Comparative Effectiveness Review Disposition of Comments Report

Research Review Title: *Menopausal Symptoms: Comparative Effectiveness Review of Therapies*

Draft review available for public comment from November 21, 2013 to December 18, 2013.

Research Review Citation: Grant MD, Marbella A, Wang AT, Pines E, Hoag J, Bonnell C, Ziegler KM, Aronson N. Menopausal Symptoms: Comparative Effectiveness of Therapies. Comparative Effectiveness Review No. 147. (Prepared by Blue Cross and Blue Shield Association Technology Evaluation Center Evidence-based Practice Center under Contract No. 290-2007-10058-I.) AHRQ Publication No. 15-EHC005-EF. Rockville, MD: Agency for Healthcare Research and Quality; March 2015. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each comparative effectiveness research review is posted to the EHC Program Web site in draft form for public comment for a 4-week period. Comments can be submitted via the EHC Program Web site, mail or email. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft comparative effectiveness research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the EHC Program Web site approximately 3 months after the final research review is published. Comments are not edited for spelling, grammar, or other content errors. Each comment is listed with the name and affiliation of the commentator, if this information is provided. Commentators are not required to provide their names or affiliations in order to submit suggestions or comments.

The tables below include the responses by the authors of the review to each comment that was submitted for this draft review. The responses to comments in this disposition report are those of the authors, who are responsible for its contents, and do not necessarily represent the views of the Agency for Healthcare Research and Quality.

Source: http://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2051
Published Online: March 5, 2015

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	none	
Peer Reviewer #1	Introduction	none	
Peer Reviewer #1	Methods	none	
Peer Reviewer #1	Results	There are two Table 11s. Below the first Table 11, Table 12 is mentioned twice, one right after each other.	Corrected
Peer Reviewer #1	Discussion	none	
Peer Reviewer #1	Conclusion	none	
Peer Reviewer #1	Figures	none	
Peer Reviewer #1	References	none	

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Appendix	The biggest problem I have with this report is that it appears that the trials involving patients with breast cancer were excluded from it. In my mind this is not appropriate. Some of the trials including breast cancer patients also included patients who did not have breast cancer. A manuscript has been written to demonstrate that the effect of different agents on hot flashes is independent of whether or not patients had a history of breast cancer or not[1]. It also shows that results are independent of whether a patient has taken tamoxifen or not. This excludes a number of trials looking at SSRIs/SNRIs, gabapentin/pregabalin, clonidine, and a progestational agent (megestrol acetate). Of note, gabapentinoids and SSRIs/SNRIs were initially shown to be effective in patients with breast cancer, and then exported to other patients. In addition, much of the clonidine work has been done in patients with a history of breast cancer. Thus, if this report is to be worth much, the data from the these trials should be included. The rationale for excluding patients with breast cancer, in my mind, is about as good as the rationale for excluding patients who happen to have red hair. [1]Bardia A, Novotny P, Sloan J, Barton D, Loprinzi C. Efficacy of nonestrogenic hot flash therapies among women stratified by breast cancer history and tamoxifen use: a pooled analysis. Menopause 2009; 16(3):477-83.	There were two primary reasons to exclude patients with breast cancer: 1) issues of comparability and exchangeability, and 2) scope and quantity of literature. The pooled percentage reduction in hot flash frequencies across a number of agents (e.g., SSRI/SNRI, gabapentin, vitamin E, DHEA, isoflavones, buproprion, and others (e.g. as in Bardia et al. 2009) have been found similar for breast cancer survivors and other women. However, hot flushes among breast cancer survivors can be more frequent and more severe. The later finding together the history of breast cancer, argues that the two populations are not exchangeable and combining them analytically would be introduce clinical heterogeneity. For example, one could not examine hormone therapies in a network analysis that included breast cancer survivors--those women would not have been eligible for hormone therapies. From a perspective of evidence synthesis, a separate set of analyses would have been required. Accordingly, the second issue was practical: including studies of breast cancer survivors would have enlarged the scope of this report beyond the impracticable size it had become. We hope that the exclusion does not diminish the worth of the report. The reviewer's point was also raised by our technical expert panel and recommends a companion report for breast cancer survivors.
Peer Reviewer #1	General		
TEP #1	Executive Summary	1. There is a newer STRAW statement that the authors apparently missed. It's in JCEM 2012 PMID 22344196.	STRAW + 10 stage definitions have been added.
		2. The stated definition of menopause is incorrect. It is not just the permanent cessation of menstruation and ovulation. It is the permanent cessation of menstruation and ovulation DUE TO OVARIAN FAILURE (see Stedman's dictionary and all major textbooks). The definition should be corrected.	This has been corrected.

Commentator & Affiliation	Section	Comment	Response
		3. Page ES-2 Figure. It seems to be incorrect that that you have excluded women with breast cancer, because many of the non-hormone prescription therapy trials were performed specifically on women with breast cancer. Did you exclude those studies? I don't think so. There is a difference between saying the report isn't specifically intended for women with breast cancer, and saying the report excluded studies in breast cancer survivors. The authors state the latter, but I believe that they intended to say the former.	Trials that stated including breast cancer survivors were excluded unless data were reported separately for women with and without breast cancer. We have added text to the ES and discussion regarding no intended applicability to breast cancer survivors.
		4. Page ES-4 Figure-what does "not relevant population mean? Not focused on women with hot flashes? Would include that info on page ES-3 as well as saying a more precise phrase rather than "not relevant population" throughout the figure.	"Not relevant population" included: studies of women with pre-existing conditions (eg, heart disease, lupus, major depression); population dietary studies; studies of women with breast cancer; studies including both pre- and postmenopausal women. A detailed list of these population exclusions is present in the Methods section of the report. Detail has been added to the ES Methods.
		5. Page ES-7. There were differences in results among the trials of SSRIs and the SNRIs. For example, evidence was probably stronger for benefit of venlafaxine than for paroxetine. So, does this sentence mean that all the SSRI and SNRIs were analyzed together in the network analysis, or were they analyzed separately? Separate analysis of the individual SSRIs and SNRI should be commented upon.	We combined SSRI and SNRIs for analysis. In pairwise analyses, the pooled SMD for SNRIs was -0.36 (95% CI: -0.55 to -0.17) and for all SSRI/SNRIs -0.35 (95% CI: -0.46 to -0.24). We have added a comment to the text (not included in the ES).
		6. Page ES-8 please verify that the definitions of high, low, and standard dose estrogen used in table A is mentioned in the ES.	Details from the methods section describing the dose categories and a referral to the appendix on dosing categories was added to the ES under the "Symptom Relief" heading.
		7. Page ES-10. The report mentions the route of estrogen (vaginal, for example), but not for testosterone. Were sexually satisfying episodes more frequent with oral testosterone versus placebo? The text regarding prevalence of sexual problems associated with the menopause transition should include the SWAN study by Avis 2009, PMID 19212271.	There were 8 testosterone studies, 7 using a patch and 1 administering oral testosterone. These details were added both to the table in ES and in the results section of the report. Information from SWAN was added as background.

Commentator & Affiliation	Section	Comment	Response
TEP #1	Executive Summary	8. Page ES-11. The authors mention that sleep disturbances may not be secondary menopausal symptoms, which is certainly true, but then it becomes unclear whether the treatment mentioned in Table F were tested in women who HAD menopausal symptoms in addition to sleep disturbance, or not. The title of the table should be amended to mention "among women with insomnia but without menopausal vasomotor symptoms", or something similar, depending on what the study populations were.	Many authors did not specify if having vasomotor symptoms was an inclusion criterion in their trials, so we were unable to conduct subgroup analyses according to presence or absence of vasomotor symptoms. Among the studies with sleep disturbance as an outcome, 46% specified requiring vasomotor symptoms to participate, 18% did not require vasomotor symptoms, and 36% did not specify. Table F includes trials with results that allowed for pairwise comparisons, regardless of whether vasomotor symptoms were required or not.
		The abstract would strongly benefit from the statement on page ES-11: No studies were identified examining the safety of the compounding practices for hormone therapies. Clinicians will very much want this information; it will be a key finding of clinical relevance from this report.	Added.
		9. Page ES-12. The issue with WHI's study population is not just related to older population, but rather that the WHI study population specifically EXCLUDED women with severe vasomotor symptoms. This should be stated.	The sentence was edited to: "In the Nelson et al report, a majority of evidence was derived from WHI trials, representing an older population without severe menopausal symptoms, but one which overlaps with the population for this review."
		10. Page ES-14 Research Gaps: "These agents are unregulated and safety data may be limited or absent." The authors' actual conclusion was that the safety data ARE limited or absent, not that they may be limited or absent, so the authors should change the wording. Ditto page 146 line 27.	Corrected.
TEP #1	Introduction	Page 1 line 9. See "General" comments regarding definition of menopause.	Corrected.
		Page 1 line 22 see "General" comments regarding newer STRAW reference that should be included.	STRAW + 10 stage definitions have been added.
TEP #1	Methods	Page 10 line 44. See "Executive Summary" comment #3 regarding exclusion of women with breast cancer.	See response to comment.
		Page 23 line 14. I believe that the authors mean to say "venous" thromboembolism, not arterial. This should be fixed throughout to avoid misunderstanding.	"Venous" has been added throughout the report.
TEP #1	Results	**KQ1 vasomotor:** page 59 line 37. Many trials have found up to 40% response of vasomotor symptoms to placebo. Would expand this range and offer citations of studies that have shown a placebo response higher than 25%.	Range has been expanded and citations added.
		KQ1 vasomotor: Table 9. The important finding that 80% of studies were of poor study quality for efficacy of treatment on vasomotor symptoms deserves to be in the executive summary abstract.	The proportion poor quality studies was added in the vasomotor symptom summary, as well as to the summaries of the other five outcomes for KQ1.

Commentator & Affiliation	Section	Comment	Response
		KQ1 vasomotor: Table 12. The citation for the definitions of high, standard, and low/ultralow definitions of estrogen doses should be footnoted, and preferably with a few examples of the more common doses. Otherwise clinicians will have trouble interpreting the table.	A footnote was added directing the reader to Appendix D which contains the dose definitions.
		KQ1 psychosocial: Figure 7 page 70. Do the authors really mean to say "depression comparisons"? I would think many of the studies actually assessed depressive symptoms, not depression by DSM criteria. The relevant table titles and text throughout the report should be carefully reviewed with that in mind, including the key points on page 78.	The reviewer is correct—the term "depression" can be interpreted to mean clinical depression. To avoid confusion, the term "depressive symptoms" has replaced "depression" when appropriate.
		KQ1 sexual function: Table 42 page 81. What does "activity" mean? Number of episodes per week? Number of satisfying episodes per month?	The last column with the estimates is labeled "Mean Difference SSE/4 Weeks" and the footnote defines SSE as 'satisfying sexual episodes'.
		KQ3: Page 128 adverse events summary. Why is liver toxicity of vitamin E not mentioned as an agent-specific toxicity? Vitamin E is permanently stored in the liver (is fat-soluble) so this should be addressed. If there are no long-term studies of vitamin E that have examined liver toxicity, this should be stated.	Liver toxicity of vitamin E was not included, as there is no evidence of an association (Ann Intern Med. 2014 Feb 25. doi: 10.7326/M14-0198. [Epub ahead of print] Vitamin, Mineral, and Multivitamin Supplements for the Primary Prevention of Cardiovascular Disease and Cancer: U.S. Preventive Services Task Force Recommendation Statement. Moyer VA.).
TEP #1	Discussion	**Research Gaps:** Page 146 line 56 The authors mean to say e.g., not i.e. Please fix.	Corrected.
TEP #1	Conclusion	none	
TEP #1	Figures	none	
TEP #1	References	none	
TEP #1	Appendix	none	
TEP #1	General	The report is excellent, clinically meaningful. I had comments about the target population/audience, see the attached comments file.	No response needed.
		Structured abstract: Review methods states "Systematic reviews, cohort, and case-control studies provided evidence." What about RCTs? SSRI//SNRI abbreviation should be defined in abstract.	RCTs provided evidence for KQ1 and the list quoted here provided evidence for KQ2 and KQ3. Abstract was reworded to clarify. SSRI/SNRI is defined in abstract.

Commentator & Affiliation	Section	Comment	Response
		Major comment 1: The authors consistently use the term progestin in instances where they mean progestogen. That is, they didn't limit their searched to progestin, did they? Didn't they consider all progestogens, including progesterone and synthetic progestins? This should be changed throughout the manuscript, tables, and titles of tables wherever it applies. If they searched for all evidence regarding progestogens (meaning synthetic progestins and also progesterone), but they only found evidence regarding progestins, that should be clearly stated, but it appears to me that a substantial portion of the time that "progestins" appears in the manuscript, the actual intended meaning is "progestogens".	The reviewer is correct that we included all progesterone and progestins in our search, thus we agree with changing the manuscript, tables, etc. from 'progestins' to 'progestogens' when appropriate. As the WHI in particular used only progestin, we have retained the term 'progestin' when referring to the WHI alone. We also retained 'progestin' when referring to other trials which only used progestins (which was a majority of the trials).
		Major comment 2: The FDA, the USPSTF, and virtually all major medical organizations recommend that menopausal HT is only indicated for menopausal symptoms, not for fracture prevention. In the section addressing fracture outcomes for menopausal HT, this should be explicitly stated. That is, it's of interest to know the fracture outcome information, but guidelines uniformly recommend AGAINST use of menopausal HT for this indication.	We have added to the ES, "Although evidence concerning potential long-term benefits are included as they are part of the decision-making process, this review did not address use of therapies for those purposes," and a similar statement in the discussion. The section on fractures in KQ is now prefaced by specific reference to the HT not being recommended.
		Major comment 3: The terminology of "hormone therapy" should be changed to "menopausal hormone therapy" where applicable throughout the entire report. There are many types of hormone therapy in this report—even DHEA is part of this report, and I'm sure that most times, the authors mean "menopausal hormone therapy" when they say "hormone therapy."	"Menopausal hormone therapy" has been inserted where applicable.
		Major comment 4: This tremendously useful and important report should be published in a form that is highly accessible and readable (i.e. really brief in length!) to practicing clinicians. All key points sections should be combined into a report and published in a journal like Annals of Internal Medicine, or something similar. (Or simultaneously in ob/gyn and family practice journal as well).	No response needed.
TEP #2	Executive Summary	none	
TEP #2	Introduction	I found this section to be adequate, but unnecessarily dryly written.	No response needed.
TEP #2	Methods	I believe the inclusion & exclusion criteria are appropriate. The reasoning behind the criteria were well described. These aspects of the report will make this review more useful than other, less clearly reported, reviews.	No response needed.
		The search strategies were clearly stated and logical.	No response needed.
		The outcome measures used were appropriate and additionally were clinically meaningful. This is also a strength of the report.	No response needed.

Commentator & Affiliation	Section	Comment	Response
TEP #2	Results	**KQ1:** I believe that the authors have presented the results of the review accurately. The trials reviewed show estrogen to be at least somewhat effective for most, if not all, of the symptoms reviewed. This will be useful information to women who are currently experiencing symptoms. However, I believe it would be useful to the lay reader to know how often trials which found estrogen to be useful for sleep and/or quality of life enrolled subjects who were currently experiencing vasomotor symptoms. Women are quite curious about whether hot flashes are the engine that drives other bothersome symptoms, and they will also be curious about whether the reported effectiveness of estrogen on sleep and QOL is driven by its effectiveness on hot flashes.	No response needed.
		KQ 2 & 3: The amount of detail was appropriate & the findings were clearly stated.	No response needed.
		KQ4: The usefulness of this question could be strengthened by the addition of a brief summary after each section. As it reads now, it is somewhat numbing. Some studies found a sub-group effect, some didn't -- the reader will have to draw their own conclusion, but it would help if the authors would share their understanding of the conflicting results.	The section has been substantively revised to include a tabular summary at the outset, followed by discussion. Overall results were added to give the subgroup findings some context. For example, there may have been a significant improvement in the whole treatment population, but analyses on age subgroups showed that only the younger ages experienced a significant improvement.
TEP #2	Discussion	Finally, a mention of hot flashes as the possible engine driving other symptoms. A very welcome 3 sentences!	No response needed.
		The discussion section is generally good. The attention to the important problem of the widespread use of compounded hormones and the lack of good evidence about these products is very good. The future research section is fine.	No response needed.
		However, in the section about clinical and policy implications, I'm concerned that the authors are inadvertently implying that evidence exists upon which to make clinical and/or policy recommendations to women trying to determine the overall net effect of long-term use for symptoms. I think the authors would agree that the evidence is lacking. However, the sentence below could be interpreted to mean that what's needed is not more research, but more communication. I recommend that the authors take another crack at clarifying this section "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).	We have revised the sentence: "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); and fill evidence gaps that prevent defining that net benefit."
TEP #2	Conclusion	none	
TEP #2	Figures	none	

Commentator & Affiliation	Section	Comment	Response
TEP #2	References	none	
TEP #2	Appendix	none	
TEP #2	General	The report deals with a topic of importance to a significant proportion of middle-aged and older women. The target population is clearly stated. The key questions are relevant to the target population, and are clearly stated.	No response needed.
		Clarity and Usability: Yes. Yes, with the caveat that some sections could be improved by editing for clarity. I believe the report will be useful in the clinical setting, and for the lay reader.	No response needed.
Peer Reviewer #2	Executive Summary	As alternatives to HT for menopausal symptoms are often considered by women with breast cancer, authors should state more clearly earlier and at various points throughout the manuscript (e.g. somewhere in Executive Summary and possibly pg. 5, Population) that studies involving women with breast cancer were excluded (first mentioned pg. 10, Exclusions).	We first mention that women with breast cancer are excluded in the study on page ES-3, under “Data Sources and Selection”. Text reinforcing this issue to both the ES and discussion.
Peer Reviewer #2	Introduction	The introduction is clear and well written. Authors do a nice job of including national guidelines and statements from relevant medical societies, including NAMS, IMS and the Endocrine Society to provide needed context for the topic.	No response needed.
Peer Reviewer #2	Methods	The literature search strategy is logical and clearly explained. The inclusion and exclusion criteria are justifiable. Modifying allowable study types and varying inclusion and exclusion criteria for each key question is appropriate and clearly described.	No response needed.
Peer Reviewer #2	Results	The amount of detail presented in the results section is extensive, but readers may refer to the Executive Summary and the many helpful tables and graphs for a simplified presentation of results. Given the length of this review, including “Key Points” for results was a highly effective way to make key messages explicit and applicable. The caterpillar plots were a particularly effective way to clearly summarize large amounts of information. Study characteristics generally were well described. Several important studies were not included, but they were published after March 2012!	No response needed. Note trials were captured in the updated search.
		When describing the many scales used to measure outcomes, authors should consistently note direction of scale (e.g. authors noted higher Kupperman score represents worse quality of life/pg. 46, but omitted describing direction of scale with SF-36/pg. 47, Beck & Hamilton/pg. 62, and WHQ pg 100).	Direction of all scales for all outcomes have been defined.

Commentator & Affiliation	Section	Comment	Response
		Authors should be careful to use term “progestin” when referring to this general class of hormones, but to describe the specific progestin when indicated. For example, “progestin skin cream doses” written when actual hormone is “progesterone” (pg 72). In tables as well, it is unclear why sometimes a specific progestin is noted (e.g. NETA), whereas in the same table “Progestin” is used with accompanying doses provided (5-60 mg), but specific progestin not named (e.g. pg 40, Table 14/pg 40, Table 32/pg 72, Table 53/pg 96).	The document was reviewed for the use of ‘progestin’, ‘progestogen’, and ‘progesterone’, and appropriate edits were made. If authors provided the specific name of the progestin, this was added in the tables. Not all authors provided this information.
		KQ1 sexual function: Authors should use term “VAGINAL ovule” when describing studies of vaginal DHEA (pg. 87)	Due to column width constraints, “V ovule” replaced “ovule” in the table, and “V: vaginal” was added in the footnote abbreviations list. Also, in the text above the table which describes the study, the term “vaginal ovule” replaced “ovule”.
		KQ2: As evidence supports a lower risk of VTE and gall bladder disease with transdermal as compared with oral estrogen, it is unclear why these studies were not included in the report (pg 111 gallbladder disease; pg 115 VTE).	KQ2 was a review of reviews, and relied on the USPSTF report, which is effectively silent on the matter.
Peer Reviewer #2	Discussion	Authors clearly state the implications of the major findings. Under limitations, Authors may wish to remind readers that women with breast cancer were excluded, so the findings of this review are not relevant to this group of symptomatic women. They also may wish to remind readers that this review is not necessarily relevant for women with primary ovarian insufficiency who experience menopause symptoms at an early age.	We have added text in appropriate places regarding lack of applicability to breast cancer survivors as well as women experiencing early menopause due to ovarian insufficiency.
		Authors would be advised to note a limitation of all extensive reviews such as this one is that they take so much time to complete that it is inevitable that there will be new research that is not included. An analysis of the comparative efficacy of both low dose paroxetine and ospemifene, well-studied agents for menopausal symptoms, both recently approved by the FDA, is noticeably absent, but much of the relevant literature was published after March 2012.	An updated literature search was conducted in January 2014, and results from relevant trials were added to the report, including the ospemifene trials.
		Authors clearly describe research gaps, including knowledge about long term benefits and harms of non-hormonal agents. Their focus on the lack of evidence on compounded HT (safety and efficacy) being used by millions of women (pg. 146, 147) makes an important statement and provides direction for future research and possible healthcare policy.	No response needed
Peer Reviewer #2	Conclusion	none	
Peer Reviewer #2	Figures	none	

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	References	none	
Peer Reviewer #2	Appendix	none	
Peer Reviewer #2	General	This thorough report addresses the comparative effectiveness of available therapies for menopausal symptoms. As the majority of publications over the past decade regarding the menopause have focused on the potential risks and benefits of hormone therapy for the prevention of the diseases of aging, it is critical to have an in depth review of the role of HT and alternatives for menopausal symptoms. As the authors clearly note in Background, the majority of women will experience symptoms at the time of menopause, so a careful comparison of available therapies will result in improved healthcare for women. The conclusion of the report that estrogen has the greatest efficacy compared with other agents for menopausal symptoms is not novel, but it is an important message to share with the healthcare community. This report is clinically meaningful, as given a decade of publications about the risks of HT in generally older, asymptomatic women, very symptomatic woman at the time of menopause are often not receiving the information they need regarding comparative risks and benefits of available options.	No response needed.
		The authors identified the key symptoms of menopausal women, including vasomotor symptoms, quality of life, psychological outcomes (depression, anxiety, global mental health), sexual function, urogenital atrophy and sleep. Although urinary incontinence is a common problem for menopausal women, it was appropriate to exclude from this review of symptoms. This review was strengthened by the authors' inclusion of compounded hormone therapy and concerns regarding limited information on both the efficacy and safety of these products. Long term effects of hormone therapy (HT) also were appropriately identified and reviewed, including breast cancer, gallbladder disease, colorectal cancer, cardiovascular disease (CVD), endometrial cancer, osteoporotic fracture and ovarian cancer. The review focused appropriately on the most widely used and studied treatments for menopausal symptoms, including HT, isoflavones and SSRI/SNRIs.	No response needed.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	General	The authors addressed four key questions, all appropriate to the topic and explicitly stated. Addressing key questions 1 (comparative effectiveness of treatments) and 2 (HT potential benefits and harms beyond symptoms) comprised the majority of the review. Key questions 3 (potential benefits and harms beyond symptoms of non-HT agents) and 4 (analysis of effectiveness and adverse effects among subgroups) are important questions, with in depth analysis limited by a paucity of high quality research.	No response needed.
		The authors describe the challenges in defining their target population. The evidence base for harms of HT comes principally from studies of older women (WHI), whereas the research on efficacy derives from studies of younger women at the time of the menopause transition. Inclusion of some studies of OCPs is a bit confusing in a review of menopause symptoms and treatments, but understandable, given that perimenopausal women often have symptoms while still desiring contraception. Authors should specifically note that this review does not address the benefits and harms of available therapies, particularly HT, for young women with primary ovarian insufficiency and symptoms of menopause prior to age 40 years.	Text has been added to the discussion noting the lack of applicability to women with primary ovarian failure.
		Clarity and Usability: The report is generally well structured and organized. As noted above, use of Key Points and graphs help present main points more clearly in such a long and detailed report. Conclusions are based on well analyzed research and may be used to inform both practice and policy decisions.	No response needed.
Peer Reviewer #3	Executive Summary	none	
Peer Reviewer #3	Introduction	Introduction clear with exception of rationale for specific outcomes in Key Question 3	The Key Questions were defined prior to undertaking the review, with outcomes for KQ3 intended to parallel KQ2.
Peer Reviewer #3	Methods	1. Eligibility criteria justifiable, search strategies logical.	No response needed.
		2. Outcome measure for vasomotor sx should be also expressed in terms of absolute reduction in HF frequency in addition to SMD as that is what patients and clinicians care about (see #2 in general comments)	We have added an analysis that transforms SMDs to a hot flush frequency reduction scale.
		3. Scales for outcome measures appear to be focused on overall QOL measures and components of QOL measures. This is a problem with evaluating effect of treatment on specific outcomes such as sleep. For example, insomnia as assessed by a single question on the Greene climacteric scale may not be a valid or comprehensive measure of insomnia symptoms.	This limitation of the evidence review, some overlapping domains, is acknowledged in the report. When scale domains were clearly distinguishable on clinical basis, analyses were reported separately—e.g., sections on sexual function and psychological symptoms. We consider sleep outcomes broadly as sleep disturbance, not as insomnia.

Commentator & Affiliation	Section	Comment	Response
		4. Could the outcome of urogenital atrophy be more clearly labeled as vaginal dryness or vaginal health complaints...urogenital atrophy may reflect a number of urinary symptoms/conditions which are clearly not meant to be considered here.	We have now addressed these outcomes as urogenital atrophy symptoms, acknowledging there are a variety of complaints and symptoms subsumed under urogenital atrophy.
Peer Reviewer #3	Results	1. Amount of detail appropriate. Characteristics of studies adequate.	No response needed.
		2. **KQ1 vasomotor:** Key message regarding primary outcome, vasomotor sx, in KQ1 not necessarily backed by evidence. Report concludes that there is strong evidence supporting that estrogens of any dose are more effective than any other comparator....however, vast majority (4/5 for high dose, 30/37 for standard dose and 40/46 for low dose) of trials evaluating estrogen rated as poor in quality (would question why in table 20 strength of evidence was not downgraded for estrogen, but was downgraded for some nonhormonal agents) and effects of standard and low dose estrogen (low dose is most commonly prescribed in clinical practice) do not appear to be statistically superior to some nonhormonal treatments. More importantly, even if difference was statistically significant, the difference in absolute terms has questionable clinical relevance and report did not consider results of pertinent trials reported since 3/2012...while the report robustly concludes that estrogen at any dose is superior to comparator agents, this statement is not clearly supported by evidence.	The reviewer raises an important point and one that can be argued from their perspective, or the approach adopted. First, we have added to the report trials updated in the recent search (1/2014). As to the matter of strength of evidence and bias, as noted in the methods we imposed one departure to the downgrading in the presence of a large number of trials (≥10) despite a majority rated poor quality *and* there was no evidence for reporting bias; we continued to downgrade for other domains. We believe the approach justified we justified because ratings of study quality were often hampered by poor reporting. The clear consistency and estimate precision support the lack of downgrading. The conclusions regarding estrogens are well supported by the evidence synthesis of all trials (clearly by magnitude effects and rankings) and also in sensitivity analysis (shown in appendix) of only the good and fair trials.
		3. Findings from studies published or reported since March 2012 would improve quality of trials included in in KQ1. These include VMS outcome Joffe H NAMS 2013 abstract, Guthrie KA NAMS 2013 abstract QOL outcome LaCroix A Maturitas 2012;73:361 Sexual Function outcome Reed SD Obstet Gynecol 2012;119:527; NAMS 2013 abstract Sleep outcome Ensrud K Menopause 2012;19:848, Newton KM NAMS 2013 abstract	An updated literature search was conducted in January 2014, and results from articles meeting inclusion criteria have been added to the database and analyzed. The Joffe, Guthrie, LaCroix, Reed, Newton, and Ensrud articles and abstracts presented updated results from the escitalopram study which was already in our database (Freeman, JAMA 2011). All of the updated results for vasomotor symptoms, quality of life, sleep, psychological symptoms, and sexual function have been added to the database under the original escitalopram study and have been included in the updated analyses
		4. Not clear to me that analyses for KQ2 add to those previously performed by 2012 Nelson AHRQ review.	The Nelson review only included RCTs. Due to the limited data available when considering only RCTs, discussions of observational studies were added to the Applicability sections to provide a broader view of the scientific literature on each outcome.

Commentator & Affiliation	Section	Comment	Response
		5. Stronger rationale for analyses for KQ3 should be provided.	We have revised the section to follow KQ2 more closely. It should be noted that the Key Questions with KQ3 intended to parallel KQ2.
Peer Reviewer #3	Discussion	1. Implications: believe evidence provided in this report does not support the statement that the reduction in VMS with any dose of estrogen (i.e. low dose ET) is clearly superior to that obtained with treatment with nonhormonal agents such as SSRI/SNRIs... if a difference exists, it may not be clinically relevant. report should provide results (difference in daily HF frequency) that are interpretable by patients and clinicians.	In addition to the rankings, nonoverlapping credible intervals, the added transformation of SMDs to hot flush frequencies are consistent with superiority to SSRI/SNRIs. Whether that difference achieves a minimally clinically important difference cannot be completely ascertained and a limitation inherent in the analysis of continuous effect measures. However, given the magnitudes of differences response rates likely follow. The discussion addresses this matter and we have included in the revised introduction to KQ1 results a better guide to interpreting pooled measures in relation to clinically important improvements.
		2. Research gaps: should call for rigorous comparative effectiveness pbo-controlled trials that simultaneously evaluate ET and nonhormonal treatments and include comprehensive assessments of the wide array of menopausal sx using validated instruments.	Added.
		3. Limitations are appropriately noted, but also include that report is not up to date and is missing results of pertinent trials reported since March 2012	An updated literature search was conducted January 24, 2014. Results from relevant literature published through that date have been incorporated into this report.
Peer Reviewer #3	Conclusion	none	
Peer Reviewer #3	Figures	none	
Peer Reviewer #3	References	none	
Peer Reviewer #3	Appendix	none	

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #3	General	1. While report considered trial data on a number of important menopausal sx in addition to vasomotor sx, the report is not current (final literature search only includes articles through March 2012). For example, results of 3 arm placebo controlled trial of low dose estradiol (0.5 mg dose equivalent to 0.3 mg CEE dose and these doses are most commonly prescribed in current clinical practice) and low dose venlafaxine (75 mg XR) that evaluated effect of active trts on outcomes of HF frequency/severity/bother/interference, self-reported sleep measures, psychological symptoms, vaginal health/sexual function are available in abstract form (NAMS 2013). Since this is the only trial that has simultaneously evaluated hormonal trt and SNRIs/SSRIs (2nd most commonly prescribed treatment for menopausal sx) as compared with pbo, its findings are paramount to consider when making clinical recommendations about relative efficacy of hormonal vs. nonhormonal drug treatment. Compared with pbo, both drugs had a similar modest benefit on HF frequency/severity/bother, insomnia sx, subjective sleep quality, etc. These findings are not in agreement with the conclusions of this report.	An updated literature search was conducted in January 2014. Data from trials meeting the inclusion criteria have been added to this report.
		2. Use of SMD. Report expresses effect of treatments on vasomotor sx in terms of SMD which is not interpretable by the vast majority of practicing clinicians prescribing treatments for peri- and postmenopausal women with HF. Unclear to me why efficacy in terms of vasomotor sx could not also be expressed in terms of absolute reduction in HF frequency. Previous AHRQ reviews first authored by Dr. Heidi Nelson on this topic published in JAMA in 2004 and 2006 presented reduction in weekly number of HF or reduction in daily number of HF. This should be possible for the major outcome of vasomotor sx considered in this updated review. Table 11 (provided but not interpreted in report) appears to convert SMD to daily HF reduction and would indicate that reported SMDs translates into daily HF reduction of -3.6 for estrogen vs. -3.0 for SSRU/SNRIs vs. -2.8 for isoflavone vs. approx -2.6 for gabapentin....since all these reductions are within 1.0 HF of each other, they do not appear to be clinically relevant differences for most patients and clinicians and would support a conclusion that estrogen therapies are modestly more effective than nonhormonal trts. The magnitude of the differences in HF frequency in absolute terms should be reported and interpreted.	We have added a transformation of SMDs to hot flush frequencies to allow better clinical interpretation. The difference between the referred to analyses and the current report was not limiting study inclusion criteria to those trials reporting hot flash frequencies. Despite the broader inclusion criteria, the conversion to hot flash frequencies is consistent with reductions described by other investigators including Dr. Nelson. For example, the 2006 JAMA report cites a 1.13 daily HF reduction with SSRIs or SNRIs. The result from our model was 1.17 for moderate to severe HFs. Whether the differences between treatments achieve thresholds sufficient to change clinical decision-making is incompletely informed by the evidence—particularly because of reporting outcomes on a continuous metric.
		3. Target population and audience clearly defined	No response needed.

Commentator & Affiliation	Section	Comment	Response
		4. Key questions clearly stated but Key Question 2 seems redundant with previously published AHRQ review and Key Question 3 needs better rationale as nonhormonal treatments are typically prescribed short term and health outcomes listed are those that have been associated with use of ET or HRT and not use of nonhormonal agents such as SSRI/SNRIs or gabapentin...For example, if concerned about harms of use of SSRI/SNRIs why would one not consider suidality, serotonin syndrome, etc...was the purpose to specifically examine whether nonhormonal agents have similar longterm benefits and harms of ET/HT?...this should be specifically stated.	Although the subject matter may have been addressed previously by AHRQ, this CER is more current. As noted previously, the Key Questions were defined prior to undertaking the review, with outcomes for KQ3 intended to parallel KQ2. We included short-term adverse effects as reported for the target population of this report. A review of evidence for all safety data for the nonhormonal agents was beyond the scope of the review.
		5. Given the lack of comparative effectiveness data (i.e. well designed trials including multiple treatment groups of estrogen, nonhormonal agent, pbo), is it prudent to strongly state that estrogen at any dose is clearly more effective than nonhormonal agents in treating VMS, especially to imply that the magnitude of the differences in effects are discernible and meaningful to patients and clinicians?	The conclusion is consistent with the results of the evidence synthesis and rating strength of evidence. We have addressed in greater depth in introduction to KQ1, the limitations of inferring conclusions concerning clinically meaningful difference.
		Clarity and Usability: Well structured and organized. Exhaustive in scope but not current (as of March 2012). Findings espec for Kq1 outcome of VMS need to be presented in manner interpretable to clinicians and policy makers. For these 2 reasons, report was rated as fair...otherwise, it would have been rated as good.	An updated literature search was conducted in January 2014 and articles meeting inclusion criteria were added to the report.
Peer Reviewer #4	Executive Summary	Page 17 of 397, lines 10-12 … Ovarian hormone secretion doesn't typically diminish gradually; rather, ovarian hormone levels are usually quite labile in perimenopause (and even early menopause.) re: Induced menopause, should state … "Menopause may be induced prematurely (before age 40 years) or early (before age 45 years) through medical interventions (e.g. chemotherapy or radiation) or surgery (e.g. bilateral oophorectomy with or without hysterectomy.)	Changes were made in the executive summary and in the background.
		Also, do you want to more clearly state that menopausal symptoms tend to come on long before the final menstrual period, and state how the findings of this report do or don't apply to women at that phase? It's buried in small parts of the report later on, but might best be discussed further up-front.	Added to the executive summary.
		In Table G, p 28-29/397, with gallbladder disease, consider specifying oral estrogen and oral estrogen/progestin; under venous thromboembolic disease, ideally specify there and in narrative section that risk is likely different between oral and transdermal estrogen therapy; under CHD, clarify that the increased risk is for older women, not women early in menopause (and not women with premature menopause.	We have added to applicability sections in KQ2 potential differences in relative risk in gallbladder and VTE accompanying transdermal administration and footnoted the table as suggested. It is our view that issues concerning age and risks are adequately highlighted.

Commentator & Affiliation	Section	Comment	Response
		On p 30/397: line18, add the word “postmenopausal” before “women of all ages” and on line 22 (23?), add “levels” after “and cholesterol ….. “. On line 47-48, delete “diagnostic” and insert “monitoring” with reference to compounded hormone therapies.	We did not add “postmenopausal” before “women of all ages”. The studies we were referring to did include women of all ages, which is why we could not include them in this report which focuses only on postmenopausal women. “levels” was added after “cholesterol”. “monitoring” replaced “diagnostic”.
		On last line of that same page, I strongly DISAGREE with your statement that “The most important previous gaps in the evidence concerning long-term effects of hormone therapy have been filled.” This is false, uninformed and misleading - particularly to funding agencies - when there are extensive research needs with regard to hormone therapy and menopausal health issues in women.	We have changed “most” to “many.”
Peer Reviewer #4	Introduction	Re: STRAW, I realize you chose the date of March 2012 as your cut-off, but it seems foolish to refer to 1991 STRAW when an update came out in early 2012. (Could you make an exception to 3/12 cut-off and reference that one? There were other places in the report where you referenced 2013 publications…)	The STRAW + 10 definitions were cited, replacing the 1991 STRAW definitions.
Peer Reviewer #4	Methods	On p 41/397, line 21, suggest changing “general medicine” to “internal medicine” – or “internal medicine, primary care and gynecology” if you involved family medicine specialists.	The suggested changes were made
Peer Reviewer #4	Results	Need to distinguish, when citing results of vaginal estrogen trials, whether vaginal estrogen was low dose (for localized, topical effect) or higher dose (for systemic effects), since there are FDA-approved vaginal estrogen products for both, and they have different effects on menopausal symptoms.	Standard dose estrogen by vaginal route included only creams, of which only 3 included a placebo comparison (all others were dose or some other comparator). We did account for dose in analyses, but to subset the vaginal administration trials further could have been problematic.
		KQ1 sexual function: Sexual interest is more commonly referred to as sexual “desire” in women. The terms interest and desire are not interchangeable. The current classification of sexual disorders in women delineates problems of desire, arousal, orgasm and pain. Unless all of your cited studies on sexual function limited their research to “sexual interest”, they would be more appropriately referred to as studies on low sexual desire, or the symptom to “low interest or desire”.	The term “sexual interest” has been replaced by “sexual desire”.
		KQ2: In gallbladder section, suggest being more explicit that only oral estrogen has been shown to have adverse effect.	We have added a footnote to the table concerning transdermal administration. In the Million Women Study the relative risk was lower with transdermal administration, but not eliminated.
		KQ2: On page 146 of 397, suggest adding an additional table presenting CHD events BY AGE group.	Because this was a review of reviews, and age issues have been highlighted, we have elected not to include a table of event rates by age.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	Discussion	Your statement "…extrapolating absolute rates from the WHI samples to the target population of this review is potentially problematic" is a gross understatement. I would suggest, at a minimum, deleting the word "potentially" here. This is really the fundamental limitation of this review, as for the USPSTF report and most hormone therapy reviews prepared for clinicians. Because of a belief that randomized clinical trials trump all other research methods, an extensive body of literature is being ignored and results of the WHI applied to women who are completely different than the ones enrolled in clinical trials. It is very important to mention somewhere in this report the now well-established conclusion that reconciling disparate results between randomized clinical trials, large, well-designed observational studies, and animal studies on postmenopausal estrogen therapy may be condensed down to the issue of timing of initiation with regard to age and the onset of menopause. There is a window of opportunity and a window of risk. Estrogen therapy administered to younger women, earlier in menopause and who are typically symptomatic, offers greater protection – not risk – against heart aging and perhaps cognitive aging. Estrogen therapy administered to older women has greater risk – for coronary heart disease, stroke and dementia. Natural menopause at an average age is not the same as induced menopause at early or premature ages in terms of long-term health risks and the risk-benefit profile of estrogen therapy. Transdermal estrogen is not the same as oral estrogen (in terms of hepatic first pass effects, clotting risk, gallbladder effects, etc.) All progestogens are not the same, particularly with regard to breast tissue effects. While there are not adequate randomized controlled trials to prove all of this, and never will be – particularly since questions about chronic disease and long-term health risk are rarely able to be adequately addressed by RCT's in isolation - there is very extensive non-RCT data which should not be ignored. The purpose of the AHRQ report obviously does not allow referencing all the data of "lesser" quality, but certainly its existence should be mentioned so that clinicians are reminded that there are important clinical considerations beyond those that can be addressed or answered by RCT's.	Potentially has been deleted. Additions made to the review concerning transdermal estrogen administration. That randomized controlled trials are the most valid approach to estimating causal effects is not debated. Observational studies appropriately analyzed can be informative, but in the case of hormone therapy discrepancies with RCT results have been well described and reasons identified. The Hernán et al 2008 analysis of the NHS is a case in point. Still, observational evidence has been included in formulating conclusions of the review. Animal studies were not considered. The reviewer argues in support of the "timing hypothesis" that has been, and continues to be, debated in the literature (note citations to the exchange between Barrett-Conner and Manson in the report's discussion). An exhaustive examination of the "timing hypothesis" was beyond the scope of the review. However, we have added to the applicability section additional material concerning timing and CHD risk; relevant analyses that do not provide substantive support. In addition, we have noted the recently reported finding from the KEEPS trial—designed to examine this hypothesis—albeit using intermediate endpoints, but results that were negative.
Peer Reviewer #4	Discussion	The WHI studied one particular hormone therapy agent (oral, non-human, single dose), in an older population (on average many years past the onset of menopause), with a notable absence of symptoms. Adding stronger warnings about the risk of extrapolating results of currently available RCT's to the whole universe of menopausal women is in order.	We believe a balanced perspective has been provided.

Commentator & Affiliation	Section	Comment	Response
		Suggest removing the reference to “diagnosis” along with saliva monitoring since salivary (and/or blood) hormone level checks are more typically done for monitoring/safety than for diagnostic purposes. (I’m not an advocate for the practice, I’m just clarifying the practice so that the report does not appear off-base …)	The suggested change has been made.
		In “Limitations of the Evidence Base on Other Benefits and Harms” this seems like an ideal place to discuss in more detail the differences in benefit vs risk for heart health if estrogen is started early in natural menopause. The reference to “selection bias and time-varying confounding” is too vague to be helpful.	Our perspective is that a lengthier discussion beyond the scope and that the major concerns have been noted.
Peer Reviewer #4	Conclusion	none	
Peer Reviewer #4	Figures	In figure A, p 18/397, suggest removing “due to natural or surgically induced menopause” since surgically induced menopause can include hysterectomy alone (menses stop but ovarian function may continue) or hysterectomy with bilateral oophorectomy (in which case menopause due to cessation of ovarian activity comes on abruptly.) What you might more optimally consider for the scope of the report is women who are symptomatic with menopausal symptoms at average age interval (then specify the age range and indicate that you’re excluding premature menopause which warrants very different management.)	We appreciate the comment and it is appropriate, but the analytic framework was developed early in the project in consultation with AHRQ and others.
Peer Reviewer #4	References	none	
Peer Reviewer #4	Appendix	none	
Peer Reviewer #4	General	Congratulations on successfully completing the monumental task of reviewing such an extensive body of literature related to the treatment of menopausal symptoms. Overall, the report is clinically meaningful and very important. I have the following questions, concerns or comments directed towards further improvement of the report:	No response needed.

Commentator & Affiliation	Section	Comment	Response
		Lack of mention of the difference between premature menopause and menopause occurring at average ages, and lack of mention that there may be differences between natural menopause and induced menopause (eg from bilateral oophorectomy) in terms of symptoms and clinical decision-making about hormone therapy is a glaring gap. For women with premature menopause, particularly following a history of bilateral oophorectomy, estrogen deprivation is associated with considerable long-term health risks. For these women, treatment with estrogen until around the age of natural menopause appears to be very important for reducing the increased long-term health risks of coronary heart disease, cognitive impairment/dementia, stroke and premature mortality, and should be considered regardless of the presence or absence of symptoms unless absolute contraindications arise. While your task was not intended to review symptom management in women with premature menopause, it will be important to clarify that findings of this systematic review should not be directly applied to women with premature menopause and women with early menopause following bilateral oophorectomy. (As it stands right now, primary care physicians are increasingly withholding estrogen from these women, jeopardizing the health of this large population of women.)	Text has been added to discussion concerning lack of applicability to women experiencing primary ovarian failure.
		Why not include non-medication treatments for menopausal symptoms in this report? (e.g. acupuncture, mindfulness, exercise, hypnosis, etc.) The report in its current form appears biased towards medication management of menopause. Why exclude menopausal symptom treatment trials which enrolled women with a past history of cancer? Many of the well-designed and important trials on non-hormonal symptom management approaches have been performed in cancer survivors. There is adequate evidence to support applicability of these findings to non-cancer populations of symptomatic menopausal women.	Adding non-medication treatments and adding subpopulations of women would have expanded the scope of this already large report.
		Need to substitute the term “progestogen” for “progestin” in most locations where used throughout the report. Progestogen is an all-inclusive term which includes synthetic progestins (e.g. medroxyprogesterone acetate/MPA/Provera, norethindrone) as well as natural progesterone (e.g. micronized progesterone/Prometrium, biochemically identical to human estrogen.) When discussing trials such as the WHI, you may use the term “progestin.” Otherwise, when discussing progestational agents in general, the term “progestogen” should be used. (Progestins and progesterone have very different physiological effects, and so shouldn’t be lumped altogether, as often erroneously occurs in the literature...)	We have changed the manuscript, tables, etc. from ‘progestins’ to ‘progestogens’ when appropriate. As the WHI in particular used only progestin, we have retained the term ‘progestin’ when referring to the WHI alone. Many of the trials used synthetic progestins, so when referring to these trials, we retained the term ‘progestin’.

Commentator & Affiliation	Section	Comment	Response
		Why not include mention of diabetes risk with hormone therapy use?	Diabetes was not an outcome selected for inclusion. As noted in the methods, "Outcomes were identified in consultation with the TEP to capture those most consequential."
TEP #3	Executive Summary	Figure A. Why not include the outcomes in the second foot note in the box itself, to improve clarity (Other benefits: osteoporotic fractures, colorectal cancer). It is unclear why the harms are depicted using a different pathway (swervy line) rather than another branch below the 2 on benefits. This is confusing as drawn. Better to simplify the figure, maybe 2 branches: one for benefits (then divide this branch into the 2 sub branches listed), another for harms. It would appear more balanced.	The wiggly line is convention to designate harms. We have changed the other benefits box as suggested for clarity.
		p ES-3, line 10. The term 'general medicine' is not entirely clear to me. Maybe replace with "Primary care medicine (internal medicine, family practice)" since these categories correspond to actual training programs?	The following clarification was made: "Input was sought from Key Informants representing clinicians (internal medicine, family practice, and gynecology), academicians, researchers, and patients during topic refinement."
		Figure B. Write our PRISMA acronym. Can you add a very concise explanation for why 7,019 were excluded? (e.g., not meeting eligibility criteria?)	PRISMA acronym written out and description added: "A total of 7019 records could be excluded in the first round of screening, because from the title and abstract, the screeners could discern that the articles did not meet one or more of the inclusion criteria relating to: study design, outcome, population, or comparator."
		ES-6, line 1: I like the clarity of this paragraph, but am confused about how you did data synthesis and analysis for the other Key Questions. As written, you only describe your analysis for Question 1. Could you add some brief explanation about what you did (and why) with the data you abstracted for the other questions.	An explanation was added, that because the evidence for the remaining key questions consisted of systematic reviews, observational studies, and a few randomized clinical trials, quantitative syntheses could not be done. Qualitative syntheses were conducted on these key questions.
		Lines 21-22: I like the concise mathematical explanation of the standardized mean difference, but a tiny bit more detail explaining how you determined which measures could be pooled. You include some explanation on lines 25-26, but a bit more on how you did this conceptually would be helpful. e.g., who did the judging for when it was appropriate, what criteria did they use, and what measures were deemed "poolable" vs not? I don't think the software programs used are necessary to include in the ES. I might add how they can be converted to OR in the ES (part of what is described on p 19).	Software references have been removed from the ES. We have added explanation as suggested to the methods section of the report. Although we agree with suggestion regarding NNT and OR conversion, we have not included that conversion in the ES because recent AHRQ guidance is cautious.

Commentator & Affiliation	Section	Comment	Response
		line 47 Funnel Plot. Isn't there controversy on the adequacy of funnel plots for assessing publication bias, especially when there are only a small # of studies used? Did you only use this method when there were greater than some threshold of studies? If so, please include.	Acknowledged, statistical tests for reporting bias are limited, and when invoked for SOE ratings are suspected. We have added that for Egger test a minimum of 10 trials were required. We were circumspect invoking potential reporting bias during the analyses, particularly because the estimation of SMDs from the published results can yield both large and small values owing to lacking an ANCOVA result.
		line 28: typo--insert Period after sentence ends.	Added.
		p ES-8 line 40. Another possibility for their null finding is their use of QoL assessment instruments that were insensitive to detect changes in QoL. As I recall, those studies did not use menopause-specific QoL ratings (though please confirm).	The reviewer is correct concerning instruments, and we have added as potential explanation.
TEP #3		es-12. The discussion of the challenges with standardized effect sizes is interesting and helpful, but moving some of the discussion on how you dealt with these challenges to the methods section would be helpful. Mention of 'other calculations' (line 13) is a bit too vague for this type of report. More detail on the actual calculations, or your thought process guiding these calculations would be good to include.	We have truncated the text to be more appropriate for ES and relegated detail to the methods section.
TEP #3	Introduction	none	
TEP #3	Methods	p. 10, lines 13-14. Did you mean ca or cancer? lines 63 Tamoxifen, alone or in combination with another treatment?	This is a list of search terms, so "ca" was used to capture both cancer and carcinoma. This has been clarified - any trial using tamoxifen, alone or in combination with another treatment, was excluded.
		p12 Header : change to Other Benefits/Harms of Hormones line 48: write out SR at beginning of sentence	The suggested changes were made.
		p 13: Better to change header to improve clarity: Q 3: Other benefits/harms of nonhormones	The suggested change was made.
		p 14. Table headers could be improved. They are very confusing as written.	Table 2 header now reads: "Inclusion/exclusion criteria for agent-specific adverse events of nonhormone therapies" Table 3 header now reads: "Inclusion/exclusion criteria for long term effects (coronary heart disease, stroke, or venous thromboembolism; gall bladder disease; osteoporotic fractures; or endometrial, breast, colorectal, or ovarian cancer) of nonhormone therapies"
		p.19, line 56. Sentence says either OR ... but then only describes OR. Remove either or add NNT etc. or rework the sentence entirely?	This paragraph has been edited for clarification.

Commentator & Affiliation	Section	Comment	Response
		Table 5 on MCID. I like this table a lot and it illustrates the quality and complexity of the work that went into this report. But I have some minor suggestions to improve its clarity. Some of the items listed in the column MCID indicate a change, but others do not. Huntley, Kupperman, Samsa. It appears confusing as to how a final score could be interpreted as a MCID. Also, the abbreviation MCID should appear in the title of the table. Abbreviations (VAS) should be defined below the table.	The table has been edited so that all reflect some measure of improvement. VAS scales have been removed.
		p 20 (or p 52), section on pooling "Appropriateness for pooling was judged on the basis of trial characteristics together with subject matter knowledge." Please add more detail on how this was done. Too much wiggle room here for someone trying to replicate your process.	Similar to the ES this has been expanded.
		p 21 (p 53), line 57. Instrument details... this appears to be a sentence fragment. Can you write it out more clearly? E.g., Below we present details on the specific instruments used to...	This paragraph has been edited for clarification.
		p 22, line 38. No need to redefine EPC at the point.	The suggested change was made.
		p 22, lines 53-57. I don't follow your reasoning. Can you be more explicit on what you are thinking.	The justification for making an SOE exception has been clarified: "We imposed one departure from the SOE domains outlined in. In the presence of a large number of trials ($n \geq 10$), even when a majority of the trials were rated poor quality, risk of bias was assigned medium rather than low. If there were >10 trials with consistent effects, and no suspected reporting bias, we concluded that low trial quality did not justify a lower strength of evidence."
		p 24. Peer Review and public commentary. It is unclear to me why this section appears here. It would make sense to put it up front before the more technical methods.	
TEP #3	Results	For the most part, sufficient detail is presented. However, more detail explaining how you determined which measures could be pooled could be included. both conceptually and logistically. Who did the judging for when it was appropriate, what criteria did they use, and what specific measures were deemed "poolable" vs not? The assessment for publication bias seemed a bit brief. See attached document.	We have added text to the ES and methods noting the inclusive approach adopted, sensitivity analyses performed, and assessment of potential reporting bias.
		p 35. Table 11. the abbreviation SMD does not match the words in the text.	The words now match the abbreviation.
TEP #3	Discussion	none	

Commentator & Affiliation	Section	Comment	Response
TEP #3	Conclusion	Clarity and Usability: Overall it is nicely structured and well organized. The only conclusion I would modify is the following: after mention that estrogen alone does not appear to increase breast cancer risk, also add 'but increase the risk of endometrial cancer.' Else this sentence could be misinterpreted.	The suggested phrase was added.
TEP #3	Figures	none	
TEP #3	References	none	
TEP #3	Appendix	none	
TEP #3	General	This is an outstanding, comprehensive report. The authors have done a commendable job in synthesizing a large and complex body of literature. The questions asked are appropriate and the authors have done a great job trying to pool disparate studies to make it more clinically relevant.	No response needed.
		The phrase Standard effect sizes is used but use the abbreviation for standard mean differences (SMD) is listed alongside those words. This is confusing. It would be more clear to pick one term and use the abbreviation that matches the words used.	We have edited the report so that the phrase 'standardized mean difference' is used consistently.
		Structured abstract, p vi, line 37. Insert in: "effective in relieving" line 39: It is not entirely clear that this sentence pertains to estrogen. Perhaps add the phrase "for estrogen" at the end of the sentence.	The suggested phrase was added.
		line 47, after mention that estrogen alone do not appear to increase breast cancer risk, also add 'but increase the risk of endometrial cancer.' Note that you mention this later "The increased risk of endometrial cancer when using estrogen-only therapies has already been established.76" on page 116.	The suggested phrase was added.