

# Evidence-based Practice Center Systematic Review Protocol

## Project Title: Menopausal Symptoms: Comparative Effectiveness Review of Therapies

Original Date of Protocol: April 3, 2012

Amendment Date: June 5, 2013

(Amendment Details – see Section VII)

### I. Background and Objectives for the Systematic Review

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 by surgical or pharmacological means. It occurs naturally between the ages of 42 and 58<sup>1-3</sup> and is a consequence of reproductive senescence. The average age at onset appears fixed, as it has been unchanged since ancient Greece.<sup>4</sup> In the United States, the number of women entering menopause (approximately 2 million per year<sup>5</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>1-3</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>1-3</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>6</sup> In longitudinal studies, 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>7-9</sup> A natural history of symptoms can be described, including the presence, severity, and time since menopause.<sup>7</sup> For example, vasomotor symptoms generally begin 2 years before menopause, peak 1 year after menopause, and then diminish over the next 10 years.<sup>10</sup> In the Penn Ovarian Aging Study,<sup>11</sup> 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. In the Study of Women’s Health Across the Nation,<sup>12</sup> the prevalence of vasomotor symptoms was greater among blacks and women with a higher body mass index.

Estrogens have been a mainstay for treating symptoms but surrounded by controversy. Approved by the U.S. Food and Drug Administration (FDA) in 1942 for treating menopausal symptoms, 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>13</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 randomized comparison of estrogen and progestin to placebo in the Women’s Health Initiative (WHI)—not only was cardiovascular risk increased but overall harms exceeded benefits.<sup>14</sup> Although subsequent evaluation of the body of evidence has indicated interpretations more complex,<sup>15</sup> particularly for the target population included in this review, the consequences for hormone therapy use in the United States remain substantial.<sup>16</sup>

Pharmacies can also compound hormones combining agents intended to meet a specific patient’s need. In contrast to equine-derived hormone preparations, such as Premarin<sup>®</sup>, these compounded hormones are often claimed to be biochemically similar or identical to endogenous hormones; the FDA does not recognize the term “bioidentical.” Compounded preparations typically contain estriol and can have variable potency.<sup>17</sup> Employing them for treating menopausal symptoms is controversial, with many clinicians advocating their use<sup>18</sup> but with FDA scrutiny.<sup>19</sup> Current evidence supporting the safety and effectiveness of compounded hormone therapies is generally described as lacking.<sup>17</sup> Nevertheless, in 2003, approximately 30 million prescriptions for all compounded products were filled<sup>20</sup> and are heavily marketed<sup>21</sup>—a \$1 billion industry and growing.<sup>22</sup>

While hormone therapy is an effective treatment for menopausal symptoms, concerns about potential risks (especially cardiovascular disease, uterine and breast cancer) provide reason to consider other agents. Nonhormone prescription medications and nonprescription agents including complementary and alternative medicine (CAM) biological therapies have been studied, in comparison with hormone therapy or placebo, primarily as treatment of vasomotor symptoms.<sup>23</sup> Nonhormone prescription therapies include selective serotonin and norepinephrine reuptake inhibitors (SSRIs, SNRIs), eszopiclone, clonidine, methyldopa, 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, and dong quai (*Angelica sinensis*).<sup>5</sup> Postulated mechanisms for SSRIs and SNRIs include central effects on serotonin, dopamine, or norepinephrine,<sup>24</sup> while the potential benefit of isoflavones is thought to be mediated through affinity for estrogen receptors. Data suggest the use of nonprescription CAM therapies is common. In the Study of Women’s Health Across the Nation,<sup>25</sup> approximately 80 percent of participants reported using some form of CAM therapy during a 6-year followup period.

The principal uncertainty for nonhormone therapies is effectiveness, whereas for hormone therapies it is the balance of benefits and harms. Although the U.S. Preventive Services Task Force in 2005 addressed the use of hormone therapy to prevent chronic conditions by recommending against routine use (D-level recommendation for combined estrogen and progestin in women with an intact uterus or unopposed estrogen for women without a uterus), they did not consider treatment of menopausal symptoms in their guideline.<sup>26</sup> The 2010 North American Menopause Society (NAMS) position statement 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>27</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>28</sup> The NAMS position statement and IMS recommendations were not accompanied by systematic reviews. Yet both express considerable certainty and are somewhat at odds with trends in hormone therapy use.<sup>16</sup> The Endocrine Society recently performed an extensive review of evidence surrounding postmenopausal hormone therapy—published as a scientific statement.<sup>29</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.

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: 1) more than one symptom may be targeted; 2) some harms are distant but of consequence (e.g., breast cancer); 3) there are potential benefits distant in time (prevention of osteoporosis and fractures) that may not be the primary goal of treatment; 4) the array of nonhormone therapies is broad and includes a number of biologic CAM and prescription agents; 5) hormone therapies vary by preparation, type, and administration route; 6) compounded hormones are unstandardized and therefore heterogeneous; and 7) women without a uterus do not require a progestational agent to prevent uterine cancer. Additionally, there are many relevant domains in the postmenopause for quality-of-life outcomes: depression, somatic, memory, vasomotor, anxiety, sexual, sleep, menstrual, and self-esteem. Hormone therapy might benefit some domains positively, but others not at all, and yet have little impact on utility (mean EQ-5D scores in the Women’s International Study of Long Duration Oestrogen after Menopause [WISDOM] trial with combined menopausal hormone therapy were improved by only 0.016 over 1 year when compared with placebo).<sup>30</sup> Further, a primary evidence base for harms derives from the WHI in an overlapping but somewhat different target population (the treatment indication being chronic disease prevention in the study sample) than 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 characteristics of the WHI trials may be relevant, as a recent report from the WHI observational study<sup>31</sup> found women experiencing early vasomotor symptoms were at the lowest risk of cardiovascular disease and cardiovascular events.

From a decisionmaking perspective (patient, provider, and policymaker), the evidence surrounding hormone therapy and alternatives presents similar challenges. Probably the most important decision for an individual menopausal woman choosing hormone agents is the question: Given symptoms, what is the balance of benefits and harms and how does timing and duration of therapy affect the balance? Accordingly, the objectives of this review include: systematically reviewing and synthesizing evidence evaluating the comparative effectiveness of treatments for menopausal symptoms, benefits from treatments other than symptom relief, and potential harms.

## II. The 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 are 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; cholecystitis; 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; cholecystitis; 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 patients defined by demographics, symptom severity, other medications, and comorbidities or according to agent, preparation, or dose?

## **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**

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)

The nonhormone therapies are listed above (KQs 1 and 3):

- **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
    - Cholecystitis
    - Osteoporotic fractures
    - Agent-specific adverse events
- **Timing**

Outcome assessment at least 12 weeks from the baseline assessment.
  - **Setting**

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

## Summary of Public Comments

**KQ 1.** Comments included suggestions to expand the list of complementary and alternative biologic treatments and to revise categorization of nonhormone treatments; concern was also expressed about the number of interventions and outcomes included in the review. As a result, the nonhormone group was modified as “Prescription” and “Nonprescription/ complementary and alternative therapies.” While the scope is inclusive, it was developed in consultation with experts.

**KQ 2.** Issues raised included the extent and available evidence for compounded products. The perspective adopted here is that the purpose of a systematic review is to both identify evidence gaps and evidence that can be synthesized. A point was raised about combining benefits and harms into a KQ. The review team considered this an organizational issue, as outcomes are evaluated independently. It was recommended that age, in addition to time-since-menopause should be considered—age was accordingly incorporated (analogously in KQ 3).

**KQ 3.** It was suggested that KQ 3 be worded similarly to KQ 2, and this was done.

**KQ 4.** The importance of examining subgroups based on symptom severity was noted and now specified. The issue of tamoxifen use was raised with respect to women with breast cancer. Given that women with breast cancer are not an included population, this was not further

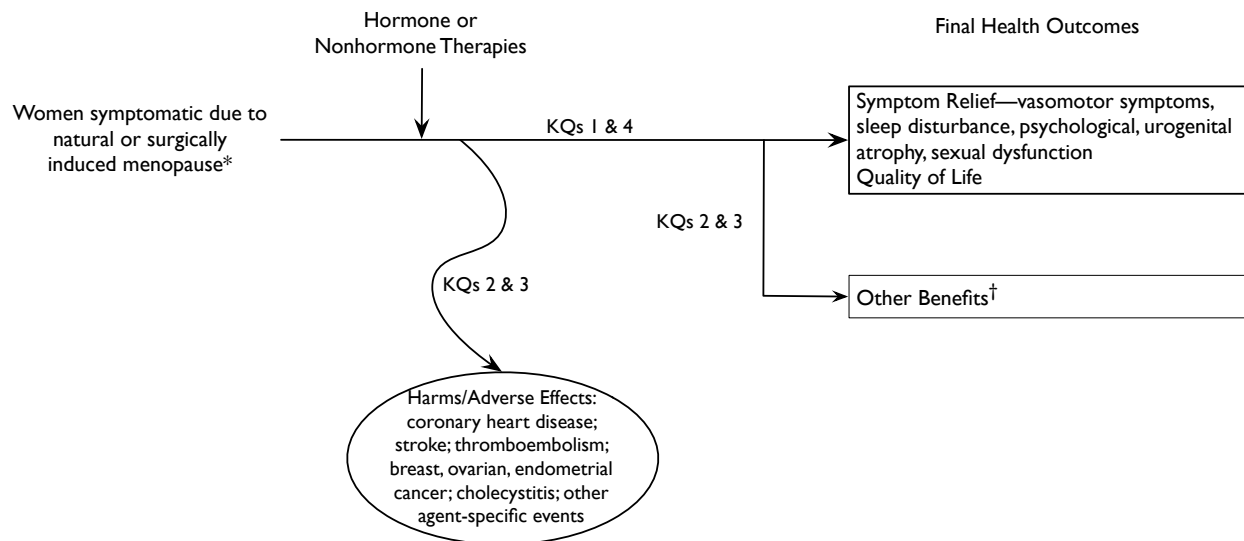
considered. That exclusion has now been made explicit in the KQs, Figure 1, and the inclusion/exclusion criteria.

Finally, ovarian cancer was added as an outcome to KQs 2 and 3, and Bellergal<sup>®</sup> was deleted to reduce the scope. Prioritization was conducted with input by the Technical Expert Panel (TEP).

### III. Analytic Framework

Figure 1 depicts the KQs illustrating how hormone and nonhormone therapies for menopausal symptoms may improve symptoms and quality of life and prevent osteoporotic fractures and colorectal cancer. Also considered are consequential adverse effects among women using hormone therapies; these adverse effects include coronary heart disease, stroke, thromboembolism, breast cancer, endometrial cancer, ovarian cancer, and cholecystitis. Finally, adverse events accompanying the use of nonhormone agents are also depicted.

Figure 1. Draft analytic framework



KQ = key question

\* Excludes women with breast cancer or receiving tamoxifen

† Osteoporotic fractures and colorectal cancer

## IV. Methods

### A. Criteria for Inclusion/Exclusion of Studies in the Review

#### Key Question 1—Symptom Relief

We will include randomized controlled trials (RCTs) with placebo or an active comparator. Because we anticipate sufficient RCTs, we do not anticipate the need to include nonrandomized studies to establish a GRADE of evidence.<sup>32,33</sup> RCTs should have at least 25 patients randomized per arm who are studied for at least 12 weeks; these conditions are minimums consistent with trials used to define efficacy for vasomotor symptoms. Other meta-analyses and systematic reviews will not be included. Bibliographies of these reviews will be examined to identify potential trials. Table 1 summarizes these inclusion and exclusion criteria.

**Table 1. Hormone and nonhormone therapies study inclusion/exclusion criteria—vasomotor symptoms, sleep disturbance, psychological, urogenital atrophy, sexual dysfunction, and quality of life**

RCTs with placebo comparator	Include <sup>a</sup>
RCTs with active comparator	Include <sup>a</sup>
Meta-analyses/systematic reviews	Exclude <sup>b</sup>
Observational studies	Exclude
Single arm/case series	Exclude
Case reports	Exclude
Minimum duration	≥ 12 weeks
Sample size	≥25 patients 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.

RCTs = randomized controlled trials

#### Key Question 2—Other Benefits/Harms Hormones

We will use a sequential approach to study inclusion as outlined in Table 2. If meta-analyses/systematic reviews of appropriate relevance are identified, they will be used as the primary evidence base<sup>34</sup> (supplemented by any more recent RCTs and observational studies when the GRADE of evidence [according to the Grading of Recommendations Assessment, Development, and Evaluation system<sup>35</sup>] provided by the meta-analyses/systematic reviews are judged low or insufficient;<sup>36</sup> see Table 2).

For some of the nine included outcomes, there are potentially many systematic reviews and meta-analyses. Systematic reviews and meta-analyses will be assessed and prioritized for inclusion in a manner informed by methods guidance<sup>37</sup> for the Evidence-based Practice Center (EPC) Program, remaining cognizant of the need to minimize potential bias and to balance that need by practical considerations. The most current and highest quality reviews, as rated by AMSTAR,<sup>38</sup> will be included.

Given the natural history of osteoporosis—as well as breast, ovarian, and colorectal cancer—a minimum study-duration criterion of 5 years will be applied to longitudinal studies ascertaining



those outcomes. While studies of very large samples (e.g., many thousands) would be preferred, 250 as a minimum will be applied to maintain sensitivity. Outcomes included were identified in consultation with the TEP to capture those most consequential. They are not intended to be an exhaustive list.

**Table 2. Hormone therapies study inclusion/exclusion—coronary heart disease, stroke, or thromboembolism; cholecystitis; osteoporotic fractures; or endometrial, breast, colorectal, or ovarian cancer**

RCTs with placebo comparator	Include if meta-analyses/systematic reviews insufficient <sup>a</sup>
RCTs with active comparator	Include if meta-analyses/systematic reviews insufficient <sup>a</sup>
Meta-analyses/systematic reviews	Include
Observational studies	Include if meta-analyses/systematic reviews insufficient <sup>a</sup>
Single arm/case series	Exclude
Case reports	Exclude
Minimum duration	5 years <sup>b</sup> 1 year <sup>c</sup>
Sample size	>250

<sup>a</sup> If evidence is sufficient to grade outcomes obtained by meta-analyses/systematic reviews, RCTs and observational designs will not be included (see Rationale section in following text for a detailed explanation).

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

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

RCTs = randomized controlled trials

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

**We will limit 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.**

**We will assess agent-specific adverse events from RCTs, meta-analyses/systematic reviews, and observational studies for each agent (Table 3). For coronary heart disease, stroke, or thromboembolism; cholecystitis; osteoporotic fractures; or endometrial, breast, colorectal, or ovarian cancer; similar study types, duration, and size criteria will be applied as described in**

Table 4 for KQ 2.

### Nonhormone Prescription Therapies

**Table 3. Nonhormone therapies study inclusion/exclusion of agent-specific adverse events**

RCTs with placebo comparator	Include
RCTs with active comparator	Include
Meta-analyses/systematic reviews	Include
Observational studies	Include
Single arm/case series	Exclude
Case reports	Exclude
Minimum duration	≥12 weeks
Sample size	≥25 patients randomized per arm

RCTs = randomized controlled trials

**Table 4. Nonhormone prescription therapies study inclusion/exclusion—coronary heart disease, stroke, or thromboembolism; cholecystitis; osteoporotic fractures; or endometrial, breast, colorectal, or ovarian cancer**

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

For nonhormone nonprescription/CAM therapies, any study design identifying agent-specific harms will be included (Table 5). Case reports will be included for life-threatening events. The approach adopted in KQ 3 to identify studies of other benefits and harms corresponding to outcomes relevant for hormones will be used (Table 6).

Due to scope issues, we limited the list of included agents, and the focus of the search strategy, which was prioritized in consultation with the TEP. It is not exhaustive. Acknowledging this approach, when trials examining efficacy for ameliorating menopausal symptoms but using agents other than those listed are identified in the literature search (e.g., other botanicals), those studies will be considered for inclusion if meeting other criteria.

### **Nonhormone Nonprescription/Complementary and Alternative Therapies**

**Table 5. Nonhormone nonprescription and complementary and alternative therapies study inclusion/exclusion criteria for agent-specific adverse events**

RCTs with placebo comparator	Include
RCTs with active comparator	Include
Meta-Analyses/systematic reviews	Include
Observational studies	Include
Single arm/case series	Include
Case reports (life-threatening events)	Include
Minimum duration	None
Sample size	None

RCTs = randomized controlled trials

**Table 6. Nonhormone nonprescription and complementary and alternative therapies study inclusion/exclusion — coronary heart disease, stroke, or thromboembolism; cholecystitis; osteoporotic fractures; or endometrial, breast, colorectal, or ovarian cancer**

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

<sup>a</sup> Longitudinal studies of colorectal cancer, breast cancer, 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

## Language

Studies will be limited to those published in the English language as those relevant to the target population are unlikely to appear in non-English periodicals

## Key Question 4—Subgroups

Subgroups (vasomotor symptom severity, years since menopause [age], ethnicity, comorbidities [smoking, obesity], estrogen dose; for harms, years since menopause [age], duration of therapy) will be selected as reported from included studies in KQs 1–3. Women with breast cancer are excluded.

## Grey Literature

Grey literature will be sought by searching [clinicaltrials.gov](http://clinicaltrials.gov) for clinical trials, the FDA Web site, and relevant conference abstracts (conferences identified by TEP members) for data pertaining to the interventions under consideration that are used to treat menopausal symptoms. Study authors will be contacted for unpublished results if the two senior team members concur that if obtained, evidence could impact results meaningfully (i.e., alter evidence GRADE). Additional and potentially unpublished evidence will be requested by the Scientific Resource Center.

## **B. Search Strategies**

Search strategies were developed (see Appendix) by an expert librarian in collaboration with the study team. No date limitations will be applied.

## **C. Data Abstraction and Data Management**

Searches from MEDLINE<sup>®</sup> will be transferred into Endnote<sup>®</sup> (Thomson Reuters, New York, NY) and subsequently into DistillerSR (Evidence Partners Inc., Manotick, ON, Canada) for selection. A decision whether to adopt initial title screening instead of title/abstract screening will be based on evaluating the agreement between senior and junior team members in applying inclusion/exclusion criteria on a training set of 100 randomly selected citations. The initial training set will be followed by at least two subsequent sets of 50 citations or until sufficient agreement is achieved (target 95%) between two independent reviewers applying specified criteria—either title or title/abstract screening as decided from the initial training set. If title screening is adopted, an analogous approach will be used for the subsequent title/abstract screening. If title screening is used, all excluded references will undergo a second screen to assure sensitivity. In the title/abstract screening phase, all references will undergo dual review by a senior and a junior team member for inclusion in the full-text review, with disagreements resolved by an independent senior team member.

For identified citations, the full text will be reviewed in the same fashion to determine their inclusion/exclusion. Reasons for exclusion for each paper retrieved as full text, but excluded from the review, will be entered in the DistillerSR database.

Data will be abstracted into tables created in DistillerSR, with elements defined by an accompanying data dictionary. A training set of five articles will be abstracted by all team members. Data abstraction either will be performed in duplicate or the initial abstraction will be independently verified, with discrepancies identified and resolved by consensus. Depending on the extent of available data, results from crossover trials will be pooled by using the methods of Curtin et al.,<sup>39, 40</sup> or if not possible, using only the first phase.

To ensure reproducibility, abstracted data will not be edited outside of DistillerSR. Data will be transferred to R<sup>41</sup> to compile study-level and summary tables in Microsoft<sup>®</sup> Excel format for inclusion in the report. Evidence synthesis will also be performed in R.

While complete specification of data to be extracted will be developed during the abstraction phase, some anticipated elements include, but are not limited to, the following: author, study year, enrollment dates, center(s), disclosures and conflicts of interest, funding, blinding, numbers of patients, age, ethnicity, surgical or natural menopause, intervention, outcome instrument, and result.

## D. Assessment of Methodological Quality of Individual Studies

In adherence with the EPC Program *Methods Guide for Effectiveness and Comparative Effectiveness Reviews* (hereafter *Methods Guide*),<sup>42</sup> the general approach to grading individual comparative studies will be performed by applying the criteria of the U.S. Preventive Services Task Force.<sup>43</sup> The quality of the abstracted studies and the body of evidence will be assessed by two independent reviewers. Ratings of good, fair, and poor are obtained (detailed in the Appendix—Application of USPSTF Grading for randomized controlled trials<sup>43</sup>). Discordant quality assessments will be resolved with input from a third reviewer, if necessary. AMSTAR, a validated tool, will be used for quality assessment of meta-analyses.<sup>38</sup>

Even with appropriate analysis, the ability of observational studies to identify unconfounded associations and causal effects<sup>44</sup> or ascertain harms<sup>45</sup> can be highly variable. Moreover, all observational data are considered lesser (low) GRADE strength.<sup>35</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 quantitative approach, we will adapt the method described by Thompson et al.<sup>46</sup>

## E. Data Synthesis

Studies employ a variety of outcome instruments (Appendix Table 1). When appropriate (e.g., similar instruments were used or standardized effect measures were relevant and interpretable), outcome measures will be pooled according to the EPC Program *Methods Guide*<sup>47</sup>,<sup>48</sup> will be synthesized in R<sup>41</sup> by using the meta<sup>49</sup> and metafor<sup>50</sup> packages. Clinical heterogeneity, and appropriateness for pooling, will be 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>51</sup> random-effects models will be used. The magnitude of statistical heterogeneity will be examined by using  $I^2$ , acknowledging potential limitations<sup>52</sup>; when values exceed 25 percent, we will consider examining heterogeneity in meta-regressions.<sup>53</sup> Evidence of possible publication bias will be explored by using funnel plots. When there is suggestion or suspicion of effects dependent on baseline risks, potential effects will be examined by using Bayesian methods to account for any correlation between baseline and relative effects.<sup>54, 55</sup> Potential subgroup-specific effects for benefits will be examined as reported but will include: vasomotor symptom severity, years since menopause (age), ethnicity, comorbidities (smoking, obesity), and estrogen dose; for harms, years since menopause (age) and duration of therapy. All results will be evaluated separately for women with and without a uterus.

Outcomes will be summarized and reported in the order specified by therapies in the KQs.

## F. Grading the Evidence for Each Key Question

Determinations of the strength of the body of evidence will be based on the Evidence-based Practice Center (EPC) approach,<sup>32, 33</sup> which is conceptually similar to the GRADE system.<sup>35</sup> The four main domains to be assessed are risk of bias, consistency, directness, and precision, with additional domains of dose-response association, confounding, strength of association, and publication bias. The body of evidence will be evaluated separately for each major outcome and

for all major comparisons of the comparative effectiveness review to derive a single GRADE of high, moderate, low, or insufficient evidence. Two reviewers will conduct the evaluations, and agreement will be reached through discussion and consensus.

The GRADE definitions are as follows:

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

## G. Assessing Applicability

The objective of this review is to facilitate an evidence-based approach to treating symptoms associated with menopause. Hence, the population of interest is women experiencing menopausal symptoms. The body of evidence, however, particularly for observational studies, may include menopausal women regardless of the presence of symptoms. Such evidence will require extrapolation to the population of interest. When such evidence is used in this report, we will clearly note it and comment on the applicability of the evidence to women with menopausal symptoms.

Other examples of anticipated limitations in interpretation of the evidence include differences in dosages between studies (which may or may not be reported), changes in prescriptions over time rendering comparisons difficult, indirect comparisons derived from both placebo-controlled and treatment controlled trials.

## V. References

1. Soules MR, Sherman S, Parrott E, et al. Executive summary: Stages of Reproductive Aging Workshop (STRAW). *Climacteric*. 2001 Dec;4(4):267-72. PMID: 11770182.
2. 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.
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. Amundsen DW, Diers CJ. The age of menopause in classical Greece and Rome. *Hum Biol*. 1970 Feb;42(1):79-86. PMID: 4910710.
5. *Menopause Practice. A Clinician's Guide*. 4th ed: North American Menopause Society; 2010.
6. McKinlay SM, Brambilla DJ, Posner JG. The normal menopause transition. *Maturitas*. 1992 Jan;14(2):103-15. PMID: 1565019.
7. Nelson HD, Haney E, Humphrey L, et al. Management of menopause-related symptoms. *Evid Rep Technol Assess (Summ)*. 2005 Mar(120):1-6. PMID: 15910013.

8. 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.
9. . 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.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. Tsai SA, Stefanick ML, Stafford RS. Trends in menopausal hormone therapy use of US office-based physicians, 2000-2009. *Menopause.* 2010 Nov 24;18(4):385-92. PMID: 21127439.
17. . Bioidentical hormones. *Med Lett Drugs Ther.* 2010 May 31;52(1339):43-4. PMID: 20508582.
18. Somers S. *Ageless: The Naked Truth About Bioidentical Hormones.* 1st ed. New York: Crown Publishers; 2006.
19. 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.
20. 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>.
21. Patsner B. Pharmacy compounding of bioidentical hormone replacement therapy (BHRT): a proposed new approach to justify FDA regulation of these prescription drugs. *Food Drug Law J.* 2008;63(2):459-91. PMID: 18561473.
22. 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&usg=AFQjCNEFjV4KhLb\\_8iC8TzbCIAMdJyUKWA&sig2=qtkkOkr8PI0jYivhRZ\\_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&usg=AFQjCNEFjV4KhLb_8iC8TzbCIAMdJyUKWA&sig2=qtkkOkr8PI0jYivhRZ_Vaw). Accessed on October 10 2011.
23. 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.

24. 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.
25. Robinson N, Blair M, Lorenc A, et al. Complementary medicine use in multi-ethnic paediatric outpatients. *Complement Ther Clin Pract*. 2008 Feb;14(1):17-24. PMID: 18243938.
26. . Hormone therapy for the prevention of chronic conditions in postmenopausal women: recommendations from the U.S. Preventive Services Task Force. *Ann Intern Med*. 2005 May 17;142(10):855-60. PMID: 15897536.
27. . 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.
28. Pines A, Sturdee DW, Birkhauser MH, et al. IMS updated recommendations on postmenopausal hormone therapy. *Climacteric*. 2007 Jun;10(3):181-94. PMID: 17487645.
29. 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.
30. 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.
31. Szmuiłowicz ED, Manson JE, Rossouw JE, et al. Vasomotor symptoms and cardiovascular events in postmenopausal women. *Menopause*. 2011 Feb 19;18(6):603-10. PMID: 21358352.
32. Owens DK, Lohr KN, Atkins D, et al. Grading the strength of a body of evidence when comparing medical interventions. In: Agency for Healthcare Research and Quality. *Methods Guide for Comparative Effectiveness Reviews*. July 2009. Rockville, MD.  
[http://www.effectivehealthcare.ahrq.gov/ehc/products/60/318/2009\\_0805\\_grading.pdf](http://www.effectivehealthcare.ahrq.gov/ehc/products/60/318/2009_0805_grading.pdf).
33. 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.
34. White CM, Ip S, McPheeter M, et al. Using existing systematic reviews to replace de novo processes in conducting Comparative Effectiveness Reviews. In: Agency for Healthcare Research and Quality. *Methods Guide for Comparative Effectiveness Reviews*. September 2009. Rockville, MD.  
<http://www.effectivehealthcare.ahrq.gov/ehc/products/123/329/SystematicReviewsReplaceDeNovo.pdf>.
35. Balslem 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.
36. Golder S, Loke YK, Bland M. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. *PLoS Med*. 2011 May;8(5):e1001026. PMID: 21559325.
37. 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.
38. 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.
39. Curtin F, Elbourne D, Altman DG. Meta-analysis combining parallel and cross-over clinical trials. II: Binary outcomes. *Stat Med*. 2002 Aug 15;21(15):2145-59. PMID: 12210630.
40. Curtin F, Altman DG, Elbourne D. Meta-analysis combining parallel and cross-over clinical trials. I: Continuous outcomes. *Stat Med*. 2002 Aug 15;21(15):2131-44. PMID: 12210629.



41. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.
42. Helfand M, Balshem H. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. August 2009. Rockville, MD.  
<http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayProduct&productID=318>.
43. 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.
44. Hernan MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health.* 2006 Jul;60(7):578-86. PMID: 16790829.
45. Vandembroucke JP. What is the best evidence for determining harms of medical treatment? *CMAJ.* 2006 Feb 28;174(5):645-6. PMID: 16505461.
46. 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.
47. 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 Apr 6; PMID: 21477993.
48. Fu R, Gartlehner G, Grant M, et al. Conducting Quantitative Synthesis When Comparing Medical Interventions: AHRQ and the Effective Health Care Program. In: Agency for Healthcare Research and Quality. Methods Guide for Comparative Effectiveness Reviews. October 2010.  
[http://www.effectivehealthcare.ahrq.gov/tasks/sites/ehc/assets/File/MethodsGuide\\_ConductingQuantitativeSynthesis.pdf](http://www.effectivehealthcare.ahrq.gov/tasks/sites/ehc/assets/File/MethodsGuide_ConductingQuantitativeSynthesis.pdf).
49. Schwarzer G. meta: Meta-Analysis with R (version 1.6-1). 2011. <http://CRAN.R-project.org/>. Accessed on June 16 2011.
50. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software.* 2010;36(3):1-48.
51. Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. *Psychol Meth.* 1998;3(4):486-504. PMID: none.
52. 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.
53. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med.* 2002 Jun 15;21(11):1559-73. PMID: 12111920.
54. Spiegelhalter D, Thomas A, Best N, et al. OpenBUGS User Manual. Version 3.1.2. September. 2010. <http://www.openbugs.info/w/>. Accessed on June 16 2011.
55. Arends LR, Hoes AW, Lubsen J, et al. Baseline risk as predictor of treatment benefit: three clinical meta-re-analyses. *Stat Med.* 2000 Dec 30;19(24):3497-518. PMID: 11122510.
56. Guttuso T, Evans M. Minimum trial duration to reasonably assess long-term efficacy of nonhormonal hot flash therapies. *Journal of women's health.* 2010 Apr;19(4):699-702. PMID: 20201698.
57. 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.
58. Prentice RL. Observational studies, clinical trials, and the women's health initiative. *Lifetime Data Anal.* 2007 Dec;13(4):449-62. PMID: 17943443.

59. 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.
60. 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.
61. 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.
62. Salanti G, Higgins JP, Ades AE, et al. Evaluation of networks of randomized trials. *Stat Methods Med Res*. 2008 Jun;17(3):279-301. PMID: 17925316.
63. Lunn D, Spiegelhalter D, Thomas A, et al. The BUGS project: Evolution, critique and future directions. *Stat Med*. 2009 Nov 10;28(25):3049-67. PMID: 19630097.
64. 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.
65. 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/>.

## VI. Definition of Terms

Not applicable.

## VII. Summary of Protocol Amendments

Date	Section	Original Protocol	Revised Protocol (italics changes)	Rationale
5/30/2012	II	For KQ 2 “cholecystitis” specified as outcome.	Replaced with “ <i>gall- bladder disease.</i> ”	To maintain consistency with USPSTF report and based on the recommendation of the clinical content expert and reporting in included studies.
5/30/2012	IV A.	RCTs should have at least 25 patients randomized per arm who are studied for at least 12 weeks; these conditions are minimums consistent with trials used to define efficacy for	RCTs should have at least 25 patients randomized per arm who are studied for at least 12 weeks; these conditions are minimums consistent with trials used to define efficacy for	Based on evidence that efficacy treating vasomotor symptoms with these agents is demonstrable by 4 to 8 weeks—and translates into similar efficacy at 12 weeks. <sup>56, 57</sup>

		vasomotor symptoms.	vasomotor symptoms. <i>For centrally acting agents (SSRI, SNRI, gabapentin, and pregabalin) minimum trial duration will be 4 weeks.</i>	
5/30/2012	IV A.	<p>We will use a sequential approach to study inclusion as outlined in Table 2. If meta-analyses/systematic reviews of appropriate relevance are identified, they will be used as the primary evidence base<sup>34</sup> (supplemented by any more recent RCTs and observational studies when the GRADE of evidence [according to the Grading of Recommendations Assessment, Development, and Evaluation system<sup>35</sup>] provided by the meta-analyses/systematic reviews are judged low or insufficient,<sup>36</sup> see Table 2).</p> <p>For some of the nine included outcomes, there are potentially many systematic reviews and meta-analyses. Systematic reviews and meta-analyses will be assessed and prioritized for inclusion in a manner</p>	<p><i>Study selection to evaluate treatment effects (i.e., causal) for KQ 2 will be limited to systematic reviews of randomized controlled trials and supplemented by individual studies as appropriate. SRs 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. During the search update, new SRs and RCTs not included in the original SRs will be assessed, appraised, and evidence will be synthesized with strength of evidence assigned.</i></p>	<p>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>58</sup> These discrepancies have been attributed to two primary reasons—selection bias and time-varying confounding.<sup>59-61</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 in including hip fractures<sup>59</sup> and colorectal cancer as well.<sup>61</sup> Relying on observational data</p>

		<p>informed by methods guidance<sup>37</sup> for the Evidence-based Practice Center (EPC) Program, remaining cognizant of the need to minimize potential bias and to balance that need by practical considerations. The most current and highest quality reviews, as rated by AMSTAR,<sup>38</sup> will be included.</p>	<p>employing standard analyses to examine these outcomes is problematic.<sup>60</sup> Accordingly, study selection for KQ2 will be limited to systematic reviews of randomized controlled trials and supplemented with individual studies if no SR or to update SRs as needed.</p> <p>For KQ2 our original goal was to: 1) identify SRs of randomized controlled trials, and 2) review those trials included in the SR of sufficient duration and sample size (&gt;250 participants) to plausibly establish causal effects: 5 years for colorectal, breast, or ovarian cancers and fracture outcomes; one year for coronary heart disease, stroke, or thromboembolism and gall bladder disease. If a SR (or reviews) is not identified we will proceed to identify randomized controlled trials meeting these criteria from our search.</p>	
5/30/2012	IV E		<p><i>Where feasible for vasomotor symptoms and QoL outcomes, network meta-analyses will be performed including the most</i></p>	<p>During the course of data abstraction it became apparent that the extent of data would likely be sufficient for these</p>

			<i>relevant comparisons with sufficient data.</i>	analyses. Network meta-analysis also formally allows quantitative indirect comparisons. While AHRQ does not currently provide guidance on the conduct and reporting of these analyses, we will adhere to generally accepted principles. <sup>62</sup> Analyses will be performed using OpenBUGS. <sup>63</sup>
5/30/2012	IV E	The magnitude of statistical heterogeneity will be examined by using $I^2$ , acknowledging potential limitations; when values exceed 25 percent, we will consider examining heterogeneity in meta-regressions. <sup>53</sup>	<i>Heterogeneity will be reported as <math>\tau^2</math>.</i> <sup>52</sup>	<i>Between-trial variances can be interpreted more intuitively on the effect estimate scale</i> <sup>52</sup>

## VIII. Review of Key Questions

For all EPC reviews, key questions were reviewed and refined as needed by the EPC 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. In addition, for Comparative Effectiveness reviews, the key questions were posted for public comment and finalized by the EPC after review of the comments.

## IX. Key Informants

Key Informants are the end-users of research, including patients 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 solicits input from Key Informants when developing questions

for systematic review or when identifying high-priority research gaps and needed new research. Key Informants are 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 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 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.

## **X. Technical Experts**

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 study 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 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 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.

## **XI. Peer Reviewers**

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 must 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.

## **XII. EPC team disclosures**

None.

### **XIII. Role of the Funder**

This project was funded under Contract No. HHS A 290-2007-10058-I from the Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services. The Task Order Officer reviewed contract deliverables for adherence to contract requirements and quality. The authors of this report are responsible for its content. Statements in the report should not be construed as endorsement by the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services.

## Appendix

**Table 1. Instruments used to assess menopausal symptoms**

Outcomes	Vasomotor	Depression/Mood	Sexual Dysfunction	Urogenital Atrophy	Sleep	Quality of Life
Instrument						
Self Report	•		•	•		
Greene Climacteric Scale	•	•	•		•	•
Kupperman Menopausal Index						
MENQOL	•	•	•			•
MQOL	•	•	•			•
Women's Health Questionnaire	•	•	•		•	•
Beck Depression Inventory		•				
Beck Anxiety Inventory		•				
Hamilton Depression Rating Scale		•				
Hamilton Anxiety Rating Scale		•				
McCoy Sex Scale Questionnaire			•			
WHI Insomnia Rating Scale					•	
SF-36		•				•

**Table 2. Instruments used to assess menopausal symptoms in the WHI**

Outcomes	Vasomotor	Depression/Mood	Sexual Dysfunction	Urogenital Atrophy	Sleep	Quality of Life
Instrument						
Self Report	•		•	•		
WHI Insomnia Rating Scale					•	
SF-36						•
Abbreviated CES-D		•				



## PubMed Searches

1. “Menopause”[Mesh] OR menopause OR menopausal OR “post-menopause” OR postmenopause OR “post-menopausal” OR postmenopausal OR climacteri\* OR perimenopausal OR “peri-menopause” OR “peri-menopausal” OR perimenopausal

2. “therapy” [Subheading] OR “Therapeutics”[Mesh] OR “Estrogen Replacement Therapy”[Mesh] OR “Drug Therapy”[Mesh] OR “drug therapy” [Subheading] OR “therapeutic use” [Subheading] OR “Hormone Replacement Therapy”[Mesh] OR “Complementary Therapies”[Mesh] OR “Estrogens”[Mesh] OR “Progestins”[Mesh] OR estrogen\* OR progestin\* OR “hormone replacement” OR antidepressant\* OR eszopiclone OR clonidine OR methylphenidate OR bupropion OR gabapentin OR pregabalin OR isoflavone\* OR “red clover” OR “black cohosh” OR cimicifuga OR “st. johns wort” OR ginseng OR flaxseed OR “vitamin E” OR “dong quai” OR “Dehydroepiandrosterone”[Mesh] OR “Androgens”[Mesh] OR DHEA OR dehydroepiandrosterone OR “androgenic agents” OR “androgenic compounds” OR androgen

3. Subset: Systematic Review OR Publication Type: Meta-analysis OR (“meta-analysis” OR metaanalysis OR “systematic review”)

4. Limits: English, Human

(1 AND 2) AND (“meta-analysis” OR metaanalysis OR “systematic review”)  
{Add those NOT in previous set and in English and relevant}

## Specific Outcomes

### Vasomotor Symptoms

hyperhidrosis (mh) OR “hot flashes”(mh) OR “vasomotor symptoms” OR “hot flashes” OR “night sweats” OR sweats OR flushes AND (1 AND 2) AND Limits: English, Human AND “randomized controlled trial”(pt)

### Sleep Disturbance

dyssomnias (mh) OR “sleep initiation and maintenance disorders” (mh) OR insomnia OR sleeplessness OR “early awakening” OR “somatic complaints” AND (1 AND 2) AND Limits: English, Human AND “randomized controlled trial” (pt)

### Psychological Symptoms

“behavioral symptoms” (mh) OR “mood disorders” (mh) OR irritability OR depression OR despair OR anxiety OR “difficulty concentrating” OR “over-reacting” OR forgetfulness OR “reminiscence lapses” OR “mood swings” OR “temper swings” OR “emotional flare-ups” OR weepiness AND (1 AND 2) AND Limits: English, Human AND “randomized controlled trial”(pt)

### Urogenital Atrophy

“female urogenital diseases” (mh) OR “urogenital system/pathology” (mh/sh) OR “urogenital disorders” OR ((vulva\* OR vagina\* OR vulvovaginal OR urinary OR genital OR urogenital) AND atrophy) OR “atrophic vaginitis” AND (1 AND 2) AND Limits: English, Human AND “randomized controlled trial” (pt)

### **Sexual Function**

“Sexual Dysfunction, Physiological” (mh) OR “Sexual Dysfunctions, Psychological libido” (mh) OR “female sexual dysfunction” OR “female sexual dysfunctions” OR “hypoactive sexual desire disorder” OR “sexual function” OR “sexual desire” OR “sexual satisfaction” AND (1 AND 2) AND Limits: English, Human AND “randomized controlled trial”(pt )

### **Quality of Life**

“quality of life” (mh) OR “quality of life” OR “well-being” AND (1 AND 2) AND Limits: English, Human AND “randomized controlled trial”(pt)

### **Osteoporotic Fractures**

“Osteoporotic fractures” (mh) OR (fracture\* AND (osteoporosis OR osteoporotic)) AND (1 AND 2) AND Limits: English, Human AND Subset: Systematic Review OR Publication Type: Meta-analysis OR (“meta-analysis” OR metaanalysis OR “systematic review”)

### **Colorectal Cancer**

“Colorectal Neoplasms” (mh) OR ((colorectal OR colon) AND (neoplasm\* OR cancer OR carcinoma)) AND (1 AND 2) AND Limits: English, Human AND Subset: Systematic Review OR Publication Type: Meta-analysis OR (“meta-analysis” OR metaanalysis OR “systematic review”)

### **Harms Other Than Specified Outcomes**

1. “Menopause”[Mesh] OR menopause OR menopausal OR “post-menopause” OR postmenopause OR “post-menopausal” OR postmenopausal OR climacteri\* OR perimenopause OR “peri-menopause” OR “peri-menopausal” OR perimenopausal

2. (((“adverse effects” [Subheading]) OR “complications” [Subheading]) OR ( “poisoning” [Subheading] OR “Poisoning”[Mesh] )) OR “drug effects” [Subheading] OR “Drug Toxicity”[Mesh]

AND

(“Ovarian Neoplasms”[Mesh] OR “Breast Neoplasms”[Mesh] OR “Colorectal Neoplasms”[Mesh] OR “Endometrial Neoplasms”[Mesh] OR ((ovarian OR breast OR colorectal OR colon OR endometrial) AND (cancer OR cancers OR carcinoma OR neoplasms)))

NOT

(“Ovarian Neoplasms/drug therapy”[Mesh] OR “Ovarian Neoplasms/therapy”[Mesh] OR “Breast Neoplasms/drug therapy”[Mesh] OR “Breast Neoplasms/therapy”[Mesh] OR “Colorectal Neoplasms/drug therapy”[Mesh] OR “Colorectal Neoplasms/therapy”[Mesh] OR “Endometrial Neoplasms/drug therapy”[Mesh] OR “Endometrial Neoplasms/therapy”[Mesh])

3. “Estrogen Replacement Therapy”[Mesh] OR “Hormone Replacement Therapy”[Mesh] OR “Estrogens”[Mesh] OR “Progestins”[Mesh] OR estrogen\* OR progestin\* OR “hormone

replacement” OR antidepressant\* OR eszopiclone OR clonidine OR methyl dopa OR bellergal OR gabapentin OR pregabalin OR isoflavone\* OR “red clover” OR “black cohosh” OR cimicifuga OR “st. johns wort” OR ginseng OR flaxseed OR “vitamin E” OR “dong quai” OR “Dehydroepiandrosterone”[Mesh]) OR “Androgens”[Mesh] OR DHEA OR dehydroepiandrosterone OR “androgenic agents” OR “androgenic compounds” OR androgen\*

4. Subset: Systematic Review OR Publication Type: Randomized controlled trial OR (“placebo-controlled” OR (placebo AND (control OR controlled))) OR (observational OR cohort OR “case-control” OR “cross-sectional”)

5. Limits: English, Human

((1 and 2) AND 3) AND (4 and 5)

## **EMBASE Searches**

#6 ‘menopause’/exp OR menopausal OR post-menopause’/exp OR ‘postmenopause’/exp OR ‘post-menopausal’ OR postmenopausal OR climacteri\* OR ‘perimenopause’/exp OR ‘peri-menopause’ OR ‘peri-menopausal’ OR perimenopausal AND [humans]/lim AND [english]/lim AND [embase]/lim

#7 ‘estrogen replacement therapy’/exp OR ‘drug therapy’/exp OR ‘hormone replacement therapy’/exp OR estrogen\* OR progestin\* OR ‘hormone replacement’/exp OR antidepressant\* OR ‘eszopiclone’/exp OR ‘clonidine’/exp OR ‘methyl dopa’/exp OR ‘bellergal’/exp OR ‘gabapentin’/exp OR ‘pregabalin’/exp OR isoflavone\* OR ‘red clover’/exp OR ‘black cohosh’/exp OR ‘cimicifuga’/exp OR ‘st johns wort’/exp OR ‘ginseng’/exp OR ‘flaxseed’/exp OR ‘vitamin e’/exp OR ‘dong quai’/exp AND (DHEA OR dehydroepiandrosterone OR “androgenic agents” OR “androgenic compounds” OR androgen\*) AND [humans]/lim AND [english]/lim AND [embase]/lim

#8 #6 AND #7

#9 ‘meta-analysis’/exp OR ‘metaanalysis’/exp OR ‘systematic review’/exp AND [humans]/lim AND [english]/lim AND [embase]/lim

#10 #8 AND #9

## **Remove Medline References from EMBASE Searches**

#11 ‘menopause’/exp OR menopausal OR ‘post-menopause’/exp OR ‘postmenopause’/exp OR ‘post-menopausal’ OR postmenopausal OR climacteri\* OR ‘perimenopause’/exp OR ‘peri-menopause’ OR ‘peri-menopausal’ OR perimenopausal AND [humans]/lim AND [english]/lim AND [medline]/lim

#12 ‘therapy’/exp OR ‘therapeutics’/exp OR ‘estrogen replacement therapy’/exp OR ‘drug therapy’/exp OR ‘therapeutic use’ OR ‘hormone replacement therapy’/exp OR ‘complementary therapies’/exp OR ‘estrogens’/exp OR ‘progestins’/exp OR estrogen\* OR progestin\* OR ‘hormone replacement’/exp OR antidepressant\* OR ‘eszopiclone’/exp OR ‘clonidine’/exp OR

'methyldopa'/exp OR 'bellergal'/exp OR 'gabapentin'/exp OR 'pregabalin'/exp OR isoflavone\* OR 'red clover'/exp OR 'black cohosh'/exp OR 'cimicifuga'/exp OR 'st. johns wort'/exp OR 'ginseng'/exp OR 'flaxseed'/exp OR 'vitamin e'/exp OR 'dong quai'/exp AND [humans]/lim AND [english]/lim AND [medline]/lim

#13 'meta-analysis'/exp OR 'metaanalysis'/exp OR 'systematic review'/exp AND [humans]/lim AND [english]/lim AND [medline]/lim

#14 #11 AND #12

#15 #13 AND #14

#16 #10 NOT #15

#17 #8 NOT #14

### **Vasomotor Symptoms**

#18 'hyperhidrosis'/exp OR 'hot flashes'/exp OR 'vasomotor symptoms' OR 'night sweats' OR 'sweats OR flushes AND [humans]/lim AND [english]/lim AND [embase]/lim

#19 #17 AND #18

#20 #19 AND 'randomized controlled trial'/de

### **Sleep Disturbance**

#22 dyssomnia\* OR 'sleep initiation and maintenance disorder' OR 'sleep initiation and maintenance disorders'/exp OR 'insomnia'/exp OR 'sleeplessness'/exp OR 'early awakening' OR 'somatic complaints' AND [humans]/lim AND [english]/lim AND [embase]/lim

#23 #17 AND #22

#25 randomized OR randomised OR random

#26 #23 AND #25

### **Psychological Symptoms**

#28 'behavioral symptoms'/exp OR 'mood disorder'/exp OR 'mood disorders'/exp OR 'irritability'/exp OR 'depression'/exp OR 'despair'/exp OR 'anxiety'/exp OR 'difficulty concentrating' OR 'over-reacting' OR forgetfulness OR 'reminiscence lapses' OR 'mood swings' OR 'temper swings' OR 'emotional flare-ups' OR weepiness AND [humans]/lim AND [english]/lim AND [embase]/lim

#29 #17 AND #28

#30 #25 AND #29

### **Urogenital Atrophy**

#31 'female urogenital disease' OR 'female urogenital diseases'/exp OR 'urogenital system'/exp OR 'urogenital disorder' OR 'urogenital disorders'/exp OR 'vulvar atrophy' OR 'vaginal atrophy'/exp OR 'vulvovaginal atrophy'/exp OR 'urinary atrophy' OR 'genital atrophy' OR 'urogenital atrophy' OR 'atrophic vaginitis' AND [humans]/lim AND [english]/lim AND [embase]/lim

#32 #17 AND #31

#33 #25 AND #32

### **Sexual Dysfunction**

#43 'sexual dysfunction'/exp OR 'sexual dysfunctions' OR 'libido'/exp OR 'female sexual dysfunction'/exp OR 'female sexual dysfunctions' OR 'hypoactive sexual desire disorder'/exp

OR 'sexual function'/exp OR 'sexual desire' OR 'sexual satisfaction'/exp AND [humans]/lim  
AND [english]/lim AND [embase]/lim

#44 #17 AND #43

#45 #25 AND #44

### **Quality of Life**

#46 'quality of life':ti OR 'well-being':ti =45417

#47 #17 AND #46 =55

#48 'quality of life':ab OR 'well-being':ab AND [humans]/lim AND [english]/lim AND  
[embase]/lim

#49 #17 AND #48

#50 #47 OR #49

#51 #25 AND #50

### **Osteoporotic Fractures**

#52 'osteoporotic fracture'/exp OR 'osteoporotic fractures'/exp OR ('fracture'/exp OR  
'fractures'/exp AND ('osteoporosis'/exp OR osteoporotic)) AND [humans]/lim AND  
[english]/lim AND [embase]/lim

#53 #17 AND #52

#54 #53 AND 'systematic review'/de

#55 #53 AND 'meta analysis'/de

#56 'meta-analysis'/exp OR 'metaanalysis'/exp OR 'systematic review'/exp AND [humans]/lim  
AND [english]/lim AND [embase]/lim

#57 #53 AND #56

#58 #54 OR #55 OR #57

### **Colorectal Cancer**

#59 'colorectal neoplasms'/exp OR (colorectal OR 'colon'/exp OR intestinal AND  
'neoplasm'/exp OR 'neoplasms'/exp OR 'cancer'/exp OR 'carcinoma'/exp) AND  
[humans]/lim AND [english]/lim AND [embase]/lim

#60 #17 AND #59

#61 #56 AND #60

### **Ovarian Cancer**

#61 'ovarian'/exp AND ('neoplasm'/exp OR 'neoplasms'/exp OR 'cancer'/exp OR  
'carcinoma'/exp) AND [humans]/lim AND [english]/lim AND [embase]/lim

#62 #17 AND #61

#63 #62

### **AMED database for CAM**

(menopaus\$ or climacter\$ or perimenopaus\$).mp. [mp=abstract, heading words, title]

## **Application of U.S. Preventive Services Task Force Grading for randomized controlled trials.**

Good: Meets all criteria outlined below.

Fair: Generally comparable groups are assembled initially but some question remains whether some (although not major) differences occurred with follow-up; measurement instruments are acceptable (although not the best) and generally applied equally; some but not all important outcomes are considered; and some but not all potential covariates are accounted for. Intention to treat analysis is performed.

Poor: Studies will be rated “poor” if any of the following flaws exists: groups assembled initially are not close to being comparable or maintained throughout the trial; unreliable or invalid measurement instruments are used or not applied at all equally among groups (including not masking outcome assessment); and key covariates are given little or no attention. Intention to treat analysis is lacking.

### Criteria

- Initial assembly of comparable groups (potential covariates appropriately distributed)
- Adequate blinding and allocation concealment
- Maintenance of comparable groups  $\approx < 20\%$  loss to follow-up in each arm
- Measurements equal, reliable, and valid
- Interventions comparable and clearly defined
- Intention to treat analysis
- Other aspects of analyses appropriate (e.g. missing data, covariate adjustment, sensitivity analyses)