18).<sup>145</sup> The intervention was an ultra-low-dose estrogen patch. Percent reporting trouble sleeping was the outcome. Subgroup analyses were performed for women with very frequent menopausal symptoms (hot flashes, vaginal dryness, trouble sleeping) at baseline and for women with somewhat frequent menopausal symptoms at baseline. No statistically significant differences in percent reporting trouble sleeping was observed between the estrogen group and the placebo group in either subgroup.

## Urogenital Atrophy

**Table 99. Evidence base for subgroup analyses reporting urogenital symptom outcomes**

Subgroup	Hormone Therapies	Nonhormone Prescription Therapies	Nonprescription Therapies
Severity of symptoms	Diem 2006 <sup>145</sup>		
Time since menopause			Osmer 2005 <sup>152</sup>

**Table 100. Quality assessment for studies reporting urogenital symptoms subgroup analyses**

Study	Comparable Groups	Researcher/ Subjects Blinded	Adequate Concealment	Comparable Groups Maintained	Differential Loss To Followup	Measures Equal, Reliable	Interventions Clear	Outcomes Defined	Intention To Treat Analysis	Study Quality
Diem 2006 <sup>145</sup>	U	Y	Y	U	U	Y	Y	Y	Y	Poor
Osmer 2005 <sup>152</sup>	U	Y	Y	Y	N	Y	Y	Y	Y	Poor

## Urogenital Symptoms by Severity of Symptoms

One trial conducted subgroup analyses on urogenital symptoms by severity of symptoms (Appendix N-Table 19).<sup>145</sup> The intervention was an ultra-low-dose estrogen patch compared with

placebo. The outcome was vaginal dryness, reported by the following subgroups: women with very frequent menopausal symptoms (hot flushes, vaginal dryness, trouble sleeping) at baseline and women with somewhat frequent menopausal symptoms at baseline. No statistically significant difference in vaginal dryness was observed between the hormone therapy group and the placebo group in either subgroup.

### **Urogenital Symptoms by Time Since Menopause**

One trial conducted subgroup analyses on urogenital symptoms by time since menopause (Appendix N-Table 20).<sup>152</sup> The intervention was a nonprescription therapy, black cohosh and the subgroups were early and late climacteric women. In both early and late climacteric women, black cohosh improved urogenital atrophy compared with placebo.

# Discussion

## Introduction

For women experiencing menopausal symptoms considering any of the agents examined here, the choice of treatment is influenced by therapeutic efficacy while considering other potential benefits and harms—particularly over the long-term (Figure 1). The results and conclusions of this review offer an evidenced-based guide to comparative efficacy as well as other important benefits and harms. In this final section, we discuss what has been learned from evidence reviewed together in relation to what is known, together with its limitations and gaps. But most importantly we place the entirety of the evidence in the context of the analytic framework incorporating the four Key Questions considered not in isolation, but as a whole to inform decisions by patients, health care providers, and policy makers.

## Symptom Relief

### Vasomotor Symptoms

A large body of evidence was identified comparing the efficacy of agents with placebo and other active treatments for the relief of vasomotor symptoms (Table 101). Trials were most numerous for estrogens, isoflavones, SSRI/SNRIs, gabapentin or pregabalin, ginseng, and black cohosh. Estrogens of any dose appeared more effective than any other comparator without apparent difference between doses or mode of administration. Few differences were apparent in the network meta-analysis among isoflavones, SSRI/SNRIs, gabapentin/pregabalin, and black cohosh. Whether ginseng might have any effect is unclear. A host of other agents have been studied, but evidence is limited to single trials.

The efficacy of estrogens in treating vasomotor symptoms is well established. The comparative effectiveness of other agents relative to estrogens had been less clear. Albeit limited by trial quality, findings from the network analysis allow us to draw conclusions concerning relative effectiveness. While nonhormone agents can ameliorate vasomotor symptoms (SMDs ranging from -0.26 to -0.41), none have estrogen’s effectiveness (SMDs ranging from -0.70 to -0.77).

**Table 101. Magnitude and strength of evidence of treatments for vasomotor symptoms; standardized effect sizes from pairwise comparisons<sup>1</sup>**

Comparisons	Comparators	Effect Size (SMD) (95 percent CI)	Effect Size Category <sup>a</sup>	Strength of Evidence
9	Estrogen High vs. Placebo	-0.72 (-0.99 to -0.44)	....	High
36	Estrogen Standard vs. Placebo	-0.79 (-0.92 to -0.66)	....	
46	Estrogen Low/Ultralow vs. Placebo	-0.70 (-0.83 to -0.58)	....	
13	Estrogen High vs. Standard	-0.15 (-0.40 to 0.09)	—	High
7	Estrogen High vs. Low/Ultralow	-0.16 (-0.39 to 0.07)	—	
21	Estrogen Standard vs. Low/Ultralow	-0.10 (-0.22 to 0.02)	—	

<sup>1</sup> To enable easy comparisons in this and the following tables, effect size categories are displayed to provide an indication of comparative efficacy. The categories are not intended to confer other significance and do not correspond to so-called small, medium, and large suggested by Cohen for the purposes of sample size calculation.

10	SSRI/SNRI vs. Placebo	-0.40 (-0.54 to -0.26)	***	High
29	Isoflavones vs. Placebo	-0.41 (-0.58 to -0.25)	***	Moderate
4	Gabapentin/Pregabalin vs. Placebo	-0.33 (-0.44 to -0.22)	**	Moderate
3	Black Cohosh vs. Placebo	-0.26 (-0.43 to -0.09)	**	Low
3	Ginseng vs. Placebo	-0.41 (-0.83 to 0.02)	***	Low
8	Estrogen mode a vs. mode b	Not estimated	—	Moderate

<sup>a</sup> • (0 to > -0.2); \*\* (-0.2 to > -0.4); \*\*\* (-0.4 to > -0.6); \*\*\*\* (< -0.6); — (equivalent)

## Quality of Life

Trials evaluating numerous agents reported some quality of life metric, but the evidence base included more than a single trial for estrogens, isoflavones, SSRI/SNRIs, ginseng, black cohosh, and DHEA. Compared with placebo, improved quality of life scores accompanied estrogens with standardized effect sizes exceeding 0.40 with moderate or high strength of evidence; effect sizes for all other agents were lesser in magnitude or low SOE. Similarly, estrogens ranked highest in the network comparison. For estrogens, there was no apparent difference in effect according to mode of administration. Quality of life scores were reported from trials of many nonprescription agents, but results from single trials do not allow conclusions concerning effects.

We found improved global quality of life scores in women taking estrogens. Yet no effect was apparent in “Women’s International Study of long Duration Oestrogen after The Menopause” (WISDOM)<sup>26</sup> or WHI.<sup>27,28</sup> Results from these trials appeared somewhat discrepant in the analyses and is likely attributable to older age and lesser symptom severity of enrolled women. For the larger body of comparisons in women receiving estrogens, despite between-trial variability, results were more consistent. The general pattern of comparative efficacy seen with quality of life scores paralleled results for other vasomotor and other symptoms.

**Table 102. Magnitude and strength of evidence of treatments for quality of life; standardized effect sizes from pairwise comparisons**

Comparisons	Comparators	Effect Size (SMD) (95 percent CI)	Effect Size Category <sup>a</sup>	Strength of Evidence
4	Estrogen High vs. Placebo	0.70 (0.40 to 1.01)	****	
21	Estrogen Standard vs. Placebo	0.63 (0.47 to 0.78)	****	High
12	Estrogen Low/Ultralow vs. Placebo	0.40 (0.24 to 0.56)	***	
6	Estrogen Standard vs. High	0.07 (-0.05 to 0.18)	—	
8	Estrogen Standard vs. Low/Ultralow	0.13 (-0.04 to 0.29)	—	High
2	Estrogen High vs. Low/Ultralow	-0.10 (-0.29 to 0.09)	—	
5	SSRI/SNRI vs. Placebo	0.27 (0.18 to 0.36)	**	
18	Isoflavones vs. Placebo	0.17 (0.06 to 0.29)	•	Moderate
3	Black Cohosh vs. Placebo	0.40 (0.18 to 0.63)	***	Moderate
3	Ginseng vs. Placebo	0.19 (0.01 to 0.36)	•	Low
3	DHEA vs. Placebo	Not estimated	i	Insufficient
7	Estrogen mode a vs. mode b	Not estimated	—	Moderate

<sup>a</sup> • (0 to < 0.2); \*\* (0.2 to < 0.4); \*\*\* (0.4 to < 0.6); \*\*\*\* (> 0.6); — (equivalent); i insufficient

## Psychological Symptoms

Just over one-third of trials examining symptom treatment reported a psychological outcome—depression, anxiety, and global mental health—and often more than one. Only half specified some psychological symptom as a primary outcome. Overall, the samples were not selected to represent populations with clinical depression or anxiety. Compared with placebo,

standardized effect sizes were in general not large (i.e., SMD between -0.5 and 0) for any of the agents studied for any psychological domain (Table 103). Furthermore, the strength of evidence was at least moderate only for some effects of estrogens and SSRI/SNRIs.

An increased risk for depression, in the absence of prior depressive illness, during the menopausal transition has been described<sup>29</sup> and may be associated with vasomotor symptoms.<sup>30</sup> The effects assessed here may provide guidance when menopausal women are experiencing psychological symptoms.

**Table 103. Magnitude and strength of evidence of treatments for psychological symptoms; standardized effect sizes from pairwise comparisons**

Comparisons	Comparators	Effect Size (SMD) (95 percent CI)	Effect Size Category <sup>a</sup>	Strength of Evidence
<b>Depression</b>				
4	Estrogen High vs. Placebo	-0.64 (-0.94 to -0.33)	****	Moderate
11	Estrogen Standard vs. Placebo	-0.19 (-0.31 to -0.07)	•	Moderate
3	Estrogen Low/Ultra-low vs. Placebo	-0.04 (-0.41 to 0.31)	i	Insufficient
8	Isoflavones vs. Placebo	-0.41 (-0.69 to -0.13)	***	Low
3	SSRI/SNRI vs. Placebo	-0.40 (-0.59 to -0.22)	***	Moderate
<b>Anxiety</b>				
2	Estrogen High vs. Placebo	-0.35 (-0.58 to 0.13)	**	Low
8	Estrogen Standard vs. Placebo	-0.16 (-0.34 to 0.03)	i	Insufficient
3	Estrogen Low/Ultra-low vs. Placebo	-0.19 (-0.41 to 0.02)	•	Low
7	Isoflavones vs. Placebo	-0.53 (-0.87 to -0.23)	***	Low
2	SSRI/SNRI vs. Placebo	-0.31 (-0.53 to -0.08)	**	Low
<b>Global</b>				
9	Estrogen Standard vs. Placebo	-0.03 (-0.10 to 0.04)	i	Insufficient
7	Estrogen Low/Ultra-low vs. Placebo	-0.24 (-0.45 to -0.02)	**	High
6	Isoflavones vs. Placebo	-0.12 (-0.26 to 0.01)	•	Low
4	SSRI/SNRI vs. Placebo	-0.39 (-0.63 to -0.15)	**	Moderate
2	Gabapentin/Pregabalin vs. Placebo	-0.22 (-0.46 to 0.03)	i	Insufficient

<sup>a</sup> • (0 to > -0.2); \*\* (-0.2 to > -0.4) ; \*\*\* (-0.4 to > -0.6) ; \*\*\*\* (< -0.6); — (equivalent); i insufficient

## Sexual Function

Some measure of sexual function was reported in less than a third of trials; half of those trials specified the outcome as primary (Table 107). Outcomes were reported in four domains: pain (dyspareunia), a global metric, activity, and interest. Vaginal estrogens improved pain most convincingly (high strength of evidence), while lower pain scores with oral estrogens were less certain (low strength of evidence). There was a modest increase in global measures with estrogens. No agent appeared to enhance measures of interest. Sexually satisfying episodes were more frequent in the comparison of testosterone with placebo—slightly more than one extra episode reported every 4 weeks (strength of evidence moderate). Overall, these results are generally consistent with evidence-informed expert clinical opinion.<sup>1</sup>

The Prevalence of Female Sexual Problems Associated with Distress and Determinants of Treatment Seeking (PRESIDE)<sup>31</sup> estimated approximately 15 percent of women aged 45 to 64

experienced some form of sexual distress. Although many trials reported sexual function outcomes during, we identified only one quantitative review. The study included literature published between 1972 and 1992.<sup>32</sup> In the analysis, standardized effect representing any domain were combined from 108 studies of estrogen therapy yielding -0.67 (SD 1.23)—somewhat larger in magnitude that obtained in this review.

**Table 104. Magnitude and strength of evidence of treatments for sexual function; standardized effect sizes from pairwise comparisons**

Comparisons	Comparators	Effect Size (SMD) (95 percent CI)	Effect Size Category <sup>a</sup>	Strength of Evidence
<b>Pain (lower is better)</b>				
10	Vaginally applied estrogens vs. placebo	-0.50 (-0.71 to -0.29)	***	High
3	Oral estrogens vs. placebo	-0.44 (-1.05 to 0.17)	***	Low
13	All estrogens vs. placebo	-0.49 (-0.69 to -0.29)	***	High
<b>Global (higher is better)</b>				
10	All estrogens vs. placebo	0.28 (0.16 to 0.41)	**	High
2	SNRIs vs. placebo	0.11 (0.02 to 0.19)	i	Insufficient
<b>Interest (higher is better)</b>				
3	All estrogens vs. placebo	0.43 (-0.02 to 0.89)	i	Insufficient
3	Isoflavones vs. placebo	0.31 (-0.24 to 0.86)	i	Insufficient
<b>Pain, Interest, Global</b>				
10	Estrogen mode a vs. mode b	Not estimated	—	Moderate
<b>Activity</b>				
		SSE/4 weeks		
4	Testosterone, no women with intact uteri/ovaries	1.05 (0.64 to 1.45)	NA	Moderate
4	Testosterone, women with/without uteri/ovaries	1.31 (0.89 to 1.72)	NA	
8	Testosterone, all trials	1.17 (0.88 to 1.46)	NA	

<sup>a</sup> For negative effect sizes • (0 to > -0.2); \*\* (-0.2 to > -0.4); \*\*\* (-0.4 to > -0.6); \*\*\*\* (< -0.6). For positive effect sizes • (0 to < 0.2); \*\* (0.2 to < 0.4); \*\*\* (0.4 to < 0.6); \*\*\*\* (> 0.6); — (equivalent); i insufficient  
SSE Satisfying sexual episodes

## Urogenital Atrophy

One quarter of trials reported urogenital atrophy outcomes—a primary outcome in 60 percent. Although multiple scales were employed, the strength of evidence was high that either oral or vaginal estrogens improve symptoms with standardized effect sizes for vaginal estrogens approximately twice that of nonvaginal estrogens. The strength of evidence was low for other agents (isoflavones and black cohosh).

The conclusions here are similar to those provided to clinicians<sup>1</sup> when considering treating symptoms that may be experienced by as many as 40 percent of postmenopausal women.<sup>33</sup> A 2006 Cochrane review including 19 trials concluded that vaginal or oral estrogens were equally effective for treating vaginal atrophy.<sup>34</sup> These results indicate, albeit indirectly based on placebo comparisons, that a greater magnitude of effect for vaginal compared with oral administration.

**Table 105. Magnitude and strength of evidence of treatments for urogenital atrophy; standardized effect sizes from pairwise comparisons**

Comparisons	Comparators	Effect Size (SMD) (95 percent CI)	Effect Size Category <sup>a</sup>	Strength of Evidence
11	Estrogen vaginal vs. placebo	-0.60 (-0.86 to -0.33)	••••	High
16	Nonvaginal estrogen vs. placebo	-0.31 (-0.40 to -0.22)	••	High
5	Isoflavones vs. placebo	-0.57 (-0.90 to -0.24)	•••	Low
2	Black Cohosh vs. placebo	-0.27 (-0.44 to -0.11)	•••	Low
7	Estrogen mode a vs. mode b	not estimated	—	Low

<sup>a</sup> • (0 to > -0.2); •• (-0.2 to > -0.4) ; ••• (-0.4 to > -0.6) ; •••• (< -0.6); — (equivalent)

## Sleep

Many trials ascertained self-reported sleep outcomes, but only a single trial examined a drug FDA-approved for use in insomnia (eszopiclone) that was highly effective. On a standardized effect scale, sleep improved with eszopiclone approximately three-fold greater than with estrogens or any other agent. This suggests that modestly improved sleep accompanies other agents, including estrogens, used to treat menopausal symptoms (Table 106).

While sleep disturbances during menopause are common,<sup>35</sup> how often they are secondary to menopausal symptoms is not well defined. Sedative hypnotics agents are not generally used to treat menopausal symptoms and so were not represented in the trials identified. Reported improvements in sleep evident with other agents such as estrogens is possibly due to treatment of vasomotor symptoms, but requires evidence not considered here.

**Table 106. Magnitude and strength of evidence of treatments for sleep; standardized effect sizes from pairwise comparisons**

Comparisons	Comparators	Effect Size (SMD) (95 percent CI)	Effect Size Category <sup>a</sup>	Strength of Evidence
25	Estrogen vs. placebo	0.36 (0.26 to 0.46)	••	High
3	Estrogen vs. estrogen different dose	-0.25 (-0.67 to 0.18)	i	Insufficient
6	Isoflavones vs. placebo	0.35 (-0.43 to 1.13)	i	Insufficient
2	Ginseng vs. placebo	not estimated	i	Insufficient

<sup>a</sup> • (0 to < 0.2); •• (0.2 to < 0.4) ; ••• (0.4 to < 0.6) ; •••• (> 0.6); i insufficient

## Limitations of the Evidence Base on Symptom Relief

The body of evidence synthesized for Key Question 1 was large with many trials rated poor quality. However, the challenges of synthesizing this evidence extends far beyond trial quality to limitations incompletely incorporated in strength of evidence assessments. These include:

- Use of different outcome scales or metrics
- Necessity of calculating standardized effect sizes and inherent difficulties estimating from publications
- Potential differences in populations represented by trial samples
- Potential for selective outcome reporting

Some two decades ago, in her review of sexual function, Myers foreshadowed the difficulties encountered here across all outcomes—variable scales, metrics, and definitions.<sup>32</sup> Even directionality of scales within the various outcomes often differed. The absence of standardized outcome reporting limits the ability to quantify effects using metrics easily and transparently translated to quantities such as clinically meaningful improvement. Given the well-described placebo effect, at least for vasomotor symptoms,<sup>99</sup> this limitation is important to consider interpreting results. An alternative approach to the one adopted here, would be to limit trials synthesized to those reporting similar outcome scales or metrics. Although appealing in many respects, if studies reporting some identical outcome metric were not representative of all trials, the potential for introducing bias exists. So while interpretive limitations accompany standardized effect sizes, their use allows including and pooling evidence from multiple trials, which would not be feasible otherwise. In many instances here, it enabled at the very least providing comparative efficacy rankings.

On the surface, calculating standardized effect sizes might appear trivial. Here it was anything but trivial. As outlined in the methods, there were a number of ways to obtain effect sizes from the continuous measures reported; trials typically did not report a between group difference and variance (standard deviation) allowing the most straightforward calculating of standardized effects. To avoid excluding trial results, other calculations were required including the use of p-values that typically were not reported exactly. Additionally, some results were reported as simply nonsignificant. In the case where results were pooled, excluding nonsignificant results lacking a p-value would introduce bias. While imputation allowed including those results, it introduces uncertainty. Fortunately, the number of p-values requiring imputation was small. A separate issue was the occasional outlier encountered because trials sometimes reported unusually large effects. Potential outliers required performing analyses to be certain effects could not be attributed to them.

A separate concern is that while trial populations included women experiencing menopause, they were differences in mean age, length of follow-up, and symptom severities. While the initial intent was to examine subgroups according to characteristics such as the presence of a uterus, lack of reporting did not allow doing so. Conclusions then apply to average women across all trials.

It is also difficult to evaluate potential selective outcome reporting from the included trials. Vasomotor symptoms were reported in about three quarters of trials but all other outcomes in fewer than half. While some trials, such as those of sexual function or vaginal atrophy, were clearly not designed to primarily assess all outcomes, insignificant results may have gone unreported. For some of the outcomes reported, in only half was the outcome reported as primary.

Finally, the results do not allow assessing whether effects on different outcomes are independent. It is conceivable that the consequence of fewer vasomotor symptoms is improved quality of life, sleep, or better psychological outcomes. Causally, it may be that the focus of therapy need not consider treatment efficacy for all outcomes, but rather a few—most likely beginning with vasomotor symptoms.

## **Compounded Hormone Therapies**

Compounded hormone therapies are commonly prescribed, often in combination with some testing for hormone levels, with effectively no direct evidence base. We identified a single randomized controlled trial examining pharmacokinetics in 40 women studied for 16 days.<sup>36</sup> No

studies were identified examining the safety of the compounding practices for hormone therapies.

## Other Benefits and Harms

### Hormone Therapy Preparations

In 1979, the National Institutes of Health convened their first consensus conference on estrogen use in postmenopausal women.<sup>335</sup> While breast and endometrial cancer were prominent in the summary, there was no mention of heart disease. Some 3 decades later there is now a robust evidence base allowing conclusions regarding both beneficial and harmful outcomes.

Evidence included in the recent review by Nelson<sup>37</sup> was assessed here with concordant conclusions. Because a majority of evidence derived from WHI trials, representing a target population overlapping the one for this review, assessing applicability of findings required considering observational study results. Still, the picture of long-term effects emerges with reasonable clarity as summarized in Table 107. The USPSTF review reported differences in event rates with estrogen/progestin or estrogen compared with placebo. However, extrapolating absolute rates from the WHI samples to the target population of this review is potentially problematic. In broad relative terms gall bladder disease is the most frequent occurrence with thromboembolic events, stroke, and breast cancer less frequent. While less frequent they are not insignificant. For example, hormone therapy in women aged 50 to 74 years has been estimated responsible for 9 percent of all strokes in women in 2012.<sup>336</sup>

**Table 107. Long-term effects of hormone therapy preparations summarized**

Outcome	Risk	Treatment vs. Placebo	Strength of Evidence	Comment
Breast Cancer	↑	Estrogen/Progestin	High	
	—	Estrogen	Low	Inconsistent
Gall bladder disease	↑	Estrogen/Progestin	Moderate	Consistency unknown with 1 trial
	↑	Estrogen	Moderate	Consistency unknown with 1 trial
Venous Thromboembolic Events	↑	Estrogen/Progestin	Moderate	Consistency unknown with 1 trial
	↑	Estrogen	High	
Stroke	↑	Estrogen/Progestin	Moderate	Consistency unknown with 1 trial
	↑	Estrogen	High	
Ovarian Cancer	↑	Estrogen/Progestin	Low	Consistency unknown with 1 trial; imprecise with few cases
Colorectal Cancer	—	Estrogen/Progestin	Moderate	Consistency unknown with 1 trial
	—	Estrogen	Moderate	Consistency unknown with 1 trial
CHD	—	Estrogen/Progestin	Low	Consistency unknown with 1 trial; imprecise
	—	Estrogen	Moderate	Consistency unknown
Endometrial Cancer	—	Estrogen/Progestin	Low	Inconsistent; imprecise
Osteoporotic Fractures	↓	Estrogen/Progestin	Moderate	Inconsistency between 2 trials
	↓	Estrogen	Moderate	Consistency unknown with 1 trial

Risk: ↑ increased, ↓ decrease, — no change

One limitation of the evidence base concerning long-term outcomes derives from necessity to rely on results of randomized controlled trials. There are well described discrepant conclusions

concerning these associations between observational studies and randomized controlled trials.<sup>41</sup> The discrepancies have been attributed to two primary reasons—selection bias and time-varying confounding.<sup>42-44</sup> While the association with cardiovascular outcomes has been most scrutinized, difficulties assessing causal effects of hormone therapy from observational data appear to extend to other outcomes including hip fractures<sup>42</sup> and colorectal cancer.<sup>44</sup> As noted throughout, trials have been conducted from a target population overlapping with the one for this review creating some challenges for assessing applicability. Still, there is considerable certainty in the effects assessed—a remarkable body of evidence accrued since the 1979 NIH consensus conference.

## Nonhormone Therapy Preparations

The evidence base informing other potential benefits and harms of nonhormone therapies is limited, but does not suggest harmful long-term effects are likely for those agents studied (Table 108). We identified large trials examining vitamin E, small trials of isoflavones, and observational studies evaluating antidepressants that did not always distinguish risks for the classes of agents used to treat symptoms (SSRI/SNRI). While no salient benefits were identified, neither were safety signals apparent. However, given the large numbers women potentially taking these agents some caution is advised particularly for nonprescription agents. For example, the possibility of increased mortality with high dose vitamin E has been raised.<sup>38</sup> Additionally, case reports of hepatotoxicity with black cohosh have been published.<sup>39</sup> This association has been debated,<sup>40</sup> but surveillance for adverse effects of nonprescription agents is generally inadequate. Safety data are also needed for the broad array of herbs and botanicals used to treat menopausal symptoms.

There are several further limitations to this evidence to consider. Many studies included women of all ages and therefore were excluded unless subgroup analyses on older women or menopausal women were specified. Much of the research available on the long-term effects of isoflavones and vitamin E consisted of population-based dietary studies and therefore did not meet inclusion criteria. Intermediate outcomes were reported in many of the studies. For example, bone density rather than osteoporotic fractures, and cholesterol rather than cardiovascular events. Finally, in studies that included all women rather than focusing on menopausal women, it was difficult to discern if exposure (to antidepressants, isoflavones) occurred during menopausal years.

**Table 108. Long-term effects of nonhormone therapy preparations summarized**

Outcome	Risk	Treatment vs. Placebo	Strength of Evidence
Breast cancer	—	Vitamin E	High
Breast cancer	—	SSRI/SNRI	Low
Colorectal Cancer	—	Vitamin E	High
Cardiovascular Events	—	Vitamin E	High
Cardiovascular Death	↓	Vitamin E	Low
Osteoporotic Fractures	↑	SSRI	Low
Osteoporotic Fractures	↓	Isoflavones	Insufficient
Ovarian Cancer	—	Vitamin E	Low
Breast, Endometrial, Ovarian Cancer	↑	Any antidepressant	Insufficient

## Symptom Relief in Subgroups

A small subset of trials identified for Key Question 1 reported subgroup analyses on symptom relief: 10 for hormone therapies, two nonhormone prescription therapies, and four nonprescription therapies. Trials with hormone therapies included analyses by age, severity of symptoms, time since menopause, and uterine status. One trial of a nonhormone prescription therapy (escitalopram) provided a subgroup analysis by race. Trials with nonprescription therapies reported outcomes by age, BMI, severity of symptoms, and time since menopause. For example, age group subpopulations were defined as <50, 50-59, and  $\geq 60$  in one trial and <55 and  $\geq 55$  in another trial. None of the subgroup analyses could be pooled, as no two trials had the same comparators, definitions of subgroups, and outcomes. The sparse evidence did not allow rating strength of evidence.

## Research Gaps

The principal gaps in the evidence on symptom relief include the following: safety data on nonprescription agents, lack of evidence on compounded hormone therapies, potential for predicting treatment response, and independence of some treatment outcomes:

- A large number of nonprescription agents were studied in individual trials. These agents are unregulated and safety data may be limited or absent. As women may elect to try these agents, those data need to be available.
- Millions of women use compounded hormone treatments. Yet there is a stark absence of evidence concerning compounded hormone therapies, and the diagnostic methods (saliva tests) used to determine the personalized dosages. While the gap is most concerning regarding safety, efficacy issues are important as well.
- The ability to individualize or personalize treatments according to some characteristics is a common goal in medicine. While the efficacy of estrogen treatment for symptom relief is so substantial that identifying some predictors of response would unlikely be fruitful. However, for the less efficacious interventions, identifying predictors could be helpful for women having reasons to forgo hormone treatments.
- As noted previously, although we considered six categories of symptom relief outcomes, the extent of correlated response (not symptom presence) among them was unclear in the evidence. While not an objective of this review, the evidence would provide little opportunity to examine that question.

The most important previous gaps in the evidence concerning long-term effects of hormone therapies have been filled. For some nonhormone therapies (Table 108), with reasonable certainty (i.e., moderate or greater strength of evidence) significant safety issues have not been apparent; the same cannot be said for the entirety of the nonprescription agents.

Finally, estrogen therapy has the greatest efficacy relieving most symptoms and is accompanied by other potentially important benefits as well as some tradeoffs in the form of harms (varying according to whether combined with progestin). Given the number of outcomes to consider with different exposure effects (e.g., duration of use); the overall risk-benefit calculus is not simple. Juxtaposing evidence concerning symptom relief (as obtained here) with models for the long-term benefits and harms<sup>45</sup> according to patient characteristics (i.e., lower risk of hip fracture in blacks) could facilitate informed decisions by women and health care providers.

## Implications for Clinical and Policy Decision-Making

The implications of the conclusions from this review for clinical decision-making are straightforward. The results provide a guide to comparative efficacy alongside potential long-term benefits and harms; all are weighed in clinical decisions. Possibly most useful, for vasomotor symptoms and quality of life, the review provides clinicians with a simple ranked efficacy comparison for the most commonly used treatments.

From the policy perspective, there are two salient issues to consider. First, since a 2007 Senate hearing, no evidence on compounded hormones has appeared. Efforts to address that absence are important. Second, is to clearly define and communicate, and translate when necessary, the net clinical benefits of hormone treatments according to duration of therapy when initiated for symptom relief (as many organizations have worked towards). Effective tools disseminating evidence in the most decision-informative could be considered.<sup>2</sup>

## Limitations of the Comparative Effectiveness Review Process

This review was a large undertaking. The variable manner in which trials reported results, multiple trial arms, multiple treatments, along with the goal of not excluding results for any a priori potentially arbitrary reason (e.g., reporting outcomes using a particular metric, or reported mean change and standard deviation) required abstracting, verifying, and managing a large amount of data—34,000 data elements and over 250 digitized figures for six outcome categories for Key Question 1. Obtaining standardized effects can be challenging.<sup>46</sup> A number of steps are required to calculate effect magnitudes often for more than one trial arm. There are multiple ways to obtain an effect measure and standard deviation for each trial arm; different approaches may not yield identical results. Furthermore, given multiple trial arms and multiple outcomes, the number of calculations required was substantial. We stipulated an order in method to perform those calculations but judgment was still required. Confidence intervals and strength of evidence ratings do not incorporate this analytical uncertainty. Whether type I error rates should be higher (e.g., calculated effect estimates higher than or accompanied with lower variances than the true values) or lower (e.g., calculated effect estimates lower than or accompanied with higher variances than the true values) is difficult to ascertain. What is clear, however, is that pooled estimates should be interpreted with this understanding. Finally, the analyses included two network and many standard pairwise meta-analyses. Network meta-analyses are not trivial undertakings.

Analyses of the multiple treatments required imposing some classification scheme that has limitations. For example, the estrogen dose categorization scheme did not consider progestin, or distinguish between combined and sequential progestin administration. Progestin use was problematic to separate because trials may have not given to women without a uterus, yet reported an effect for the entire sample.

Finally, interpreting network and pairwise meta-analyses deserves comment. In the pairwise meta-analysis the randomized comparison is entirely preserved when pooling. Underlying the network of comparisons is an assumed exchangeability (similarity) of patient samples or the population from which they were drawn. All enrolled women were menopausal or peri-

---

<sup>2</sup> For example, the Choosing Wisely® program.

menopausal, but there were some differences in the samples as noted in the review. Despite this, the closeness of almost all the network and pairwise estimates argues any discrepancies likely small. In the end, both analyses are informative and can be viewed as complementary.

## **Conclusions**

Women experiencing symptoms of menopause can consider a number of potential treatments of varying efficacy. From a large body of evidence, there is considerable certainty that estrogens are most effective relieving the common symptoms. Estrogens are accompanied by other potential long-term benefits and harms that require considering. Compared with estrogen, other agents have lesser efficacy and limited evidence on long-term benefits and harms.

## References

1. Society NAM. Menopause Practice: A Clinician's Guide. 4th Edition ed. Mayfield Heights, OH: North American Menopause Society; 2010.
2. Soules MR, Sherman S, Parrott E, et al. Executive summary: Stages of Reproductive Aging Workshop (STRAW). *Climacteric : the journal of the International Menopause Society*. 2001 Dec;4(4):267-72. PMID: 11770182.
3. Soules MR, Sherman S, Parrott E, et al. Executive summary: Stages of Reproductive Aging Workshop (STRAW) Park City, Utah, July, 2001. *Menopause*. 2001 Nov-Dec;8(6):402-7. PMID: 21141649.
4. Soules MR, Sherman S, Parrott E, et al. Executive summary: Stages of Reproductive Aging Workshop (STRAW). *Fertil Steril*. 2001 Nov;76(5):874-8. PMID: 11704104.
5. McKinlay SM, Brambilla DJ, Posner JG. The normal menopause transition. *Maturitas*. 1992 Jan;14(2):103-15. PMID: 1565019.
6. Nelson HD, Haney E, Humphrey L, et al. Management of menopause-related symptoms. Evidence report/technology assessment. 2005 Mar(120):1-6. PMID: 15910013.
7. Woods NF, Mitchell ES. Symptoms during the perimenopause: prevalence, severity, trajectory, and significance in women's lives. *Am J Med*. 2005 Dec 19;118 Suppl 12B:14-24. PMID: 16414323.
8. . Proceedings from the NIH State-of-the-Science Conference on Management of Menopause-Related Symptoms, March 21-23, 2005, Bethesda, Maryland, USA. *Am J Med*. 2005 Dec 19;118 Suppl 12B:1-171. PMID: 16414320.
9. Politi MC, Schleinitz MD, Col NF. Revisiting the duration of vasomotor symptoms of menopause: a meta-analysis. *J Gen Intern Med*. 2008 Sep;23(9):1507-13. PMID: 18521690.
10. Freeman EW, Sammel MD, Lin H, et al. Duration of menopausal hot flashes and associated risk factors. *Obstet Gynecol*. 2011 May;117(5):1095-104. PMID: 21508748.
11. Gold EB, Block G, Crawford S, et al. Lifestyle and demographic factors in relation to vasomotor symptoms: baseline results from the Study of Women's Health Across the Nation. *Am J Epidemiol*. 2004 Jun 15;159(12):1189-99. PMID: 15191936.
12. Organization MMaSS. Introductory Guide to MeDRA Version 14.0. Chantilly, VA: International Federation of Pharmaceutical Manufacturers and Associations; 2011.
13. Helfand M, Balshem H. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. August 2009. Rockville, MD. [http://www.effectivehealthcare.ahrq.gov/index.cfm/s\\_earch-for-guides-reviews-and-reports/?pageaction=displayProduct&productID=318](http://www.effectivehealthcare.ahrq.gov/index.cfm/s_earch-for-guides-reviews-and-reports/?pageaction=displayProduct&productID=318).
14. Harris RP, Helfand M, Woolf SH, et al. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med*. 2001 Apr;20(3 Suppl):21-35. PMID: 11306229.
15. Fu R, Gartlehner G, Grant M, et al. Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. *J Clin Epidemiol*. 2011 Nov;64(11):1187-97. PMID: 21477993.
16. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2011.
17. Schwarzer G. meta: Meta-Analysis with R (version 1.6-1). 2011. <http://CRAN.R-project.org/>. Accessed on June 16 2011.
18. Re ACD. compute.es: Compute Effect Sizes. 2012.
19. Wickham H. ggplot2: elegant graphics for data analysis: Springer New York; 2009.
20. Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. *Psychol Meth*. 1998;3(4):486-504. PMID: none.
21. Rucker G, Schwarzer G, Carpenter JR, et al. Undue reliance on I(2) in assessing heterogeneity may mislead. *BMC Med Res Methodol*. 2008;8:79. PMID: 19036172.
22. Song F, Altman DG, Glenny AM, et al. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ*. 2003 Mar 1;326(7387):472. PMID: 12609941.
23. Owens DK, Lohr KN, Atkins D, et al. AHRQ series paper 5: grading the strength of a body of evidence when comparing medical interventions--agency for healthcare research and quality and the effective health-care program. *J Clin Epidemiol*. 2010 May;63(5):513-23. PMID: 19595577.
24. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of

- evidence. *J Clin Epidemiol*. 2011 Apr;64(4):401-6. PMID: 21208779.
25. Treadwell JR, Uhl S, Tipton K, et al. Assessing equivalence and noninferiority. *J Clin Epidemiol*. 2012 Nov;65(11):1144-9. PMID: 22732455.
  26. Welton AJ, Vickers MR, Kim J, et al. Health related quality of life after combined hormone replacement therapy: randomised controlled trial. *BMJ*. 2008;337:a1190. PMID: 18719013.
  27. Brunner, R L, Gass, et al. Effects of conjugated equine estrogen on health-related quality of life in postmenopausal women with hysterectomy: results from the Women's Health Initiative Randomized Clinical Trial. *Arch Intern Med*. 2005 Sep 26(17):1976-86. PMID: 16186467
  28. Hays, J, Ockene, et al. Effects of estrogen plus progestin on health-related quality of life. *N Engl J Med*. 2003 May 8(19):1839-54. PMID: 12642637
  29. Freeman EW, Sammel MD, Lin H, et al. Associations of hormones and menopausal status with depressed mood in women with no history of depression. *Arch Gen Psychiatry*. 2006 Apr;63(4):375-82. PMID: 16585466.
  30. Cohen LS, Soares CN, Vitonis AF, et al. Risk for new onset of depression during the menopausal transition: the Harvard study of moods and cycles. *Arch Gen Psychiatry*. 2006 Apr;63(4):385-90. PMID: 16585467.
  31. Shifren JL, Monz BU, Russo PA, et al. Sexual problems and distress in United States women: prevalence and correlates. *Obstet Gynecol*. 2008 Nov;112(5):970-8. PMID: 18978095.
  32. Myers LS. Methodological review and meta-analysis of sexuality and menopause research. *Neurosci Biobehav Rev*. 1995 Summer;19(2):331-41. PMID: 7630587.
  33. Krychman ML. Vaginal estrogens for the treatment of dyspareunia. *J Sex Med*. 2011 Mar;8(3):666-74. PMID: 21091878.
  34. Suckling J, Lethaby A, Kennedy R. Local oestrogen for vaginal atrophy in postmenopausal women. *Cochrane Database Syst Rev*. 2006(4):CD001500. PMID: 17054136.
  35. National Sleep Foundation. Women and sleep. 2007.  
[http://www.sleepfoundation.org/sites/default/files/Summary\\_of\\_Findings%20-%20FINAL.pdf](http://www.sleepfoundation.org/sites/default/files/Summary_of_Findings%20-%20FINAL.pdf)
  36. Sood R, Warndahl RA, Schroeder DR, et al. Bioidentical compounded hormones: a pharmacokinetic evaluation in a randomized clinical trial. *Maturitas*. 2013 Apr;74(4):375-82. PMID: 23384975.
  37. Nelson HD, Walker M, Zakher B, et al. Menopausal Hormone Therapy for the Primary Prevention of Chronic Conditions: Systematic Review to Update the 2002 and 2005 U.S. Preventive Services Task Force Recommendations. Rockville (MD); 2012.
  38. Miller ER, 3rd, Pastor-Barriuso R, Dalal D, et al. Meta-analysis: high-dosage vitamin E supplementation may increase all-cause mortality. *Ann Intern Med*. 2005 Jan 4;142(1):37-46. PMID: 15537682.
  39. Mahady GB, Low Dog T, Barrett ML, et al. United States Pharmacopeia review of the black cohosh case reports of hepatotoxicity. *Menopause*. 2008 Jul-Aug;15(4 Pt 1):628-38. PMID: 18340277.
  40. Teschke R, Schmidt-Taenzer W, Wolff A. Spontaneous reports of assumed herbal hepatotoxicity by black cohosh: is the liver-unspecific Naranjo scale precise enough to ascertain causality? *Pharmacoepidemiol Drug Saf*. 2011 Jun;20(6):567-82. PMID: 21702069.
  41. Prentice RL. Observational studies, clinical trials, and the women's health initiative. *Lifetime Data Anal*. 2007 Dec;13(4):449-62. PMID: 17943443.
  42. Tanaka S, Matsuyama Y, Shiraki M, et al. Estimating the effects of time-varying treatments: incidence of fractures among postmenopausal Japanese women. *Epidemiology*. 2007 Sep;18(5):529-36. PMID: 17700241.
  43. Hernan MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology*. 2008 Nov;19(6):766-79. PMID: 18854702.
  44. Morois S, Fournier A, Clavel-Chapelon F, et al. Menopausal hormone therapy and risks of colorectal adenomas and cancers in the French E3N prospective cohort: true associations or bias? *Eur J Epidemiol*. 2012 Jun;27(6):439-52. PMID: 22644109.
  45. Salpeter SR, Buckley NS, Liu H, et al. The cost-effectiveness of hormone therapy in younger and older postmenopausal women. *Am J Med*. 2009 Jan;122(1):42-52 e2. PMID: 19114171.
  46. Gotzsche PC, Hrobjartsson A, Maric K, et al. Data extraction errors in meta-analyses that use standardized mean differences. *JAMA*. 2007 Jul 25;298(4):430-7. PMID: 17652297.

47. Amundsen DW, Diers CJ. The age of menopause in classical Greece and Rome. *Human biology*. 1970 Feb;42(1):79-86. PMID: 4910710.
48. Keating NL, Cleary PD, Rossi AS, et al. Use of hormone replacement therapy by postmenopausal women in the United States. *Ann Intern Med*. 1999 Apr 6;130(7):545-53. PMID: 10189323.
49. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA*. 2002 Jul 17;288(3):321-33. PMID: 12117397.
50. Prentice RL, Langer R, Stefanick ML, et al. Combined postmenopausal hormone therapy and cardiovascular disease: toward resolving the discrepancy between observational studies and the Women's Health Initiative clinical trial. *Am J Epidemiol*. 2005 Sep 1;162(5):404-14. PMID: 16033876.
51. Tsai SA, Stefanick ML, Stafford RS. Trends in menopausal hormone therapy use of US office-based physicians, 2000-2009. *Menopause*. 2011 Apr;18(4):385-92. PMID: 21127439.
52. Somboonporn. Testosterone for peri and postmenopausal women. *Cochrane Database Syst Rev*. 2005(4)PMID: 16235365
53. . The 2012 Hormone Therapy Position Statement of The North American Menopause Society. *Menopause: The Journal of The North American Menopause Society*. 2012;19(3):257-71.
54. . Bioidentical hormones. *The Medical letter on drugs and therapeutics*. 2010 May 31;52(1339):43-4. PMID: 20508582.
55. Siyam T, Yuksel N. Abstract: Bioidentical Hormone Therapy: A Survey of Pharmacists Beliefs. *Menopause*. 2011;18(12):1.
56. Iftikhar S, Shuster L, Johnson R, et al. Abstract: Use of Bio-identical Hormones in Patients Presenting to a Women's Health Clinic. *Menopause*. 207;14(6):1.
57. Galson SK. Federal and State Role in Pharmacy Compounding and Reconstitution: Exploring the Right Mix to Protect Patients. 2003. <http://www.fda.gov/NewsEvents/Testimony/ucm115010.htm>.
58. Patsner B. What every scientist should know about the FDA. 2004. [http://www.google.com/url?sa=t&rct=j&q=WHAT%2BEVERY%2BSCIENTIST%2BSHOULD%2BKNOW%2BABOUT%2BFDA&source=web&cd=1&ved=0CCUQFjAA&url=http%3A%2F%2Fwww.bcm.edu%2Fpda%2Findex.cfm%3Fpmid%3D9920&ei=9jmTTtyPCLOpsAKZg5mJCw&usq=AFQjCNEFjV4KhLb\\_8iC8TzbC1AMdJyUKWA&sig2=qtTkOkr8PI0jYivhRZ\\_Vaw](http://www.google.com/url?sa=t&rct=j&q=WHAT%2BEVERY%2BSCIENTIST%2BSHOULD%2BKNOW%2BABOUT%2BFDA&source=web&cd=1&ved=0CCUQFjAA&url=http%3A%2F%2Fwww.bcm.edu%2Fpda%2Findex.cfm%3Fpmid%3D9920&ei=9jmTTtyPCLOpsAKZg5mJCw&usq=AFQjCNEFjV4KhLb_8iC8TzbC1AMdJyUKWA&sig2=qtTkOkr8PI0jYivhRZ_Vaw). Accessed on October 10 2011.
59. Nelson HD, Vesco KK, Haney E, et al. Nonhormonal therapies for menopausal hot flashes: systematic review and meta-analysis. *JAMA*. 2006 May 3;295(17):2057-71. PMID: 16670414.
60. Loprinzi CL, Sloan JA, Perez EA, et al. Phase III evaluation of fluoxetine for treatment of hot flashes. *J Clin Oncol*. 2002 Mar 15;20(6):1578-83. PMID: 11896107.
61. Bair. Use of complementary and alternative medicine during the menopause transition: longitudinal results from the Study of Women's Health Across the Nation. *Menopause*. 2008;15(1):32-43. PMID: 18090037
62. . Estrogen and progestogen use in postmenopausal women: 2010 position statement of The North American Menopause Society. *Menopause*. 2010 Mar;17(2):242-55. PMID: 20154637.
63. Pines A, Sturdee DW, Birkhauser MH, et al. IMS updated recommendations on postmenopausal hormone therapy. *Climacteric*. 2007 Jun;10(3):181-94. PMID: 17487645.
64. Santen RJ, Allred DC, Ardoin SP, et al. Postmenopausal hormone therapy: an Endocrine Society scientific statement. *J Clin Endocrinol Metab*. 2010 Jul;95(7 Suppl 1):s1-s66. PMID: 20566620.
65. . The American College of Obstetricians and Gynecologists Committee Opinion: Compounded Bioidentical Hormones. *Obstetrics & Gynecology*. 2005;106(5):2.
66. U.S. FDA. Compounded Menopausal Hormone Therapy Questions and Answers. US Food and Drug Administration; 2010. <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/PharmacyCompounding/ucm183088.htm#FDAsEnforcementAction>. Accessed on June 18 2012.
67. Szmuilowicz ED, Manson JE, Rossouw JE, et al. Vasomotor symptoms and cardiovascular events in postmenopausal women. *Menopause*. 2011 Jun;18(6):603-10. PMID: 21358352.
68. Loprinzi CL, Diekmann B, Novotny PJ, et al. Newer antidepressants and gabapentin for hot

- flashes: a discussion of trial duration. *Menopause*. 2009 Sep-Oct;16(5):883-7. PMID: 19295449.
69. The Endocrine Society Position Statement: Bioidentical Hormones. 2006. <http://www.endo-society.org/advocacy/policy/upload/BH-Position-Statement-final-10-25-06-w-Header.pdf>. Accessed on June 18 2012.
70. American Association of Clinical Endocrinologists (AACE) Reproductive Medicine Committee Position Statement on Bioidentical Hormones. 2007. <https://www.aace.com/files/AACEBHStatement071507.pdf>. Accessed on June 18 2012.
71. SPECIAL COMMITTEE ON AGING UNITED STATES SENATE. Bioidentical Hormones: Sound Science or Bad Medicine. 2007. [http://aging.senate.gov/hearing\\_detail.cfm?id=321322&](http://aging.senate.gov/hearing_detail.cfm?id=321322&). Accessed on July 21 2011.
72. Somers S. *Ageless: The Naked Truth About Bioidentical Hormones*. 1st ed. New York: Crown Publishers; 2006.
73. Guttuso T, Jr, Kurlan R, et al. Gabapentin's effects on hot flashes in postmenopausal women: a randomized controlled trial. *Obstet Gynecol*. 2003 Feb(2):337-45. PMID: 12576259
74. Prentice RL, Manson JE, Langer RD, et al. Benefits and risks of postmenopausal hormone therapy when it is initiated soon after menopause. *Am J Epidemiol*. 2009 Jul 1;170(1):12-23. PMID: 19468079.
75. Atkins D, Chang S, Gartlehner G, et al. Assessing the Applicability of Studies When Comparing Medical Interventions. *Methods Guide for Comparative Effectiveness Reviews*. AHRQ Publication No. 11-EHC019-EF. January 2011. <http://effectivehealthcare.ahrq.gov/>.
76. Furness S, Roberts H, Marjoribanks J, et al. Hormone therapy in postmenopausal women and risk of endometrial hyperplasia. *Cochrane Database Syst Rev*. 2009(2):CD000402. PMID: 19370558
77. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*. 2007;7:10. PMID: 17302989.
78. Viswanathan. Assessing the risk of bias of individual studies in systematic reviews of health care interventions. *Agency for Healthcare Research and Quality Methods Guide for Comparative Effectiveness Reviews*. AHRQ Publication No. 12-EHC047-EF. 2012. [www.effectivehealthcare.ahrq.gov/](http://www.effectivehealthcare.ahrq.gov/).
79. Hernan MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006 Jul;60(7):578-86. PMID: 16790829.
80. Vandembroucke JP. What is the best evidence for determining harms of medical treatment? *CMAJ*. 2006 Feb 28;174(5):645-6. PMID: 16505461.
81. Thompson S, Ekelund U, Jebb S, et al. A proposed method of bias adjustment for meta-analyses of published observational studies. *Int J Epidemiol*. 2011 Jun;40(3):765-77. PMID: 21186183.
82. da Costa BR, Rutjes AW, Johnston BC, et al. Methods to convert continuous outcomes into odds ratios of treatment response and numbers needed to treat: meta-epidemiological study. *Int J Epidemiol*. 2012 Oct;41(5):1445-59. PMID: 23045205.
83. Furukawa TA. From effect size into number needed to treat. *Lancet*. 1999 May 15;353(9165):1680. PMID: 10335798.
84. Huntley AL, Ernst E. A systematic review of herbal medicinal products for the treatment of menopausal symptoms. *Menopause*. 2003 Sep-Oct;10(5):465-76. PMID: 14501609.
85. Kupperman HS, Wetchler BB, Blatt MH. Contemporary therapy of the menopausal syndrome. *J Am Med Assoc*. 1959 Nov 21;171:1627-37. PMID: 14412840.
86. Morrison MF, Kallan MJ, Ten Have T, et al. Lack of efficacy of estradiol for depression in postmenopausal women: a randomized, controlled trial. *Biol Psychiatry*. 2004 Feb 15;55(4):406-12. PMID: 14960294.
87. Gelfand MM, Moreau M, Ayotte NJ, et al. Clinical assessment and quality of life of postmenopausal women treated with a new intermittent progestogen combination hormone replacement therapy: a placebo-controlled study. *Menopause*. 2003 Jan-Feb;10(1):29-36. PMID: 12544674.
88. Zollner YF, Acquadro C, Schaefer M. Literature review of instruments to assess health-related quality of life during and after menopause. *Qual Life Res*. 2005 Mar;14(2):309-27. PMID: 15892422.
89. Hilditch JR, Lewis J, Peter A, et al. A menopause-specific quality of life questionnaire: development and psychometric properties. *Maturitas*. 2008 Sep-Oct;61(1-2):107-21. PMID: 19434884.

90. Lewis JE, Hilditch JR, Wong CJ. Further psychometric property development of the Menopause-Specific Quality of Life questionnaire and development of a modified version, MENQOL-Intervention questionnaire. *Maturitas*. 2005 Mar 14;50(3):209-21. PMID: 15734602.
91. Wyrwich KW, Fihn SD, Tierney WM, et al. Clinically important changes in health-related quality of life for patients with chronic obstructive pulmonary disease: an expert consensus panel report. *J Gen Intern Med*. 2003 Mar;18(3):196-202. PMID: 12648251.
92. Samsa G, Edelman D, Rothman ML, et al. Determining clinically important differences in health status measures: a general approach with illustration to the Health Utilities Index Mark II. *Pharmacoeconomics*. 1999 Feb;15(2):141-55. PMID: 10351188.
93. Levine DW, Dailey ME, Rockhill B, et al. Validation of the Women's Health Initiative Insomnia Rating Scale in a multicenter controlled clinical trial. *Psychosom Med*. 2005 Jan-Feb;67(1):98-104. PMID: 15673630.
94. DeRogatis LR, Graziottin A, Bitzer J, et al. Clinically relevant changes in sexual desire, satisfying sexual activity and personal distress as measured by the profile of female sexual function, sexual activity log, and personal distress scale in postmenopausal women with hypoactive sexual desire disorder. *J Sex Med*. 2009 Jan;6(1):175-83. PMID: 19170847.
95. Lee JS, Hobden E, Stiell IG, et al. Clinically important change in the visual analog scale after adequate pain control. *Acad Emerg Med*. 2003 Oct;10(10):1128-30. PMID: 14525749.
96. Tashjian RZ, Deloach J, Porucznik CA, et al. Minimal clinically important differences (MCID) and patient acceptable symptomatic state (PASS) for visual analog scales (VAS) measuring pain in patients treated for rotator cuff disease. *J Shoulder Elbow Surg*. 2009 Nov-Dec;18(6):927-32. PMID: 19535272.
97. Atkins. Assessing applicability when comparing medical interventions: AHRQ and the Effective Health Care Program. *J of Clinical Epidemiology*. 2011;64:1198-207. PMID: 21463926
98. Moher D. Reprint - Preferred Reporting Items for Systematic Reviews and Meta-analyses: the PRISMA Statement. *Phys Ther*. 2009;89(9):873-80.
99. Sloan JA, Loprinzi CL, Novotny PJ, et al. Methodologic lessons learned from hot flash studies. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2001 Dec 1;19(23):4280-90. PMID: 11731510.
100. Oktem M, Eroglu D, Karahan HB, et al. Black cohosh and fluoxetine in the treatment of postmenopausal symptoms: a prospective, randomized trial. *Adv Ther*. 2007 Mar-Apr(2):448-61. PMID: 17565936
101. Wiklund IK, Mattsson LA, Lindgren R, et al. Effects of a standardized ginseng extract on quality of life and physiological parameters in symptomatic postmenopausal women: a double-blind, placebo-controlled trial. *Swedish Alternative Medicine Group. Int J Clin Pharmacol Res*. 1999(3):89-99. PMID: 10761538
102. Hartley DE, Elsabagh S, File SE. Gincosan (a combination of Ginkgo biloba and Panax ginseng): the effects on mood and cognition of 6 and 12 weeks' treatment in post-menopausal women. *Nutr Neurosci*. 2004 Oct-Dec(5-6):325-33. PMID: 15682929
103. Rozenberg S, Ylikorkala O, Arrenbrecht S. Comparison of continuous and sequential transdermal progestogen with sequential oral progestogen in postmenopausal women using continuous transdermal estrogen: vasomotor symptoms, bleeding patterns, and serum lipids. *Int J Fertil Womens Med*. 1997;3:76-87. PMID: 9397385
104. Gambacciani M, Spielmann D, Genazzani AR. Efficacy on climacteric symptoms of a continuous combined regimen of 1 mg 17beta-estradiol and trimegestone versus two regimens combining 1 or 2 mg 17beta-estradiol and norethisterone acetate. *Gynecol Endocrinol*. 2005 Aug(2):65-73. PMID: 16294457
105. Benster B, Carey A, Wadsworth F, et al. A double-blind placebo-controlled study to evaluate the effect of progestelle progesterone cream on postmenopausal women. *Menopause Int*. 2009 Jun(2):63-9. PMID: 19465671.
106. Wren BG, Champion SM, Willetts K, et al. Transdermal progesterone and its effect on vasomotor symptoms, blood lipid levels, bone metabolic markers, moods, and quality of life for postmenopausal women. *Menopause*. 2003 Jan-Feb(1):13-8. PMID: 12544672
107. Leonetti HB, Longo S, Anasti JN. Transdermal progesterone cream for vasomotor symptoms and postmenopausal bone loss. *Obstet Gynecol*. 1999 Aug(2):225-8. PMID: 10432132
108. Joffe H, Petrillo L, Viguera A, et al. Eszopiclone improves insomnia and depressive and anxious

- symptoms in perimenopausal and postmenopausal women with hot flashes: a randomized, double-blinded, placebo-controlled crossover trial. *American Journal of Obstetrics & Gynecology*. 2010;202(2):11. PMID: 20035910
109. Clayden JR, Bell JW, Pollard P. Menopausal flushing: double-blind trial of a non-hormonal medication. *Br Med J*. 1974 Mar 9;1(5905):409-12. PMID: 4593554.
110. Odabasi AR, Yuksel H, Demircan SS, et al. A prospective randomized comparative study of the effects of intranasal and transdermal 17 beta-estradiol on postmenopausal symptoms and vaginal cytology. *J Postgrad Med*. 2007 Oct-Dec(4):221-7. PMID: 18097108
111. Serrano D, Mariani L, Mora S, et al. Quality of life assessment in a chemoprevention trial: fenretinide and oral or transdermal HRT. *Maturitas*. 2006 Aug 20(1):69-75. PMID: 16500052
112. Davis SR, Davison SL, Wilson S, et al. Intranasal versus transdermal matrix oestrogen replacement in Australasian women. *Maturitas*. 2005 Jun 16(2):163-71. PMID: 15917157
113. Ozsoy M, Oral B, Ozsoy D. Clinical equivalence of intranasal estradiol and oral estrogens for postmenopausal symptoms. *Int J Gynaecol Obstet*. 2002 Nov(2):143-6. PMID: 12427399
114. Lopes P, Rozenberg S, Graaf J, et al. Aerodiol versus the transdermal route: perspectives for patient preference. *Maturitas*. 2001 Jun 15:S31-9. PMID: 11390122
115. Mattsson LA, Christiansen C, Colau JC, et al. Clinical equivalence of intranasal and oral 17beta-estradiol for postmenopausal symptoms. *Am J Obstet Gynecol*. 2000 Mar(3):545-52. PMID: 10739506
116. Studd JW, McCarthy K, Zamblera D, et al. Efficacy and tolerance of Menorest compared to Premarin in the treatment of postmenopausal women. A randomised, multicentre, double-blind, double-dummy study. *Maturitas*. 1995 Sep(2):105-14. PMID: 8538478
117. Parsey K, Ellman H, Rahman M. Randomised, controlled comparison of transdermal estradiol with oral conjugated estrogens for the relief of hot flushes. *Clinical Drug Investigation*. 2000(4):207-14.
118. Haines CJ, Lam PM, Chung TK, et al. A randomized, double-blind, placebo-controlled study of the effect of a Chinese herbal medicine preparation (Dang Gui Buxue Tang) on menopausal symptoms in Hong Kong Chinese women. *Climacteric*. 2008 Jun(3):244-51. PMID: 18568789
119. Garcia JT, Gonzaga F, Tan D, et al. Use of a multibotanical (Nutrafem) for the relief of menopausal vasomotor symptoms: a double-blind, placebo-controlled study. *Menopause*. 2010 Mar(2):303-8. PMID: 19934777
120. Sluijs CPvd, Bensoussan A, Chang S, et al. A randomized placebo-controlled trial on the effectiveness of an herbal formula to alleviate menopausal vasomotor symptoms. *Menopause*. 2009 Mar-Apr(2):336-44. PMID: 19057416
121. Yang HM, Liao MF, Zhu SY, et al. A randomised, double-blind, placebo-controlled trial on the effect of Pycnogenol on the climacteric syndrome in peri-menopausal women. *Acta Obstet Gynecol Scand*. 2007(8):978-85. PMID: 17653885
122. Mucci M, Carraro C, Mancino P, et al. Soy isoflavones, lactobacilli, Magnolia bark extract, vitamin D3 and calcium. Controlled clinical study in menopause. *Minerva Ginecol*. 2006 Aug(4):323-34. PMID: 16957676
123. Heger M, Ventskovskiy BM, Borzenko I, et al. Efficacy and safety of a special extract of *Rheum rhaponticum* (ERr 731) in perimenopausal women with climacteric complaints: a 12-week randomized, double-blind, placebo-controlled trial. *Menopause*. 2006 Sep-Oct(5):744-59. PMID: 16894335
124. Winther K, Rein E, Hedman C. Femal, a herbal remedy made from pollen extracts, reduces hot flushes and improves quality of life in menopausal women: a randomized, placebo-controlled, parallel study. *Climacteric*. 2005 Jun(2):162-70. PMID: 16096172
125. Davis SR, Briganti EM, Chen RQ, et al. The effects of Chinese medicinal herbs on postmenopausal vasomotor symptoms of Australian women. A randomised controlled trial. *Med J Aust*. 2001 Jan 15(2):68-71. PMID: 11245505
126. Hirata JD, Swiersz LM, Zell B, et al. Does dong quai have estrogenic effects in postmenopausal women? A double-blind, placebo-controlled trial. *Fertil Steril*. 1997 Dec(6):981-6. PMID: 9418683
127. Chenoy R, Hussain S, Tayob Y, et al. Effect of oral gamolenic acid from evening primrose oil on menopausal flushing. *BMJ*. 1994 Feb 19(6927):501-3. PMID: 8136666
128. Hsu CC, Kuo HC, Chang SY, et al. The assessment of efficacy of *Dioscorea alata* for menopausal symptom treatment in Taiwanese women. *Climacteric*. 2011 Feb(1):132-9. PMID: 20653397

129. Auerbach L, Rakus J, Bauer C, et al. Pomegranate seed oil in women with menopausal symptoms: a prospective randomized, placebo-controlled, double-blinded trial. *Menopause*. 2012 Jan 11; PMID: 22240636
130. Chang A, Kwak BY, Yi K, et al. The Effect of Herbal Extract (EstroG-100) on Pre-, Peri- and Post-Menopausal Women: A Randomized Double-blind, Placebo-controlled Study. *Phytother Res*. 2011 Sep 2; PMID: 21887807
131. Xia Y, Zhao Y, Ren M, et al. A randomized double-blind placebo-controlled trial of a Chinese herbal medicine preparation (Jiawei Qing'e Fang) for hot flashes and quality of life in perimenopausal women. *Menopause*. 2012(2):234-44. PMID: 22089177
132. Chung DJ, Kim HY, Park KH, et al. Black cohosh and St. John's wort (GYNO-Plus) for climacteric symptoms. *Yonsei Med J*. 2007 Apr 30(2):289-94. PMID: 17461529
133. Verhoeven MO, Mooren MJvd, Weijer PHvd, et al. Effect of a combination of isoflavones and *Actaea racemosa* Linnaeus on climacteric symptoms in healthy symptomatic perimenopausal women: a 12-week randomized, placebo-controlled, double-blind study. *Menopause*. 2005 Jul-Aug(4):412-20. PMID: 16037756
134. Uebelhack R, Blohmer JU, Graubaum HJ, et al. Black cohosh and St. John's wort for climacteric complaints: a randomized trial. *Obstet Gynecol*. 2006 Feb(2 Pt 1):247-55. PMID: 16449108
135. Die MDv, Burger HG, Bone KM, et al. *Hypericum perforatum* with *Vitex agnus-castus* in menopausal symptoms: a randomized, controlled trial. *Menopause*. 2009 Jan-Feb(1):156-63. PMID: 18791483
136. Barnhart KT, Freeman E, Grisso JA, et al. The effect of dehydroepiandrosterone supplementation to symptomatic perimenopausal women on serum endocrine profiles, lipid parameters, and health-related quality of life. *J Clin Endocrinol Metab*. 1999 Nov(11):3896-902. PMID: 10566625
137. Dodin S, Lemay A, Jacques H, et al. The effects of flaxseed dietary supplement on lipid profile, bone mineral density, and symptoms in menopausal women: a randomized, double-blind, wheat germ placebo-controlled clinical trial. *J Clin Endocrinol Metab*. 2005 Mar(3):1390-7. PMID: 15613422
138. Andrikoula M, Baker D, Nestic J, et al. The effects of micronutrient supplementation on vasomotor symptoms in postmenopausal women. *Climacteric*. 2011(5):544-50. PMID: 21413865
139. Chandeying V, Sangthawan M. Efficacy comparison of *Pueraria mirifica* (PM) against conjugated equine estrogen (CEE) with/without medroxyprogesterone acetate (MPA) in the treatment of climacteric symptoms in perimenopausal women: phase III study. *J Med Assoc Thai*. 2007 Sep(9):1720-6. PMID: 17957910
140. Virojchaiwong P, Suvithayasiri V, Itharat A. Comparison of *Pueraria mirifica* 25 and 50 mg for menopausal symptoms. *Archives of Gynecology and Obstetrics*. 2011(2):411-9. PMID: 20872225
141. Agosta C, Atlante M, Benvenuti C. Randomized controlled study on clinical efficacy of isoflavones plus *Lactobacillus sporogenes*, associated or not with a natural anxiolytic agent in menopause. *Minerva Ginecol*. 2011 Feb(1):11-7. PMID: 21311416
142. Zervoudis SI, G.; Navrozoglou, I.; Tsonis, C.; Chaniotis, D.; Diakakis, I. Soja Isoflavones with *Lactobacillus Sporogenes* in Recent Menopause Women: Effects on Clinical Symptoms and Bone Density: A Comparative Prospective Study of 62 Patients. 12th World Congress on Menopause; 2008 Madrid, Spain. *Climacteric*; 11. pp. 223 (Abstract PS-32).
143. Hedrick RE, Ackerman RT, Koltun WD, et al. Transdermal estradiol gel 0.1% for the treatment of vasomotor symptoms in postmenopausal women. *Menopause*. 2009 Jan-Feb(1):132-40. PMID: 18971794
144. Akhila V, Pratakumar. A comparison of transdermal and oral HRT for menopausal symptom control. *Int J Fertil Womens Med*. 2006 Mar-Apr(2):64-9. PMID: 16881381
145. Diem S, Grady D, Quan J, et al. Effects of ultralow-dose transdermal estradiol on postmenopausal symptoms in women aged 60 to 80 years. *Menopause*. 2006 Jan-Feb(1):130-8. PMID: 16607109
146. Hidalgo MJC, Castelo-Branco C, Blumel JE, et al. Effect of a compound containing isoflavones, primrose oil and vitamin E in two different doses on climacteric symptoms. *J Obstet Gynaecol*. 2006 May(4):344-7. PMID: 16753687
147. Kerwin J, Gordon P, Senf J. The variable response of women with menopausal hot flashes when treated with sertraline. *Menopause: The Journal of The North American Menopause Society*. 2007;14(5):5. PMID: 17413648

148. Plotnikoff GA, Watanabe K, Torkelson C, et al. The TU-025 keishibukuryogan clinical trial for hot flash management in postmenopausal women: Results and lessons for future research. *Menopause*. 2011(8):886-92. PMID: 21738077
149. Baksu B, Baksu A, Goker N, et al. Do different delivery systems of hormone therapy have different effects on psychological symptoms in surgically menopausal women? A randomized controlled trial. *Maturitas*. 2009 Feb 20(2):140-5. PMID: 19179026
150. Hidalgo LA, Chedraui PA, Morocho N, et al. The effect of red clover isoflavones on menopausal symptoms, lipids and vaginal cytology in menopausal women: a randomized, double-blind, placebo-controlled study. *Gynecol Endocrinol*. 2005 Nov(5):257-64. PMID: 16373244
151. Han KK, Soares JM, Jr, et al. Benefits of soy isoflavone therapeutic regimen on menopausal symptoms. *Obstet Gynecol*. 2002 Mar(3):389-94. PMID: 11864664
152. Osmers R, Friede M, Liske E, et al. Efficacy and safety of isopropanolic black cohosh extract for climacteric symptoms. *Obstet Gynecol*. 2005 May(5 Pt 1):1074-83. PMID: 15863547
153. Brunner RL, Gass M, Aragaki A, et al. Effects of conjugated equine estrogen on health-related quality of life in postmenopausal women with hysterectomy: results from the Women's Health Initiative Randomized Clinical Trial. *Arch Intern Med*. 2005 Sep 26(17):1976-86. PMID: 16186467
154. Hays J, Ockene JK, Brunner RL, et al. Effects of estrogen plus progestin on health-related quality of life. *N Engl J Med*. 2003 May 8(19):1839-54. PMID: 12642637
155. Panjari M, Bell RJ, Jane F, et al. A randomized trial of oral DHEA treatment for sexual function, well-being, and menopausal symptoms in postmenopausal women with low libido. *J Sex Med*. 2009 Sep(9):2579-90. PMID: 19619146
156. Labrie F, Archer D, Bouchard C, et al. Effect of intravaginal dehydroepiandrosterone (Prasterone) on libido and sexual dysfunction in postmenopausal women. *Menopause*. 2009 Sep-Oct(5):923-31. PMID: 19424093
157. Lubbert H, Nauert C. Continuous versus cyclical transdermal estrogen replacement therapy in postmenopausal women: influence on climacteric symptoms, body weight and bleeding pattern. *Maturitas*. 1997 Dec 15(2):117-25. PMID: 9522319
158. Polvani F, Zichella L, Bocci A, et al. A randomized comparative study for the clinical evaluation of hormone replacement by transdermal and oral routes. *Clin Exp Obstet Gynecol*. 1991(4):207-13. PMID: 1790601
159. Soares CN, Almeida OP, Joffe H, et al. Efficacy of estradiol for the treatment of depressive disorders in perimenopausal women: a double-blind, randomized, placebo-controlled trial. *Arch Gen Psychiatry*. 2001 Jun(6):529-34. PMID: 11386980
160. Lewis JE, Nickell LA, Thompson LU, et al. A randomized controlled trial of the effect of dietary soy and flaxseed muffins on quality of life and hot flashes during menopause. *Menopause*. 2006 Jul-Aug(4):631-42. PMID: 16837885
161. Veerus P, Fischer K, Hovi SL, et al. Symptom reporting and quality of life in the Estonian Postmenopausal Hormone Therapy Trial. *BMC Womens Health*. 2008;5. PMID: 18366766
162. Frei-Kleiner S, Schaffner W, Rahlfs VW, et al. Cimicifuga racemosa dried ethanolic extract in menopausal disorders: a double-blind placebo-controlled clinical trial. *Maturitas*. 2005 Aug 16(4):397-404. PMID: 16039414
163. Liske E, Hanggi W, Zepelin HHH-v, et al. Physiological investigation of a unique extract of black cohosh (*Cimicifuga racemosa* rhizoma): a 6-month clinical study demonstrates no systemic estrogenic effect. *J Womens Health Gend Based Med*. 2002 Mar(2):163-74. PMID: 11975864
164. Sammartino A, Tommaselli GA, Gargano V, et al. Short-term effects of a combination of isoflavones, lignans and *Cimicifuga racemosa* on climacteric-related symptoms in postmenopausal women: a double-blind, randomized, placebo-controlled trial. *Gynecol Endocrinol*. 2006 Nov(11):646-50. PMID: 17145651
165. Sammartino A, Carlo CD, Mandato VD, et al. Effects of genistein on the endometrium: ultrasonographic evaluation. *Gynecol Endocrinol*. 2003 Feb(1):45-9. PMID: 12724018
166. JG G. Constructing a standard climacteric scale. *Maturitas*. 1998;29:25-31. PMID: 9643514
167. Beck. An inventory for measuring clinical anxiety: psychometric properties. *J Consult Clin Psychol*. 1988;56:893-7. PMID: 3204199
168. Beck. Beck Depression Inventory. *Archives of General Psychiatry*. 1961;4(6):567-71.
169. Hamilton. The assessment of anxiety states by rating. *British Medical Journal*. 1959;32:50-5.
170. Hamilton. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23(1):56-62.

171. Hays. The RAND-36 measure of health-related quality of life. *Ann Med.* 2001;33:350-7. PMID: 11491194
172. Kupperman HS. Comparative clinical evaluation of estrogenic preparations by the menopausal and amenorrheal indices. *J Clin Endocrinol Metab.* 1953;13:688-703. PMID: 13061588
173. Campbell, S, Whitehead, et al. Oestrogen therapy and the menopausal syndrome. *Clin Obstet Gynaecol.* 1977 Apr(1):31-47. PMID: 322905
174. Speroff L. Efficacy and tolerability of a novel estradiol vaginal ring for relief of menopausal symptoms. *Obstet Gynecol.* 2003 Oct(4):823-34. PMID: 14551014
175. Rudolph I, Palombo-Kinne E, Kirsch B, et al. Influence of a continuous combined HRT (2 mg estradiol valerate and 2 mg dienogest) on postmenopausal depression. *Climacteric.* 2004 Sep(3):301-11. PMID: 15669555
176. Wiklund I, Karlberg J, Mattsson LA. Quality of life of postmenopausal women on a regimen of transdermal estradiol therapy: a double-blind placebo-controlled study. *Am J Obstet Gynecol.* 1993 Mar(3 Pt 1):824-30. PMID: 8456888
177. Derman RJ, Dawood MY, Stone S. Quality of life during sequential hormone replacement therapy - a placebo-controlled study. *Int J Fertil Menopausal Stud.* 1995 Mar-Apr(2):73-8. PMID: 7599662
178. Saletu B, Brandstatter N, Metka M, et al. Double-blind, placebo-controlled, hormonal, syndromal and EEG mapping studies with transdermal oestradiol therapy in menopausal depression. *Psychopharmacology (Berl).* 1995 Dec(4):321-9. PMID: 8657828
179. Bech P, Munk-Jensen N, Obel EB, et al. Combined versus sequential hormonal replacement therapy: a double-blind, placebo-controlled study on quality of life-related outcome measures. *Psychother Psychosom.* 1998 Jul-Oct(4-5):259-65. PMID: 9693354
180. Polo-Kantola P, Erkkola R, Irjala K, et al. Effect of short-term transdermal estrogen replacement therapy on sleep: a randomized, double-blind crossover trial in postmenopausal women. *Fertil Steril.* 1999 May(5):873-80. PMID: 10231049
181. Yang TS, Wang HL, Chen YJ, et al. Effect of continuous administration of conjugated estrogen plus medroxyprogesterone acetate (Premelle) in postmenopausal women in Taiwan. *J Chin Med Assoc.* 2004 Jul(7):336-43. PMID: 15510930
182. Onalan G, Onalan R, Selam B, et al. Mood scores in relation to hormone replacement therapies during menopause: a prospective randomized trial. *Tohoku J Exp Med.* 2005 Nov(3):223-31. PMID: 16210834
183. Hassa H, Tanir HM, Oge T. Is placebo as effective as estrogen regimens on vasomotor symptoms in women with surgical menopause? *Clin Exp Obstet Gynecol.* 2010(2):135-7. PMID: 21077506
184. Demetrio FN, J. Renno J, Gianfaldoni A, et al. Effect of estrogen replacement therapy on symptoms of depression and anxiety in non-depressive menopausal women : A randomized double-blind, controlled study. *Archives of Women's Mental Health.* 2011(6):479-86. PMID: 22016254
185. Strickler R, Stovall DW, Merritt D, et al. Raloxifene and estrogen effects on quality of life in healthy postmenopausal women: a placebo-controlled randomized trial. *Obstet Gynecol.* 2000 Sep(3):359-65. PMID: 10960626
186. Rozenbaum H, Chevallier O, Moyal M, et al. Efficacy and tolerability of pulsed estrogen therapy: a 12-week double-blind placebo-controlled study in highly symptomatic postmenopausal women. *Climacteric.* 2002 Sep(3):249-58. PMID: 12419083
187. Haines CJ, Yim SF, Chung TK, et al. A prospective, randomized, placebo-controlled study of the dose effect of oral oestradiol on menopausal symptoms, psychological well being, and quality of life in postmenopausal Chinese women. *Maturitas.* 2003 Mar 28(3):207-14. PMID: 12648884
188. Nielsen TF, Ravn P, Pitkin J, et al. Pulsed estrogen therapy improves postmenopausal quality of life: a 2-year placebo-controlled study. *Maturitas.* 2006 Jan 20(2):184-90. PMID: 16368471
189. Coope J. Is oestrogen therapy effective in the treatment of menopausal depression? *J R Coll Gen Pract.* 1981 Mar(224):134-40. PMID: 6268783
190. Gambacciani M, Ciapponi M, Cappagli B, et al. Effects of low-dose, continuous combined estradiol and noretisterone acetate on menopausal quality of life in early postmenopausal women. *Maturitas.* 2003 Feb 25(2):157-63. PMID: 12590012
191. Pefanco MA, Kenny AM, Kaplan RF, et al. The effect of 3-year treatment with 0.25 mg/day of micronized 17beta-estradiol on cognitive function in older postmenopausal women. *J Am Geriatr Soc.* 2007 Mar(3):426-31. PMID: 17341247
192. Baerug U, Winge T, Nordland G, et al. Do combinations of 1 mg estradiol and low doses of

- NETA effectively control menopausal symptoms? *Climacteric*. 1998 Sep(3):219-28. PMID: 11907946
193. Yaffe K, Vittinghoff E, Ensrud KE, et al. Effects of ultra-low-dose transdermal estradiol on cognition and health-related quality of life. *Arch Neurol*. 2006 Jul(7):945-50. PMID: 16831962
194. Simon JA, Bouchard C, Waldbaum A, et al. Low dose of transdermal estradiol gel for treatment of symptomatic postmenopausal women: a randomized controlled trial. *Obstet Gynecol*. 2007 Mar(3):588-96. PMID: 17329509
195. Kotsopoulos D, Dalais FS, Liang YL, et al. The effects of soy protein containing phytoestrogens on menopausal symptoms in postmenopausal women. *Climacteric*. 2000 Sep(3):161-7. PMID: 11910617
196. Tice JA, Ettinger B, Ensrud K, et al. Phytoestrogen supplements for the treatment of hot flashes: the Isoflavone Clover Extract (ICE) Study: a randomized controlled trial. *JAMA*. 2003 Jul 9(2):207-14. PMID: 12851275
197. Kok L, Kreijkamp-Kaspers S, Grobbee DE, et al. A randomized, placebo-controlled trial on the effects of soy protein containing isoflavones on quality of life in postmenopausal women. *Menopause*. 2005 Jan-Feb(1):56-62. PMID: 15668601
198. Sousa-Munoz RLd, Filizola RG. Efficacy of soy isoflavones for depressive symptoms of the climacteric syndrome. *Maturitas*. 2009 May 20(1):89-93. PMID: 19339127
199. Lipovac M, Chedraui P, Gruenhut C, et al. Improvement of postmenopausal depressive and anxiety symptoms after treatment with isoflavones derived from red clover extracts. *Maturitas*. 2010 Mar(3):258-61. PMID: 19948385
200. Jassi HK, Jain A, Arora S, et al. Effect of soy proteins Vs soy isoflavones on lipid profile in postmenopausal women. *Indian Journal of Clinical Biochemistry*. 2010(2):201-7. PMID: 23105910
201. Evans M, Elliott JG, Sharma P, et al. The effect of synthetic genistein on menopause symptom management in healthy postmenopausal women: a multi-center, randomized, placebo-controlled study. *Maturitas*. 2011 Feb(2):189-96. PMID: 21163595
202. Casini ML, Marelli G, Papaleo E, et al. Psychological assessment of the effects of treatment with phytoestrogens on postmenopausal women: a randomized, double-blind, crossover, placebo-controlled study. *Fertil Steril*. 2006 Apr(4):972-8. PMID: 16580383
203. Ho SC, Chan AS, Ho YP, et al. Effects of soy isoflavone supplementation on cognitive function in Chinese postmenopausal women: a double-blind, randomized, controlled trial. *Menopause*. 2007 May-Jun(3 Pt 1):489-99. PMID: 17308499
204. Basaria S, Wisniewski A, Dupree K, et al. Effect of high-dose isoflavones on cognition, quality of life, androgens, and lipoprotein in post-menopausal women. *J Endocrinol Invest*. 2009 Feb(2):150-5. PMID: 19411814
205. Suvanto-Luukkonen E, Koivunen R, Sundstrom H, et al. Citalopram and fluoxetine in the treatment of postmenopausal symptoms: a prospective, randomized, 9-month, placebo-controlled, double-blind study. *Menopause*. 2005 Jan-Feb(1):18-26. PMID: 15668596
206. Kornstein SG, Jiang Q, Reddy S, et al. Short-term efficacy and safety of desvenlafaxine in a randomized, placebo-controlled study of perimenopausal and postmenopausal women with major depressive disorder. *J Clin Psychiatry*. 2010 Aug;71(8):1088-96. PMID: 20797382.
207. Pfizer. Study Evaluating The Safety And Efficacy Of Desvenlafaxine Succinate For Vasomotor Symptoms In Menopausal Women. NCT00683800. 2011.
208. Evans ML, Pritts E, Vittinghoff E, et al. Management of postmenopausal hot flashes with venlafaxine hydrochloride: a randomized, controlled trial. *Obstet Gynecol*. 2005 Jan(1):161-6. PMID: 15625158
209. Kalay A, Demir B, Haberal A, et al. Efficacy of citalopram on climacteric symptoms. *Menopause: The Journal of The North American Menopause Society*. 2007;14(2):7. PMID: 17224858
210. Bouchard P, Panay N, Villiers TJD, et al. Randomized placebo- and active-controlled study of desvenlafaxine for menopausal vasomotor symptoms. *Climacteric*. 2012(1):12-20. PMID: 22066790
211. Butt D, Deng L, Lewis J, et al. Minimal decrease in hot flashes desired by postmenopausal women in family practice. *Menopause: The Journal of The North American Menopause Society*. 2007;14(2):5. PMID: 17099324
212. Nappi RE, Malavasi B, Brundu B, et al. Efficacy of *Cimicifuga racemosa* on climacteric complaints: a randomized study versus low-dose transdermal estradiol. *Gynecol Endocrinol*. 2005 Jan(1):30-5. PMID: 15969244
213. Buckler H, Al-Azzawi F. The effect of a novel vaginal ring delivering oestradiol acetate on

- climacteric symptoms in postmenopausal women. *BJOG*. 2003 Aug(8):753-9. PMID: 12892687
214. Nathorst-Boos J, Floter A, Jarkander-Rolff M, et al. Treatment with percutaneous testosterone gel in postmenopausal women with decreased libido--effects on sexuality and psychological general well-being. *Maturitas*. 2006 Jan 10(1):11-8. PMID: 16183220
215. Shen CL, Chyu MC, Pence BC, et al. Green tea polyphenols supplementation and Tai Chi exercise for postmenopausal osteopenic women: safety and quality of life report. *BMC Complement Altern Med*. 2010;76. PMID: 21143878
216. Soares CN. Desvenlafaxine and escitalopram for the treatment of postmenopausal women with major depressive disorder. *Menopause*. 2010;17(4):700-11. PMID: 20539246
217. Pitkin J, Smetnik VP, Vadasz P, et al. Continuous combined hormone replacement therapy relieves climacteric symptoms and improves health-related quality of life in early postmenopausal women. *Menopause Int*. 2007 Sep(3):116-23. PMID: 17785037
218. Bachmann G, Bouchard C, Hoppe D, et al. Efficacy and safety of low-dose regimens of conjugated estrogens cream administered vaginally. *Menopause*. 2009 Jul-Aug(4):719-27. PMID: 19436223
219. Freedman M, Kaunitz AM, Reape KZ, et al. Twice-weekly synthetic conjugated estrogens vaginal cream for the treatment of vaginal atrophy. *Menopause*. 2009 Jul-Aug(4):735-41. PMID: 19252451
220. Gast MJ, Freedman MA, Vieweg AJ, et al. A randomized study of low-dose conjugated estrogens on sexual function and quality of life in postmenopausal women. *Menopause*. 2009 Mar-Apr(2):247-56. PMID: 19034054
221. Dessole S, Rubattu G, Ambrosini G, et al. Efficacy of low-dose intravaginal estriol on urogenital aging in postmenopausal women. *Menopause*. 2004 Jan-Feb(1):49-56. PMID: 14716182
222. Simunic V, Banovic I, Ciglar S, et al. Local estrogen treatment in patients with urogenital symptoms. *Int J Gynaecol Obstet*. 2003 Aug(2):187-97. PMID: 12873780
223. Parsons A, Merritt D, Rosen A, et al. Effect of raloxifene on the response to conjugated estrogen vaginal cream or nonhormonal moisturizers in postmenopausal vaginal atrophy. *Obstet Gynecol*. 2003 Feb(2):346-52. PMID: 12576260
224. Casper F, Petri E. Local treatment of urogenital atrophy with an estradiol-releasing vaginal ring: a comparative and a placebo-controlled multicenter study. *Vaginal Ring Study Group. Int Urogynecol J Pelvic Floor Dysfunct*. 1999(3):171-6. PMID: 10430010
225. Eriksen PS, Rasmussen H. Low-dose 17 beta-estradiol vaginal tablets in the treatment of atrophic vaginitis: a double-blind placebo controlled study. *Eur J Obstet Gynecol Reprod Biol*. 1992 Apr 21(2):137-44. PMID: 1587379
226. Vestergaard P, Hermann AP, Stilgren L, et al. Effects of 5 years of hormonal replacement therapy on menopausal symptoms and blood pressure--a randomised controlled study. *Maturitas*. 2003 Oct 20(2):123-32. PMID: 14559383
227. Simon JA, Reape KZ, Wininger S, et al. Randomized, multicenter, double-blind, placebo-controlled trial to evaluate the efficacy and safety of synthetic conjugated estrogens B for the treatment of vulvovaginal atrophy in healthy postmenopausal women. *Fertil Steril*. 2008 Oct;90(4):1132-8. PMID: 18053998.
228. Osmanagaoglu MA, Atasarl T, Baltaci D, et al. Effect of different preparations of hormone therapy on sexual dysfunction in naturally postmenopausal women. *Climacteric*. 2006 Dec(6):464-72. PMID: 17085380
229. Haines C, Yu SL, Hiemeyer F, et al. Micro-dose transdermal estradiol for relief of hot flashes in postmenopausal Asian women: a randomized controlled trial. *Climacteric*. 2009 Oct(5):419-26. PMID: 19479489
230. Huang A, Yaffe K, Vittinghoff E, et al. The effect of ultralow-dose transdermal estradiol on sexual function in postmenopausal women. *Am J Obstet Gynecol*. 2008 Mar(3):265 e1-7. PMID: 18313446
231. Raghunandan C, Agrawal S, Dubey P, et al. A comparative study of the effects of local estrogen with or without local testosterone on vulvovaginal and sexual dysfunction in postmenopausal women. *J Sex Med*. 2010 Mar(3):1284-90. PMID: 20102444
232. Davis SR, Mooren MJvd, Lunsen RHv, et al. Efficacy and safety of a testosterone patch for the treatment of hypoactive sexual desire disorder in surgically menopausal women: a randomized, placebo-controlled trial. *Menopause*. 2006 May-Jun(3):387-96. PMID: 16735935

233. Braunstein GD, Sundwall DA, Katz M, et al. Safety and efficacy of a testosterone patch for the treatment of hypoactive sexual desire disorder in surgically menopausal women: a randomized, placebo-controlled trial. *Arch Intern Med.* 2005 Jul 25(14):1582-9. PMID: 16043675
234. Simon J, Braunstein G, Nachtigall L, et al. Testosterone patch increases sexual activity and desire in surgically menopausal women with hypoactive sexual desire disorder. *J Clin Endocrinol Metab.* 2005 Sep(9):5226-33. PMID: 16014407
235. Buster JE, Kingsberg SA, Aguirre O, et al. Testosterone patch for low sexual desire in surgically menopausal women: a randomized trial. *Obstet Gynecol.* 2005 May(5 Pt 1):944-52. PMID: 15863529
236. Panay N, Al-Azzawi F, Bouchard C, et al. Testosterone treatment of HSDD in naturally menopausal women: the ADORE study. *Climacteric.* 2010 Apr(2):121-31. PMID: 20166859
237. Penteado SR, Fonseca AM, Bagnoli VR, et al. Effects of the addition of methyltestosterone to combined hormone therapy with estrogens and progestogens on sexual energy and on orgasm in postmenopausal women. *Climacteric.* 2008 Feb(1):17-25. PMID: 18202961
238. Davis SR, Moreau M, Kroll R, et al. Testosterone for low libido in postmenopausal women not taking estrogen. *N Engl J Med.* 2008 Nov 6(19):2005-17. PMID: 18987368
239. Shifren JL, Davis SR, Moreau M, et al. Testosterone patch for the treatment of hypoactive sexual desire disorder in naturally menopausal women: results from the INTIMATE NM1 Study. *Menopause.* 2006 Sep-Oct(5):770-9. PMID: 16932240
240. Cieraad D, Conrads C, Jesinger D, et al. Clinical study comparing the effects of sequential hormone replacement therapy with oestradiol/dydrogesterone and conjugated equine oestrogen/norgestrel on lipids and symptoms. *Arch Gynecol Obstet.* 2006 May(2):74-80. PMID: 16491367
241. Utian WH, Speroff L, Ellman H, et al. Comparative controlled trial of a novel oral estrogen therapy, estradiol acetate, for relief of menopause symptoms. *Menopause.* 2005 Nov-Dec(6):708-15. PMID: 16278614
242. Loh FH, Chen LH, Yu SL, et al. The efficacy of two dosages of a continuous combined hormone replacement regimen. *Maturitas.* 2002 Feb 26(2):123-31. PMID: 11836043
243. Limpaphayom KK, Darmasetiawan MS, Hussain RI, et al. Differential prevalence of quality-of-life categories (domains) in Asian women and changes after therapy with three doses of conjugated estrogens/medroxyprogesterone acetate: the Pan-Asia Menopause (PAM) study. *Climacteric.* 2006 Jun(3):204-14. PMID: 16766434
244. Barentsen R, Weijer PHvd, Schram JH. Continuous low dose estradiol released from a vaginal ring versus estradiol vaginal cream for urogenital atrophy. *Eur J Obstet Gynecol Reprod Biol.* 1997 Jan(1):73-80. PMID: 9031963
245. Ayton RA, Darling GM, Murkies AL, et al. A comparative study of safety and efficacy of continuous low dose oestradiol released from a vaginal ring compared with conjugated equine oestrogen vaginal cream in the treatment of postmenopausal urogenital atrophy. *Br J Obstet Gynaecol.* 1996 Apr(4):351-8. PMID: 8605133
246. Hilditch JR, Lewis J, Ross AH, et al. A comparison of the effects of oral conjugated equine estrogen and transdermal estradiol-17 beta combined with an oral progestin on quality of life in postmenopausal women. *Maturitas.* 1996 Jul(3):177-84. PMID: 8844631
247. Weisberg E, Ayton R, Darling G, et al. Endometrial and vaginal effects of low-dose estradiol delivered by vaginal ring or vaginal tablet. *Climacteric.* 2005 Mar(1):83-92. PMID: 15804736
248. Long CY, Liu CM, Hsu SC, et al. A randomized comparative study of the effects of oral and topical estrogen therapy on the vaginal vascularization and sexual function in hysterectomized postmenopausal women. *Menopause.* 2006 Sep-Oct(5):737-43. PMID: 16946685
249. Henriksson L, Stjernquist M, Boquist L, et al. A comparative multicenter study of the effects of continuous low-dose estradiol released from a new vaginal ring versus estradiol vaginal pessaries in postmenopausal women with symptoms and signs of urogenital atrophy. *Am J Obstet Gynecol.* 1994 Sep(3):624-32. PMID: 8092207
250. Lobo RA, Rosen RC, Yang HM, et al. Comparative effects of oral esterified estrogens with and without methyltestosterone on endocrine profiles and dimensions of sexual function in postmenopausal women with hypoactive sexual desire. *Fertil Steril.* 2003 Jun(6):1341-52. PMID: 12798881
251. Foidart JM, Vervliet J, Buytaert P. Efficacy of sustained-release vaginal oestradiol in alleviating urogenital and systemic climacteric complaints. *Maturitas.* 1991 Jun(2):99-107. PMID: 1921739

252. Bachmann G, Lobo RA, Gut R, et al. Efficacy of low-dose estradiol vaginal tablets in the treatment of atrophic vaginitis: a randomized controlled trial. *Obstet Gynecol*. 2008 Jan(1):67-76. PMID: 18165394
253. Chung TK, Yip SK, Lam P, et al. A randomized, double-blind, placebo-controlled, crossover study on the effect of oral oestradiol on acute menopausal symptoms. *Maturitas*. 1996 Oct(2):115-23. PMID: 8905602
254. Holst Tv, Salbach B. Efficacy and tolerability of a new 7-day transdermal estradiol patch versus placebo in hysterectomized women with postmenopausal complaints. *Maturitas*. 2000 Feb 15(2):143-53. PMID: 10714909
255. Schurmann R, Holler T, Benda N. Estradiol and drospirenone for climacteric symptoms in postmenopausal women: a double-blind, randomized, placebo-controlled study of the safety and efficacy of three dose regimens. *Climacteric*. 2004 Jun(2):189-96. PMID: 15497908
256. Lee BS, Kang BM, Yoon BK, et al. Efficacy and tolerability of estradiol 1 mg and drospirenone 2 mg in postmenopausal Korean women: a double-blind, randomized, placebo-controlled, multicenter study. *Maturitas*. 2007 Aug 20(4):361-9. PMID: 17467203
257. Lin SQ, Sun LZ, Lin JF, et al. Estradiol 1mg and drospirenone 2mg as hormone replacement therapy in postmenopausal Chinese women. *Climacteric*. 2011(4):472-81. PMID: 21469973
258. Lee J, Kim KW, Kim HK, et al. The effect of Rexflavone (*Sophorae fructus* extract) on menopausal symptoms in postmenopausal women: a randomized double-blind placebo controlled clinical trial. *Arch Pharm Res*. 2010 Apr(4):523-30. PMID: 20422360
259. Cianci A, Cicero AF, Colacurci N, et al. Activity of isoflavones and berberine on vasomotor symptoms and lipid profile in menopausal women. *Gynecol Endocrinol*. 2012 Feb 8 PMID: 22313171
260. Dugal R, Hesla K, Sordal T, et al. Comparison of usefulness of estradiol vaginal tablets and estriol vagitories for treatment of vaginal atrophy. *Acta Obstet Gynecol Scand*. 2000 Apr(4):293-7. PMID: 10746845
261. Shulman LP, Yankov V, Uhl K. Safety and efficacy of a continuous once-a-week 17beta-estradiol/levonorgestrel transdermal system and its effects on vasomotor symptoms and endometrial safety in postmenopausal women: the results of two multicenter, double-blind, randomized, controlled trials. *Menopause*. 2002 May-Jun(3):195-207. PMID: 11973443
262. Donne ML, Caruso C, Mancuso A, et al. The effect of vaginally administered genistein in comparison with hyaluronic acid on atrophic epithelium in postmenopause. *Arch Gynecol Obstet*. 2011 Jun(6):1319-23. PMID: 20577750
263. Al-Azzawi F, Buckler HM. Comparison of a novel vaginal ring delivering estradiol acetate versus oral estradiol for relief of vasomotor menopausal symptoms. *Climacteric*. 2003 Jun(2):118-27. PMID: 12841882
264. MS H. The Women's Health Questionnaire (WHQ): frequently asked questions (FAQ). *Health and Quality of Life Outcomes*. 2003;1(41):1-5. PMID: 14521718
265. Panay N, Ylikorkala O, Archer DF, et al. Ultra-low-dose estradiol and norethisterone acetate: effective menopausal symptom relief. *Climacteric*. 2007 Apr(2):120-31. PMID: 17453860
266. Pornel B, Spielmann D. A study of the control of climacteric symptoms in postmenopausal women following sequential regimens of 1 mg 17beta-estradiol and trimegestone compared with a regimen containing 1 mg estradiol valerate and norethisterone over a two-year period. *Gynecol Endocrinol*. 2005 Aug(2):74-81. PMID: 16294458
267. Sood R, Shuster L, Schroeder D, et al. Abstract: Bioidentical Compounded Hormones - A pharmacokinetic evaluation in the setting of a randomized controlled trial. *Menopause*. 2010;17(6):1.
268. White CM, Ip S, McPheeters M, et al. Using Existing Systematic Reviews To Replace De Novo Processes in Conducting Comparative Effectiveness Reviews. *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. Rockville (MD); 2008.
269. Greiser CM, Greiser EM, Doren M. Menopausal hormone therapy and risk of ovarian cancer: systematic review and meta-analysis. *Hum Reprod Update*. 2007 Sep-Oct(5):453-63. PMID: 17573406
270. Chlebowski RT, Anderson GL, Gass M, et al. Estrogen plus progestin and breast cancer incidence and mortality in postmenopausal women. *JAMA*. 2010 Oct 20(15):1684-92. PMID: 20959578
271. Cirillo. Effect of estrogen therapy on gallbladder disease. *JAMA*. 2005;293:330-9. PMID: 15657326
272. Heiss. Health risks and benefits 3 years after stopping randomized treatment with estrogen and

- progesterin. *JAMA*. 2008;299(9):1036-45. PMID: 18319414
273. Anderson GL. Effects of estrogen and progesterin on gynecologic cancers and associated diagnostic procedures: the Women's Health Initiative randomized trial. *JAMA*. 2003;290(13):1739-48. PMID: 14519708
274. Anderson GL. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *JAMA*. 2004;291(14):1701-12. PMID: 15082697
275. LaCroix AZ, Chlebowski RT, Manson JE, et al. Health outcomes after stopping conjugated equine estrogens among postmenopausal women with prior hysterectomy: a randomized controlled trial. *JAMA*. 2011 Apr 6(13):1305-14. PMID: 21467283
276. Shumaker SA. Estrogen plus progesterin and the incidence of dementia and mild cognitive impairment in postmenopausal women: the Women's Health Initiative Memory Study: a randomized controlled trial. *JAMA*. 2003;289(20):2651-62. PMID: 12771112
277. Espeland. Long-term effects of conjugated equine estrogen therapies on domain-specific cognitive function: results from the Women's Health Initiative Study of Cognitive Aging extension. *J Am Geriatr Soc*. 2010;58(7):1263-71.
278. Grady D. The Heart and Estrogen/progesterin Replacement Study: design, methods, and baseline characteristics. *Control Clin Trials*. 1998;19:314-35. PMID: 9683309
279. Hulley. Noncardiovascular disease outcomes during 6.8 years of hormone therapy: Heart and Estrogen/progesterin Replacement Study follow-up (HERS II). *JAMA*. 2002;288(1):58-66. PMID: 12090863
280. Vickers MR. The Women's International Study of Long-Duration Oestrogen After Menopause (WISDOM): a randomized controlled trial. *BMC Womens Health*. 2007;7:2. PMID: 17324282
281. Cherry. Oestrogen therapy for prevention of reinfarction in postmenopausal women: a randomized placebo controlled trial. *Lancet*. 2002;360(9350):2001-8. PMID: 12504395
282. Tierney. A randomized double-blind trial of the effects of hormone therapy on delayed verbal recall in older women. *Psychoneuroendocrinology*. 2009;34(7):1065-74. PMID: 19297102
283. Ettinger B. Effects of ultra-low dose transdermal estradiol on bone mineral density: a randomized clinical trial. *Obstet Gynecol*. 2004;2004(104):3. PMID: 15339752
284. Anderson GL, Chlebowski RT, Aragaki AK, et al. Conjugated equine oestrogen and breast cancer incidence and mortality in postmenopausal women with hysterectomy: extended follow-up of the Women's Health Initiative randomised placebo-controlled trial. *The Lancet Oncology*. 2012PMID: 22401913
285. Prentice RL, Chlebowski RT, Stefanick ML, et al. Estrogen plus progesterin therapy and breast cancer in recently postmenopausal women. *Am J Epidemiol*. 2008 May 15(10):1207-16. PMID: 18372396
286. Beral V. Breast cancer risk in relation to the interval between menopause and starting hormone therapy. *J Natl Cancer Inst*. 2011;103(4):296-305. PMID: 21278356
287. Prentice RL, Chlebowski RT, Stefanick ML, et al. Conjugated equine estrogens and breast cancer risk in the Women's Health Initiative clinical trial and observational study. *Am J Epidemiol*. 2008 Jun 15(12):1407-15. PMID: 18448442
288. Hersh. National use of postmenopausal hormone therapy. *JAMA*. 2004;291(1):47-53. PMID: 14709575
289. Jemal. Recent trends in breast cancer incidence rates by age and tumor characteristics among U.S. women. *Breast Cancer Res*. 2007;9(3):R28. PMID: 17477859
290. Grodstein F. Postmenopausal hormone use and cholecystectomy in a large prospective study. *Obstet Gynecol*. 1994;83:5-11. PMID: 8272307
291. Liu B, Beral V, Balkwill A, et al. Gallbladder disease and use of transdermal versus oral hormone replacement therapy in postmenopausal women: prospective cohort study. *BMJ*. 2008;a386. PMID: 18617493
292. Boland. Hyperinsulinemia, dyslipidemia, and obesity as risk factors for hospitalized gallbladder disease: a prospective study. *Ann Epidemiol*. 2002;12(2):131-40. PMID: 11880221
293. Johnson. Menopausal hormone therapy and risk of colorectal cancer. *Cancer Epidemiol Biomarkers Prev*. 2009;18(1):196-203. PMID: 19124498
294. Grodstein F. Postmenopausal hormone use and risk for colorectal cancer and adenoma. *Annals of Internal Med*. 1998;128(9):705-12. PMID: 9556463

295. Rennert. Use of hormone replacement therapy and the risk of colorectal cancer. *J Clinical Oncology*. 2009;27(27):4542-7. PMID: 19704062
296. Schierbeck LL, Rejnmark L, Tofteng CL, et al. Effect of hormone replacement therapy on cardiovascular events in recently postmenopausal women: randomised trial. *BMJ*. 2012;345:e6409. PMID: 23048011.
297. Allen NE. Menopausal hormone therapy and risk of endometrial carcinoma among postmenopausal women in the European Prospective Investigation into Cancer and Nutrition. *Amer J Epidemiol*. 2010;172(12):1394-403. PMID: 20961969
298. Karageorgi. Reproductive factors and postmenopausal hormone use in relation to endometrial cancer risk in the Nurses' Health Study cohort 1976-2004. *Int J Cancer*. 2010;126(1):208-16. PMID: 19551854
299. Tsilidis. Menopausal hormone therapy and risk of ovarian cancer in the European prospective investigation into cancer and nutrition. *Cancer Causes Control*. 2011;22:1075-84. PMID: 21637986
300. Hildebrand. Postmenopausal hormone use and incident ovarian cancer: associations differ by regimen. *Int J Cancer*. 2010;127:2928-35. PMID: 21351271
301. Chien C LC, Heckbert SR, et al. Antidepressant use and breast cancer risk. *Breast Cancer Res and Treatment*. 2006;95:131-40. PMID: 16322894
302. Davis S MD. Medication use and the risk of breast cancer. *Eur J Epidemiol*. 2007;22:319-25. PMID: 17487555
303. Wernli KJ HJ, Trentham-Dietz A, Newcomb PA. Antidepressant medication use and breast cancer risk. *Pharmacoepidemiol Drug Saf*. 2009;18(4):284-90. PMID: 19226540
304. Lonn E, Bosch J, Yusuf S, et al. Effects of long-term vitamin E supplementation on cardiovascular events and cancer: a randomized controlled trial. *JAMA*. 2005 Mar 16;293(11):1338-47. PMID: 15769967.
305. Lee IM, Cook NR, Gaziano JM, et al. Vitamin E in the primary prevention of cardiovascular disease and cancer: the Women's Health Study: a randomized controlled trial. *JAMA*. 2005 Jul 6;294(1):56-65. PMID: 15998891.
306. Lin. Vitamins C and E and beta carotene supplementation and cancer risk: a randomized controlled trial. *J Natl Cancer Inst*. 2009;101:14-23. PMID: 19116389
307. Spangler L, Scholes D, Brunner RL, et al. Depressive symptoms, bone loss, and fractures in postmenopausal women. *J Gen Intern Med*. 2008 May(5):567-74. PMID: 18286345
308. Maugeri D, Panebianco P, Russo MS, et al. Ipriflavone-treatment of senile osteoporosis: results of a multicenter, double-blind clinical trial of 2 years. *Arch Gerontol Geriatr*. 1994 Nov-Dec;19(3):253-63. PMID: 15374271.
309. Passeri. Effects of 2-year therapy with ipriflavone in elderly women with established osteoporosis. *Ital J Mineral Electrolyte Metab*. 1995;9:137-44. PMID: 9263613
310. Alexandersen P, Toussaint A, Christiansen C, et al. Ipriflavone in the treatment of postmenopausal osteoporosis: a randomized controlled trial. *JAMA*. 2001 Mar 21(11):1482-8. PMID: 11255425
311. Cook NR. A randomized factorial trial of vitamins C and E and beta carotene in the secondary prevention of cardiovascular events in women. *Arch Intern Med*. 2007;167(15):1610-8. PMID: 17698683
312. Lloyd-Jones. Accuracy of death certificates for coding coronary heart disease as the cause of death. *Annals of Internal Med*. 1998;129(12):1020-6. PMID: 9867756
313. Clarkson TB, Utian WH, Barnes S, et al. The role of soy isoflavones in menopausal health: Report of the North American Menopause Society/Wulf H. Utian Translational Science Symposium in Chicago, IL (October 2010). *Menopause*. 2011(7):732-53. PMID: 21685820
314. Horn-Ross PL, John EM, Canchola AJ, et al. Phytoestrogen intake and endometrial cancer risk. *J Natl Cancer Inst*. 2003 Aug 6(15):1158-64. PMID: 12902445
315. Bandera EV, Williams MG, Sima C, et al. Phytoestrogen consumption and endometrial cancer risk: a population-based case-control study in New Jersey. *Cancer Causes Control*. 2009 Sep;20(7):1117-27. PMID: 19353280.
316. Xu. Soya food intake and risk of endometrial cancer among Chinese women in Shanghai: population based case-control study. *BMJ*. 2004;328:1285. PMID: 15136343
317. Bolanos R, Francia J. Isoflavones versus hormone therapy for reduction of vertebral fracture risk: indirect comparison. *Menopause*. 2010 Nov-Dec(6):1201-5. PMID: 20613672
318. Moher D, Jadad AR, Nichol G, et al. Assessing the quality of randomized controlled trials: an

- annotated bibliography of scales and checklists. *Control Clin Trials*. 1995 Feb;16(1):62-73. PMID: 7743790.
319. Salanti. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Journal of the Royal Society of Medicine*. 2012.
320. Archer DF, Dupont CM, Constantine GD, et al. Desvenlafaxine for the treatment of vasomotor symptoms associated with menopause: a double-blind, randomized, placebo-controlled trial of efficacy and safety. *Am J Obstet Gynecol*. 2009 Mar(3):238 e1- e10. PMID: 19167693
321. Archer DF, Seidman L, Constantine GD, et al. A double-blind, randomly assigned, placebo-controlled study of desvenlafaxine efficacy and safety for the treatment of vasomotor symptoms associated with menopause. *Am J Obstet Gynecol*. 2009 Feb(2):172 e1-10. PMID: 19110224
322. Speroff L, Gass M, Constantine G, et al. Efficacy and tolerability of desvenlafaxine succinate treatment for menopausal vasomotor symptoms: a randomized controlled trial. *Obstet Gynecol*. 2008 Jan(1):77-87. PMID: 18165395
323. Depomed. Study of Gabapentin Extended Release (G-ER) in the Treatment of Vasomotor (Hot Flashes/Hot Flushes) Symptoms in Postmenopausal Women. NCT00755417. 2012.
324. Freeman EW, Guthrie KA, Caan B, et al. Efficacy of escitalopram for hot flashes in healthy menopausal women: a randomized controlled trial. *JAMA*. 2011 Jan 19;305(3):267-74. PMID: 21245182.
325. Albertazzi P, Pansini F, Bonaccorsi G, et al. The effect of dietary soy supplementation on hot flashes. *Obstet Gynecol*. 1998 Jan(1):6-11. PMID: 9464712
326. Levis S, Strickman-Stein N, Ganjei-Azar P, et al. Soy isoflavones in the prevention of menopausal bone loss and menopausal symptoms: A randomized, double-blind trial. *Archives of Internal Medicine*. 2011(15):1363-9. PMID: 21824950
327. Radhakrishnan G, Rashmi, Agarwal N, et al. Evaluation of isoflavone rich soy protein supplementation for postmenopausal therapy. *Pakistan Journal of Nutrition*. 2009(7):1009-17. PMID: 15754123
328. Newton KM, Reed SD, LaCroix AZ, et al. Treatment of vasomotor symptoms of menopause with black cohosh, multibotanicals, soy, hormone therapy, or placebo: a randomized trial. *Ann Intern Med*. 2006 Dec 19(12):869-79. PMID: 17179056
329. Aso T, Uchiyama S, Matsumura Y, et al. A natural S-equol supplement alleviates hot flashes and other menopausal symptoms in equol nonproducing postmenopausal Japanese women. *J Womens Health (Larchmt)*. 2012 Jan;21(1):92-100. PMID: 21992596.
330. Hedrick RE, Ackerman RT, Koltun WD, et al. Estradiol gel 0.1% relieves vasomotor symptoms independent of age, ovarian status, or uterine status. *Menopause*. 2010 Nov-Dec(6):1167-73. PMID: 20720511
331. Maki PM, Gast MJ, Vieweg AJ, et al. Hormone therapy in menopausal women with cognitive complaints: a randomized, double-blind trial. *Neurology*. 2007 Sep 25(13):1322-30. PMID: 17893293
332. Rigano A, Rigano M, Cancellieri F, et al. Sexually and well-being in early menopause. Effect of transdermal estradiol therapy. *Panminerva Med*. 2001 Jun(2):115-8. PMID: 11449182
333. Simon JA, Stevens RE, Ayres SA, et al. Perimenopausal women in estrogen vasomotor trials: contribution to placebo effect and efficacy outcome. *Climacteric*. 2001 Mar(1):19-27. PMID: 11379374
334. Mattsson LA, Skouby S, Rees M, et al. Efficacy and tolerability of continuous combined hormone replacement therapy in early postmenopausal women. *Menopause Int*. 2007 Sep(3):124-31. PMID: 17785038
335. NIH. Estrogen Use and Postmenopausal Women. National Institutes of Health Consensus Development Conference Statement September 13-14, 1979. 1979. <http://consensus.nih.gov/1979/1979EstrogenPostmenopause018html.htm>.
336. Go AS, Mozaffarian D, Roger VL, et al. Heart disease and stroke statistics--2013 update: a report from the American Heart Association. *Circulation*. 2013 Jan 1;127(1):e6-e245. PMID: 23239837.

## List of Abbreviations

<b>Acronym</b>	<b>Definition</b>
AHRQ	Agency for Healthcare Research and Quality
BMI	body mass index
CAM	complementary and alternative medicine
CE	conjugated estrogen
CEE	conjugated equine estrogen
CER	comparative effectiveness review
CES-D	Center for Epidemiologic Studies Depression Scale
CHD	coronary heart disease
CI	confidence interval
CV	cardiovascular
DHEA	dehydroepiandrosterone
EPC	evidence-based practice center
E2V	estradiol valerate
FDA	Food and Drug Administration
GCS	Greene Climacteric Scale
GRADE	Grading of Recommendations Assessment, Development, and Evaluation
HF	hot flushes
HFNS	hot flushes and night sweats
HOPE	Health Outcomes Prevention Evaluation trial
HOPE-TOO	Health Outcomes Prevention Evaluation- The Ongoing Outcomes trial
HR	hazards ratio
HRT	hormone replacement therapy
IMS	International Menopause Society
IU	international unit
KI	Kupperman Index
MARIE	Mamma carcinoma Risk factor Investigation
MENQOL	Menopause-specific Quality of Life
MI	myocardial infarction
MPA	medroxyprogesterone acetate
MRS	Menopause Rating Scale
MSHF	moderate-to-severe hot flushes
MSHFNS	moderate-to-severe hot flushes and night sweats
MSVS	moderate-to-severe vasomotor symptoms
N	number
NAMS	North American Menopause Society
NETA	norethindrone acetate
NPNH	nonprescription nonhormone
NR	not reported
PICOTS	Population(s), Interventions, Comparators, Outcomes, Timing, and Setting
PMS	premenstrual syndrome
PND	postnatal depression
QOL	quality of life
RCT	randomized controlled trial

<b>Acronym</b>	<b>Definition</b>
RR	relative risk
SD	standard deviation
SNRI	serotonin–norepinephrine reuptake inhibitor
SRC	Scientific Resource Center
SSRI	selective serotonin reuptake inhibitor
STRAW	Stages of Reproductive Aging Workshop
TEP	Technical Expert Panel
THF	total hot flushes
THFNS	total hot flushes and night sweats
TOO	task order officer
USPSTF	United States Preventive Services Task Force
VAS	Visual Analog Scale
WHI	Women’s Health Initiative
WHQ	Women’s Health Questionnaire
WHS	Women’s Health Study
WISDOM	Women’s International Study of Long Duration Oestrogen after Menopause