Comparative Effectiveness Research Review Disposition of Comments Report

Research Review Title: *Comparative Effectiveness of Medications to Reduce Risk of Primary Breast Cancer in Women*

Research Review Citation: Nelson HD, Fu R, Humphrey L, Smith ME, Griffin JC, Nygren P. Comparative Effectiveness of Medications To Reduce Risk of Primary Breast Cancer in Women. Comparative Effectiveness Review No. 17. (Prepared by Oregon Evidence-based Practice Center under Contract No. 290-2007-10057-1.) Rockville, MD: Agency for Healthcare Research and Quality. September 2009. Available at: http://effectivehealthcare.ahrq.gov/reports/final.cfm.

Comments to Research Review

The Effective Health Care Program encourages the public to participate in the development of its research projects. Each comparative effectiveness research review is posted to the Web site in draft form for public comment for a four-week period. Comments may be submitted through the Effective Health Care Web site or by postal mail. At the conclusion of the public comment period, comparative effectiveness research reviews are revised by the authors in response to the comments received.

Comments on draft reviews and the authors' responses to the comments will be posted publicly on the Effective Health Care Web site within three months after the final review is published. Comments are not edited for spelling, grammar, or other content errors. The table below includes the response by the authors of the review to each comment submitted for the draft review.

Section	Comment	Response
Executive Summary	I suggest that the researchers have followed the classic failures of the "evidence based" approach to effective treatment, in this case, of "preventive approaches" to primary breast cancer. Now widely publicized (see New York Times, 12/30/08 article by Andrew Pollack) is the opinion that such approaches are a "one size fits all" -- an effort to find the "winning treatment mode" which is then recommended for everyone, but which in actuality benefits a fortunate few. Specifically, tamoxifen is converted in the body to endoxifen by an enzyme CYP2D6. Up to 7% of people have an inactive enzyme, up to 40% have only modestly active levels of this enzyme. The implications of this are profound, the FDA advisory panel recommended 2 years ago that the 2D6 test should be noted.	The CER presents the evidence addressing the key questions. This evidence can be used by stakeholders for clinical applications and for making practice recommendations as appropriate.
Executive summary	PES-5, line 42. Gail-2 score is not defined until much later. Perhaps a footnote or omitting the 2 would help at this time. ES-6, line 5. Remove 'the'	We agree with referring to the Gail-2 model as the "Gail model" and adjusted the tables, figures, and text accordingly.
Executive Summary	Raloxifene in indicated to: "reduce in risk of invasive breast cancer in postmenopausal women with osteoporosis and "reduce in risk of invasive breast cancer in postmenopausal women at high risk for invasive breast cancer." Tamoxifen is indicated to "reduce the incidence of breast cancer in high risk women." The term prevention is not used in the labeled indications. The decision to use the term "risk reduction" was deliberate by FDA. This fact should be strongly considered in revisions made to the report.	To avoid misunderstandings with some of these terms, the title has been changed to, "Comparative Effectiveness of Medications to Reduce Risk of Primary Breast Cancer in Women." These terms are used in relevant sections of the review.
Executive Summary	Recommend title of report be changed to: "Comparative Effectiveness of Agents Evaluated for Reducing the Risk of Primary Breast Cancer in Women," and also further appropriate changes to the text.	The title has been changed to, "Comparative Effectiveness of Medications to Reduce Risk of Primary Breast Cancer in Women."
Executive summary	The inability to determine which women would optimally benefit from chemoprevention. The lack of information about who is most likely to experience adverse affects from the drugs. This last point is the most important one in many ways, especially if we recall that we're treating risk not disease, and folks may be much more willing to incur the risk of an adverse side-effect to treat an illness than they are to reduce the risk of an illness.	These are current limitations of the evidence and have been highlighted in the discussion and future research sections.
Executive Summary	Title of Report: Question the use of terms chemotherapy and prevention. Although any synthetic drug can be labeled "chemotherapy" in general, medical, and oncologic parlance, the term is reserved for anti-cancer agents that conform to one of the established classes of cytotoxic agents. The labeling of two drugs for which the major indications are not oncologic (raloxifene for osteoporosis and tibolone for HRT) and tamoxifen that is the paradigm of hormonal therapy of cancer, as "chemotherapy" with the negative implications that is attached to the	To avoid misunderstandings with some of these terms, the title has been changed to, "Comparative Effectiveness of Medications to Reduce Risk of Primary Breast Cancer in Women." These terms are used in relevant sections of the review.

Section	Comment	Response
	word, unfortunate. Nowhere in the medical literature are these drugs considered "chemotherapy".	
Executive Summary	While overall I found this report exceedingly clear, some specific problems include: 1. p. ES-3, table a, column 2: the line for non invasive cancer reads "more with ral" and is quite misleading, given the title states the table is about benefits. What you are trying to convey is that there are more noninvasive cancers with ral, but it appears to suggest that there is more benefit with ral. I would change the wording to be more clear "more induced with ral." 2. p.ES-4, table b. Why weren't vasomotor symptoms included in this table? Later it is mentioned as a major reason for nonadherence. 2. pES-4, line 16. This section is about side- effects, yet you describe a beneficial effect of tibolone on hot flashes here. Does this belong elsewhere?	1. These tables have been revised to include event rates per 1,000 women years to more clearly indicate the differences between raloxifene and tamoxifen and between medications and placebo. 2. Vasomotor symptoms were not included in the table because the trials reported these outcomes descriptively rather than with risk ratios like the other outcomes in the tables. 3. This finding can be more effectively presented when it is contrasted with results of the other medications which increase hot flashes.
Executive Summary	Page ES-1: In the Background section, first paragraph (lines 3-4), 2008 statistics for breast cancer incidence and mortality estimates are written in the future tense. They should now be changed to the past tense.	These statistics have been updated.
Executive Summary	Page ES-6: In the Applicability section, second paragraph, I believe that the last sentence needs clarification. As written, the sentence leaves the impression that raloxifene, unlike tibolone, results apply to older as well as younger women. However, all of the raloxifene trials have been conducted in postmenopausal women. Of the three agents, only tamoxifen has had extensive experience in both pre- and postmenopausal women.	We agree that the raloxifene trials have only been conducted in postmenopausal women, however, some of these women were in their 30s and 40s (9% under 50 in the STAR trial). The statement that tamoxifen is the only drug evaluated in premenopausal women has been added to this section.
Executive Summary	In the Key Points of the Executive Summary, it would be helpful to include a brief description of the studies or reference the table with the study descriptions. Some readers will likely not read past the Executive Summary and therefore may miss this important information.	It is difficult to describe these briefly in the Executive Summary. A statement, "Trials are described further in the review" has been added where they are first discussed for key question 1 conclusions.
Executive Summary	On page ES3 in the table footnotes, is there a difference between + and x? The definition listed for both is the same. If not, perhaps only one symbol is needed.	These tables have been revised for clarity for the final report.
Executive Summary	On page ES5, under the 3rd bullet for Key Question 4, is information available to specify whether women make decisions based on their perceived vs. their objective breast cancer risk?	No, as detailed further in the report and in the final bullet for key question 4, studies are lacking about treatment choice.
Executive Summary	Under the fourth bullet for Executive Summary Key Question 1, specifying that this text refers to invasive cancer would help clarify this point.	This has been added to the bullet in the executive summary and results section.

Section	Comment	Response
Executive Summary	Under the third bullet for Executive Summary Key Question 1, some discussion about the timeframe for treatment and that this timeframe varied across studies would be helpful.	The bullet in the executive summary and results section has been changed to include this information.
Executive Summary	It would informative in the summaries to include follow-up time when RRs are reported.	This information is detailed in the report itself.
Introduction	"The target population includes women without pre-existing breast cancer, noninvasive breast cancer, or precursor conditions who are not known carriers of breast cancer susceptibility mutations (BRCA, BRCA2, or others)." The description of the target population contradicts Key Question 3. Key Question 3 does include women with precursor conditions (LCIS, atypical hyperplasia).	Correct, this may be confusing. The target population for the comparative effectiveness review is as stated. Inclusion criteria for the trials were slightly different. To resolve these differences, we evaluated trial outcomes for women with and without these precursor conditions.
Introduction	1. P10, line 1. I think you meant to say that oophorectomy (not hysterectomy) decreases risk of breast cancer. To my knowledge, hysterectomy alone does not affect the risk of breast cancer. 2. Line 2. You might also add that these differences also may affect the observed rates of side effects (i.e., hysterectomy and endometrial pathology). 3. Line 12. Write out 4 as you did 2 lines above.	1. Changed to oophorectomy. 2. Trials generally reported rates of hysterectomy and other uterine outcomes based on women who had a uterus, so this concern was adequately addressed. 3. The “4” refers to “4 years” which is a measure—current conventions require this to be numeric rather than spelled out.
Introduction	The use of the phrase "primary prevention" is extremely problematic, as it has been since before the NSABP-1 results were published. "Primary prevention" is a medical term that means to keep a disease from occurring. This is also what the public understands by the term "prevention," since the concepts of secondary and tertiary prevention are not widely understood in the public domain. The problem with the word "prevention" in the context of chemoprevention is that the drugs, if they work in the intended way for someone, reduce that person's risk of getting breast cancer, but do not guarantee that the person won't get the disease. After all, some women on all of the drugs got breast cancer.	To avoid misunderstandings with some of these terms, the title has been changed to, “Comparative Effectiveness of Medications to Reduce Risk of Primary Breast Cancer in Women.” The CER refers to risk reduction rather than prevention throughout the text. The term “primary prevention” is used in the text in referring to the trials to distinguish them from trials of drugs to prevent recurrences in women with breast cancer. This is the term most relevant to the research itself.
Introduction	To quote the last line of the executive summary: "However, use of chemoprevention for breast cancer is low in the United States.” My question is compared to which other countries? I am sure it is used less in other countries? Not sure what this statement quantifies.	We could find no data about other countries, and low usage in the U.S. is based on experts’ experiences rather than evidence.
Introduction	Women need to understand what their own risks are of death as a result of breast cancer and of all the unwanted affects of these drugs before they consider chemoprevention of breast cancer. This review should say that outright. Women think they are most likely to die of breast cancer and we know that their risk of heart disease and stroke for most individuals in the over 50 age group should be a bigger worry.	This point has been added to the future research section.

Section	Comment	Response
Introduction	Page 1: In the last paragraph of the Background section on that page, risk factors are listed. I believe that older age at first birth should be added as a risk factor for breast cancer. It is one of the risk markers in most risk calculators, including the various iterations of the Gail model.	This has been added to the text in the introduction.
Introduction	Page 1: Same comment as on page ES-1 regarding tense of the sentence summarizing 2008 breast cancer statistics.	These statistics have been updated.
Methods	2-3 of the large studies that were included in the meta-analyses were categorized as 'Fair' (Veronesi, Powles); by pooling these fair studies with the good studies, many times the 'signal' gets lost. (i.e., the difference between tam and ral on DCIS, as well as some adverse events). I think it would be helpful to present sensitivity analyses where poorer quality studies are separated from high quality studies.	We added statements to the text to highlight situations in which the meta-analysis differs from single, good quality trials. This occurs rarely, but there are some outcomes where results of the NSABP P-1 trial and the meta-analysis are different, but because the NSABP P-1 trial is large, good-quality, and U.S. based, its results may be more relevant to U.S. populations (e.g., noninvasive breast cancer, cataracts).
Methods	Are the inclusion and exclusion criteria for the review clear? N Women 50 and under are not covered in the research. This should be encouraged.	Most trials enrolled women age 50 and under and this was the mean age of subjects in the tamoxifen trials. Younger women are less represented in the raloxifene and tibolone trials. The point that future research should evaluate outcomes across multiple age groups, including younger women, has been added.
Methods	As stated on pate 9, there was little if any participation by women of color in these trials, so there are no data relevant to racial or ethnic groups other than white women. This is a continuing problem in all trials, and reinforces the concern that this draft not be published as relevant to all women.	This point is also discussed in the future research section. The review does not state that the results are relevant to all women.
Methods	I am sure you are including this, but just don't forget to get a list of all meds and alternative treatments for the tam arm. Seems the CYP2d6 chart changes monthly.	Noted, women using active medications require careful monitoring by their clinicians.
Methods	I found 1 tiny typo- page 8, line 13. Either "we drew" or "were drawn" should be deleted.	This refers to our use of a simulation method to calculate number needed to treat or harm. This is usually described as “drawing” random samples from normal distributions.
Methods	Is there no information on the baseline risk for smokers vs nonsmokers and if any of the drugs affect the risk for breast cancer differently in these two groups?	This information was generally not available in the published papers, and outcomes were not reported by smoking status.
Methods	It is not clear why observational studies are ignored in answering the first Key question, particularly since such studies have produced contrary information on the risk of breast cancer in tibolone users (see the Million Women Study).	Women using tibolone for relief of menopausal symptoms in observational studies differ significantly from women randomized to tibolone in primary prevention trials. The Technical Expert Panel for this comparative effectiveness review advised the researchers to use only randomized controlled trials to determine efficacy or effectiveness. The point about the Million Women Study has been added to the discussion for context.

Section	Comment	Response
Methods	It is not clear why tibolone is included in this review, since that drug is not approved for the market in the U.S. and it appears that the likelihood of approval is getting smaller, not larger. At the recent San Antonio Breast Cancer Symposium there was a presentation of the LIBERATE trial bone sub-study, looking at tibolone on breast cancer recurrence, particularly in a subgroup of patients followed for bone mineral density. The trial closed early because of an increased risk of breast cancer recurrence on tibolone compared to placebo. Here's the link to the abstract, which is number 66: http://www.abstracts2view.com/sabcs/view.php?nu=SABCS08L 244	Tibolone was included for completeness. The LIBERATE study has been reviewed by our research team. The LIBERATE trial was designed to investigate whether tibolone is effective and safe to use in women with a history of breast cancer, which is outside the scope of this review. This study is discussed in the text for key question 2.
Methods	Meta-analyses: It's clear from the discussion of the studies relied upon that they are not really comparable for a number of reasons. Doesn't using meta-analysis just cover up the limitations?	We listed several major differences between trials, such as enrollment of dissimilar groups of women and different treatment and follow-up times. However, these differences mainly exist *between* tamoxifen and raloxifene trials. The four placebo-controlled tamoxifen trials are reasonably comparable to each other and can be reliably combined. The two raloxifene trials are also similar enough to combine. We did not attempt to combine trials of different drugs. The review includes extensive figures detailing the outcomes of each trial separately and combined to avoid covering up any information.
Methods	Search strategy -- p.4 It is certainly appropriate to focus on clinical trials published on MedLine and elsewhere, but in light of the concerns outlined in "remaining issues," it seems uninformative to have ignored all the lay literature about why women do and don't take these drugs, and what their concerns are. Maybe that's outside AHRQ's purview, but it's still seems odd.	These issues are important, however, evaluating information from the lay literature was outside the scope of this comparative effectiveness review. This source of information may be included in other AHRQ reviews.
Methods	The inclusion of tibolone in the analysis is problematic since it is not approved for prevention in the US, was minimally evaluated and the major study cited was stopped for an excess of strokes on the drug.	Tibolone was included for completeness.
Methods	The study criticizes (on p. 9) the Italian tamoxifen study for excluding women who had had hysterectomies, noting that "women with oophorectomies may be at lower than average risk for breast cancer." The statement is true, but not everyone who has a hysterectomy also has ovaries removed. In addition, and maybe more importantly, by excluding women who had had hysterectomies from their study, the Italians avoided any risk of endometrial cancer in the treatment group. The draft does not apply the same critical analysis to the raloxifene trials, which were focused on women with osteoporosis. Women with osteoporosis, as the study notes on page 10, may be at reduced risk of breast cancer because they have lower circulating estrogen levels. Yet, the draft does not criticize the raloxifene studies as it does the Italian study.	We agree that not all women with hysterectomies also have oophorectomies, but it is difficult to sort this out in the Italian trial. This trial was well designed and included in all parts of the comparative effectiveness review. Its quality rating was downgraded because some women in the trial continued to use estrogen--an important risk factor for developing breast cancer. These points are important to note when examining the findings and comparing them to the other trials. Relevant research design and methodological issues were detailed for all the trials—this is a necessary component of comparative effectiveness reviews.

Section	Comment	Response
Methods	Inclusion of Tibolone data may just muddy the waters especially for consumers (patients).	Tibolone was included for completeness.
Methods	It may be helpful to emphasize both to providers and consumers (patients) that data was for the most part extracted from white participants.	This is discussed in the results and conclusions sections, and is now included as an area for more research in the future research section.
Methods	The target population is defined as excluding women with prior noninvasive breast cancer or precursor lesions. Some clarification about the exclusion of precursor lesions would be helpful. Previous studies have shown that women with atypical hyperplasia experienced some of the greatest risk reductions from tamoxifen.	The target population for the comparative effectiveness review is average risk women because it is unclear what their benefit/risk tradeoffs are. Women with precursor lesions have an elevated risk that would be considered differently in making clinical decisions than average risk women. Indeed, they would likely have greater benefit.
Results	"Tamoxifen (RR 0.70; 0.59, 0.82; 4 trials), raloxifene (RR 0.44; 0.27, 0.71; 2 trials),".. reduce invasive breast cancer in midlife and older women by approximately 30% to 68%; tamoxifen and raloxifene have similar effects in the STAR head-to-head trial." A major problem here and throughout the analysis is the absence of conclusions about efficacy in premenopausal women. This is significant since given differences in study design: studies of raloxifene were limited to post menopausal women and the tamoxifen registration trial, P1 stratified by age (50) not menopausal status. Nevertheless in the Royal Marsden and IBIS 1 trials there were analyses by menopausal status and there was benefit to the premenopausal group on tamoxifen as was also the case for tamoxifen patients in P1 in the in the <50 group. Although in the report is the statement, "We detected no significant differences (in efficacy of tamoxifen) between pre and postmenopausal women by subgroup comparison analysis" this does not appear in any conclusion (only the implication that for older women age was not a discriminant) and most importantly is not further discussed when the observation is made that in P1 there was no excess of serious complications of tamoxifen in women under 50.	We agree with these points, but we are limited to reporting data that the trials provide in their publications. Details of our subgroup analysis by menopausal status is provided in Figure 21. The second bullet point in the conclusions for key question 3 indicates that tamoxifen reduces breast cancer outcomes in subgroups based on age, menopausal status, etc ."
Results	A single clinical trial of tibolone is not adequate for inclusion. The population of patients with osteoporosis and the primary outcome of fracture make this trial inadequate for any broad statement about the efficacy of this drug in breast cancer prevention. The LIFT trial is of very short duration and follow-up, and not adequate for a comparable evaluation of tibolone. There is no data in this trial on ER status and many other important variables. I do not think the inclusion of tibolone in this report is well defended. There is only one placebo controlled trial and no discussion is presented on the contrary reports from observational studies on the risk of breast cancer in tibolone users. There is just no enough data to justify including	Tibolone was included for completeness.

Section	Comment	Response
	tibolone along side tamoxifen and raloxifene.	
Results	Another important omission in this section is the omission of age stratification for stroke. Several other analyses (and the trials included) have distinguished between women > vs < 50 and found a higher stroke risk only for older women. Can you perform this subgroup analysis?	We only found stroke results stratified by age (> vs < 50) in the NSABP P-1 trial and could not perform a meta-analysis.
Results	As noted on page 9 at the bottom is the reference to the statement that most women over 60 years of age have a Gail model [score 1.67% or greater]; risk models should help people understand their risk of getting breast cancer over the next 5 years, and most of the models exclude too many people, or don't narrow the number of years over which risk ranges.	Correct, the model is limited. Some of these issues are now addressed in the future research section.
Results	1. Consideration should be given to the presentation of the mortality data with the implied disappointment that there was no decreased mortality related to the drugs. While none of the studies showed decreased mortality on the intervention, none should even a trend toward increased mortality and invasive breast cancer was the primary endpoint of the prevention trials. 2. The inclusion of the fracture data as a positive endpoint is appropriate but is not emphasized as an additional reason for using the drugs in women at high risk for breast cancer.	1. Noted, mortality data indicate no differences between comparators, however, follow-up for some of the trials is likely too short to show this. 2. We tried to present all the benefits and harms in a balanced way. New tables with event rates and numbers needed to treat or harm have been added to the final version that may be useful in considering all the outcomes.
Results	I like the consistent presentation of results by drug (tam vs. ralox, tam vs. placebo, ralox by placebo, tibolone by placebo). It was clear and easy to follow.	Noted.
Results	It would have been much easier for me to evaluate the tables and charts if they had been incorporated into the text document. There were also a lot of black lines going through the document that confused me. I didn't know if you were blocking out information or if this was a design element. It confused me as a reviewer.	Noted; unclear where the black lines came from, these were not present in the pdf we posted.
Results	My main concern with the presentation (a minor one) relates to comparison of rates among subgroups, where it is not clear if the differences are due to differences in baseline risk for that subgroup, or difference in relative risk. Ex, p ES-5, lines 10-11: does tam cause more events in older women because their baseline risk is higher, or because the RR for stroke is higher in older vs younger women? Please always report the RR (CI) for each relevant subgroup to avoid confusion.	We agree, sorting out baseline risks from elevated relative risks is confusing. We have included relative risk estimates for each subgroup in the text when they are calculated. However, trials report these results differently. The subgroup event rates are detailed in the meta-analysis figures.
Results	One conclusion, "All models (of breast cancer risk) have low discriminatory accuracy in predicting the probability of breast cancer in an individual. Most models perform only slightly better than age alone as a risk predictor," is a telling and not widely enough appreciated. The section is good and unbiased.	Noted.

Section	Comment	Response
Results	P11, 25-27. I suspect that if you separated out good from fair trials you might come to a different conclusion on the DCIS question.	This is one of the few situations where the results of the largest, good-quality U.S. trial had results that differed from the meta-analysis of 4 trials (NSABP P-1). The revised version highlights this discrepancy and the strength of evidence was downgraded because of inconsistency.
Results	P17, line 46. This is another area where I think it would be helpful to separate out good from fair studies to see if there is a different signal from the good studies (esp. given the ascertainment issues described, and that the Marsden study included a relatively young population who were at very low risk for stroke).	For stroke outcomes, results of the tamoxifen trials are not significantly different from placebo whether considered individually or combined. Figure 16 provides these details.
Results	P21, lines 43 on down: I am a bit confused by the label of outcomes affecting QOL, given that previously you discussed several outcomes that only affected QOL (such as vaginal symptoms, cystitis, incontinence, on p 19). Are those not QOL issues as well? Perhaps some mention that other QOL issues are dealt with elsewhere.	There are many reasonable ways to organize these outcomes. The key question groups them in this way so related symptoms can be considered together. Most of these also affect QOL. Symptoms that do not fit these groupings but are important for QOL are collected under that title.
Results	P22, line 41. The 10% threshold is mentioned, but it is not clear if this is 10% more than in the placebo group? Nor is it clear if the differences between treatment and placebo were statistically significant. If >10% in treatment group were affected, this could be consistent with rates of 11% in treatment and 9% in placebo, or 50% in treatment and 1% in placebo. More clarity is needed here.	We are limited to reporting data that are included in published papers.
Results	P23, line 17. The term modified Gail model is used but I can't find where it is defined. Is this the same as the Gail 2 mentioned earlier? Could you simply call it Gail and use a footnote to explain (with more detail) what it is you mean, preferably with a reference to which Gail model you refer. line 38. Typo- remove the stray '.' before 39.	These terms have been clarified in the text.
Results	P24, lines 4-9: see above comment about lack of clarity with subgroup comparisons. Be consistent and clear about whether you are trying to communicate that the RR differ across these subgroups (which I think is what you are talking about) vs that the absolute rates are higher. In several of these sentences, the specific RR should be mentioned for each subgroup. The last sentence, where you revert to talking about event rates, it is confusing why you do this, given that the event rate will be largely driven by difference in baseline risk (older age) versus relative risk of treatment.	We reported data available from the published trials. Unfortunately, they described this in various ways. We detail the more rigorous results in Figures 20-24.
Results	P25, line 11. Again report the RR (CI) for each of the BMI groups as it is not clear what you are trying to communicate.	These details are provided in Figure 24 for studies that reported them.
Results	P30, line 6. Please convert the fraction to a percent; this is the fraction of physicians sampled.	This sentence has been changed to read “A mailed survey of 350 physicians indicated that 27% prescribed tamoxifen…”

Section	Comment	Response
Results	The draft says there is mortality data, but most of the trials did not run long enough to actually look effectively at mortality, either from breast cancer, or all causes. And, in "prevention" trials, isn't the most important issue whether you reduce the risk of death from both the disease you are seeking to prevent and the others that may be caused by the treatment?	The point that the trials may not be long enough to capture mortality benefit is included in the text. We tried to present all the benefits and harms in a balanced way. New tables with event rates and numbers needed to treat or harm have been added to the final version that may be useful in considering all the outcomes.
Results	The potential benefit of tamoxifen vs. other agents to the high-risk premenopausal women should be discussed.	Tamoxifen is the only medication currently approved for premenopausal women at this time, and the only medication with trial data from premenopausal women.
Results	The results are at best misleading. Using terms such as "chemoprevention" and the suggestion that women should submit to this sort of "treatment" in ignorance of recent research and the corollary comments by the insurance industry seems to be irresponsible.	The CER presents the evidence addressing the key questions. This can be used by stakeholders for clinical applications and for making practice recommendations as appropriate.
Results	This question is very general in attempting to address concordance, adherence, and persistence. Although by design this study was limited to an analysis of prevention trials (with limited information available) the data for non-adherence to tamoxifen in established cancer patients is extensive, well known, and eye opening and is a major impediment to its efficacy in the real world. The lack of discussion of this data is problematic to its broader application, albeit methodologically justified. Furthermore a discussion of the patients screened for enrollment but not actually enrolled on P1, STAR would have been useful. Including at minimum a statement or references to tamoxifen non-adherence would be beneficial.	As noted, it would not be methodologically appropriate to include studies of women taking tamoxifen to treat breast cancer. We do not have data regarding screen failures for any of the trials. We included one descriptive study reported reasons for women not selecting risk reducing medication for either the tamoxifen or the STAR trial.
Results	1. This was a reasonably straightforward discussion of the adverse events, however, the importance of the observation that tamoxifen-related serious adverse events in younger (premenopausal) women is not increased over placebo is not discussed as offering a different risk benefit ratio than exists for older women. 2. Further, it is implied that tamoxifen is the preferred and only preventative for younger women, raloxifene or tamoxifen for older women.	1. The discussion makes the point that risks and benefits need to be determined on an individual basis. Age is an important factor in considering this risk benefit ratio. 2. Raloxifene is approved only for postmenopausal women of any age.
Results	Page 15 (Key Question 2, Key Points, Bullet #4): The difference in endometrial cancer between raloxifene and tamoxifen was not statistically significant in the STAR trial (95% confidence interval 0.35, 1.08). The wording of the sentence should be changed to reflect lack of statistical significance.	This statement has been changed.
Results	Page 19 (Genitourinary Outcomes, Tamoxifen vs. Placebo): same comment as on page 15.	This statement has been changed.

Section	Comment	Response
Results	On page 11, under the first Key Point for Key Question 1, the authors might consider adding a sentence that most of the tamoxifen trials and the STAR trial included only women with elevated risk by family history of Gail score. Although the fact that women with lower Gail scores were not well represented in studies and the implications of this are noted in the Applicability section, including this briefly in the Results as well would help the reader not to miss this information.	The inclusion criteria for the trials are detailed in the beginning of the results section as well as in Table 2 before the results of the trials are presented so readers will be able to interpret them with this information.
Results	On page 21 of the text, if true, it would be helpful to clarify the statement that raloxifene does not cause cataracts to indicate that findings were not significant. Otherwise it may read as suggesting that no cataracts occurred.	This has been clarified, “no more cataracts than placebo.”
Results	On page 25 under exogenous estrogen use, was information available to examine estrogen with progestin use separately, since this combination has been associated with increased breast cancer risk?	This point is correct, based on results of the Women’s Health Initiative, however, trials did not report outcomes by type of estrogen regimen.
Results	1. Among the studies that were excluded was a study by Abrams which used tissue from NSABP P1 which sought to identify individuals most likely to develop thromboembolic events based on existing factors, i.e. Factor 5 Leiden. It is not clear why it was marked as "wrong population." 2. Also on the "excluded" list is a trial by Veronesi related to tamoxifen use in the adjuvant and in healthy women (at risk of breast cancer).	1. The Abramson study has been included in the revision. 2. The Veronesi study does not meet inclusion criteria because it compares women taking tamoxifen for cancer treatment with women in the NSABP P-1 trial (mixed population of tamoxifen and placebo users) and does not provide data addressing a key question.
Results	Breast cancer incidence was a primary endpoint in RUTH.	Correct, the text has been changed in the results section that describes the primary prevention trials.
Results	For raloxifene, data does not suggest everyone benefits - it excludes premenopausal women.	Correct, this point is emphasized in the text.
Results	It would be informative to report the ER-positive reduction post treatment where available.	This information is detailed in Figure 6 as well as in the results text.
Results	There is inadequate discussion related to the lack of validation of the tools for special populations.	This point has been added to the future research section.
Discussion	Are the limitations of the review/studies described adequately? N Is the future research section clear and easily translated into new research? N	These sections have been expanded.
Discussion	I disagree that the outcome of breast cancer risk is of moderate strength in the case of tibolone.	Noted, the rationale for the moderate rating is described in the text, but there may be other interpretations. We applied criteria based on EPC GRADE.

Source: www.effectivehealthcare.ahrq.gov
Published Online: March 10, 2010

Section	Comment	Response
Discussion	I wouldn't necessarily call it "important literature" but there's no comment about physician attitudes toward or acceptance of breast cancer chemoprevention. There's not much literature available on this but it might be mentioned, as a limiting factor of the use of chemoprevention.	A single survey of physicians describing prescribing practices for risk reduction medications was included in our review. We did not identify other sources of information regarding physician attitudes.
Discussion	The most important aspect of the trials on tamoxifen and raloxifene is the head to head comparison in the STAR trial. More discussion on the need for such trials, proper power calculations, duration of treatment, and duration of observation, as well as outcome variables, would be useful.	These points have been added to the future research section.
Discussion	There are no data available for healthy young women with or without a family history. These are the women who are most likely to want to keep their ovaries and use chemo prevention before choosing surgical interventions. They need to know that there are no relevant data for them.	The mean ages of women in the four major tamoxifen vs placebo trials were 47 to 51 years. These trials enrolled many premenopausal women in their 30s and 40s and results apply to this age group. How these results apply to women in their 20s is not clear. Raloxifene is currently only approved for use in postmenopausal women of any age. These points are highlighted in the text.
Discussion	This draft simply is inadequate to justify any clinical practice recommending or decisions by individuals to take pills to "prevent" breast cancer.	The CER presents the evidence addressing the key questions. This evidence can be used by stakeholders for clinical applications and for making practice recommendations as appropriate.
Discussion	We really don't know much about premenopausal women. That should be stated repeatedly and clearly. Under Summary of Results, harms, q's 2 and 3 you state, "Subgroup analysis indicated that stroke was higher for older >70 women than young women." Women in their 20s & 30s are not included in your subgroups. You really need to be clear about who you are able to draw conclusions about. Is there any literature about any of these drugs with women younger than 50? I did not see a section re: future research. Obviously, I would want to see data on premenopausal women.	See above comment. A call for more studies across multiple age groups has been included in the future research section. The comment about stroke refers to findings from the tibolone trial that was conducted in older women.
Discussion	Page 35 (Precision section): The first part of the paragraph states, "Precision is the degree of certainty surrounding an estimate of effect for specific outcomes. We considered estimates precise if they provided statistically significant differences between drugs, or between drugs and placebo, for major clinical outcomes that would support clinical decisions." I think that this leaves the impression that statistical significance is equivalent to precision of effect size estimate. Additionally, the last part of the sentence is closer to the concept of "conceptual confidence" than it is to either statistical significance or estimate precision. The next sentence goes on to say, "Estimates were also considered precise if they showed no statistical significant differences between comparators, and confidence intervals did not range beyond 0.50 to 1.50." (As an aside, this is an asymmetric confidence interval.) This latter sentence is closer to the usual meaning of precision in	The CER applies the definition of precision as defined for the EPC GRADE table. This methodology is detailed in the Appendix. Although GRADE does not provide operational guidance, it emphasizes the need to include both clinical and statistical considerations. We believe it is reasonable to consider both 1.) conceptual confidence in the clinical importance of an outcome, and 2.) statistical precision of effect estimation (including statistical significance). For this CER, we considered estimates precise if the outcome indicated a statistically significant difference between comparators. We accepted this as adequate evidence that this level of statistical precision can be used to support clinical decisions. If an outcome was not statistically significantly different from a comparator, we then

Section	Comment	Response
	estimation. In other words, the paragraph is conflating hypothesis testing (i.e., statistical significance) with conceptual confidence in the clinical importance and then with precision of effect estimation.	examined statistical precision. We clarified this in the text. In this CER, we do not have comparisons where the results are statistically significant, but the confidence intervals are so wide to be considered imprecise because we have meta-analysis of several large trials. We agree that 0.50 to 1.50 is an asymmetric interval. We have changed it to a symmetric interval (0.67 to 1.50).
Discussion	Page 35 (Strength of Evidence, line 7): The paragraph suggests that the evidence is strong that tamoxifen does not decrease the incidence of noninvasive disease. However, I am not convinced that the evidence is strong. I don't think it's consistent, even though Table 1 states that there is no inconsistency among the 4 placebo controlled trials addressing this endpoint. The Breast Cancer Prevention Trial (BCPT) showed a statistically significant decrease in non-invasive cancers according to both your Figure 6 and to the 2005 BCPT paper in JNCI (RR=0.68; 95% CI = 0.45 to 0.89), while the other trials that reported non-invasive cancer did not. The meta-analysis did not show a statistically significant decrease in noninvasive cancers, but the lack of consistency would, in my opinion, downgrade the strength of evidence to moderate rather than strong.	We re-evaluated the strength of evidence based on our refined definition of precision and the valid point about inconsistency within the tamoxifen trials. Our rating dropped to "low" strength of evidence based on these criteria. The text and tables have been changed to reflect this.
Discussion	Page 37 (Summary of Results--Benefits, end of the first paragraph): The number needed to treat (NNT) is the inverse of differences in absolute rates. Since absolute rates and absolute rate differences change over time, the time frame for the NNT should be provided.	We assumed that women take the drug for 5 years and NNT estimates were calculated for a 5-year period. This information has been added to the text.
Discussion	On page 35 under Strength of Evidence (SOE), criteria for high and moderate Strength of Evidence are described and Appendix C-3 is referenced. This Appendix provides the interpretation for each level of SOE. We would also suggest adding to this table the review team's criteria for each level (e.g. low risk of bias, consistency, precision, etc).	This information has been expanded in the text.
Discussion	On page 35 under Strength of Evidence, some clarification about how STAR data are being used to "confirm" placebo trial data would be helpful. This seems somewhat unclear given that the placebo trials indicate whether and effect was present for a drug such as tamoxifen, while the STAR trial indicated whether the effect differed significantly from that of raloxifene.	These points have been clarified in the text.
Figures	Figure 24. Calibrations of Breast Cancer Risk Models - difficult to discern - many abbreviations.	We tried to reduce abbreviations in the final version; all abbreviations are defined in the figure.

Section	Comment	Response
Figures	Figure 25. Discriminatory Accuracy of Breast Cancer Risk Models - better summarized in paragraph format.	The figures are provided to readers who desire this level of detail; most readers may prefer the text description.
Tables	Please include tables in the body of the text for the draft review. It is too difficult to figure out what goes where.	Noted.
Tables	Table 11: the row for tibolone. This may reflect my ignorance, but it is not clear to me why the tibolone trial is rated so highly given it is a single study in a limited population with a not so large N.	Tibolone results for the overall strength of evidence have been revised based on a number of criteria highlighted in the GRADE Appendix and text.
Tables	Table 4: in the cell for STAR trial and DCIS, there are 2 rows of data. Why? This is confusing.	The second row refers to DCIS specifically, this has been clarified.
Tables	Endometrial cancer is listed as not reported from the LIFT trial in Table 5, although it is listed as reported in Table 3, and in the text on Key Points on p. 15.	The LIFT trial reported cases (0 vs. 4, p=0.06), not risk ratios, so it was not included in Table 5 with all the other risk ratio results. LIFT results have been added as a footnote to Table 5.
Tables	Table 3 lists ER+ and ER- outcomes as not reported in the STAR trial, although results are given for these in STAR in Table 4.	This has been corrected.
Tables	P1 did not allow the use of hormone replacement during the trial. There was a wash out period for enrollees.	Correct, Table 2 now includes this information.
References	No comments submitted.	
General	Can you mention the several prevention trials underway examining the impact of aromatase inhibitors and retinoids on breast cancer risk (perhaps in the future work section?)	This has been added to the future research section.
General	1. In summary, I would not include tibolone in this evaluation of the primary question as the data on this drug is too limited and contrary. 2. I would devote some effort to structuring a way forward, rather than dropping it at the end. Was there any attempt to contact those involved in the reported trials to see if there was further information available? What further analyses could be done that would address the key questions, and have the trial groups been asked to undertake such analyses?	1. Tibolone was included for completeness. 2. The CER researchers contacted the principal investigators of all the trials to determine if additional analyses were underway and to obtain data to specifically address the key questions. They provided no new data for the CER. The tamoxifen trial investigators indicated new analyses using results from all 4 trials are underway.
General	In the description of the MORE and CORE trial, my understanding was that there was a gap of over 10 months between CORE and MORE when participants received no treatment. If this is correct, this might be worth mentioning in the sections summarizing the trials.	This was the mean time; this detail has been added.

Source: www.effectivehealthcare.ahrq.gov
Published Online: March 10, 2010

Section	Comment	Response
General	1. The "future research" section is inadequate. 2. Potential approaches to primary prevention of breast cancer are of vital importance, particularly for high risk women. The report could outline the next reasonable steps to actual rational clinical practice. In the absence of such, the drugs will be used too arbitrarily and without creating useful results for further decision making. The conclusion of adverse events dictating decision making, is precisely the point about the need for a rational, step by step outline of approaches that could provide definitive information on primary prevention.	1. The future research section has been expanded. 2. The CER presents the evidence addressing the key questions. This can be used by stakeholders for clinical applications and for making practice recommendations.
General	1. The report needs to state often that you are discussing chemoprevention and not treatment for cancer. 2. The risks and unwanted effects seem extensive enough to warrant caution that when using these drugs as chemoprevention, this population is trading one disease for another. That needs to be stated clearly to the community.	1. To avoid misunderstandings with some of these terms, the title has been changed to, "Comparative Effectiveness of Medications to Reduce Risk of Primary Breast Cancer in Women." Risk reduction is emphasized throughout the review. 2. The CER presents the evidence addressing the key questions. This can be used by stakeholders for clinical applications and for making practice recommendations as appropriate.
General	The report stops short of providing much clarity about future work in this very important area for clinical cancer research.	The future research section has been expanded.
General	To summarize the scientific/clinical analysis of the study, it is accurate but the choice of wording and what is not discussed while perhaps accidental does seem prejudicial against the use of tamoxifen or raloxifene. Particularly from the point of view of a general physician or a high risk woman, after reading the report, without knowledge of the totality of the data, they would be very unlikely to choose preventative SERM. Finally missing from this analysis is context. How do the preventative effects of the SERMs measure up to that of stains or antihypertensives and CV events? Including comparisons of numbers needed to treat to prevent data, would have been helpful in putting this is context. Although beyond the scope of the study, a cost effectiveness analysis (with a comparison to that of statins/antihypertensives) would also have been useful.	These issues are important but extend beyond the scope of the comparative effectiveness review. Some of these points have been added to the future research section. The review presents the evidence addressing the key questions. This evidence can be used by stakeholders for clinical applications and for making practice recommendations as appropriate.
General	Finally, I always like to see a Figure of the Analytic Framework ("Causal Pathway") in these systematic reviews, as outlined in the methodology papers of the USPSTF. I did not see one here.	An analytic framework has been added to the final version.
General	As mentioned in the *Draft* patients are as much influenced by concerns about harms of treatment as risk of breast cancer. I believe the presentation format of potential adverse events can alter the patient's or provider's perception of that event. Thus, if I tell a patient twice as many women get uterine cancer with tamoxifen than placebo it seems more ominous than if I say the incidence is increased only by 1 in several	Tables of event rates per 1,000 women years and estimates of number needed to treat have been added to provide absolute estimates of risk that are more useful clinically than risk ratios.

Section	Comment	Response
	hundreds and those are early cancers. The adverse events seem to be described in ways that accentuate their probability.	
General	From a provider standpoint I found the references and outline comprehensive and well organized.	Noted.
General	If part of the target audience is in fact the consumer, summaries of the data are important. The Executive Summary is most helpful in this regard. I do not think the average busy primary care provider wants to know in general the level of detail found in for example the Methods section; but may want to know where he/she can obtain this information. The Contents section makes this acquisition easy.	Dissemination products for clinicians and consumers are being prepared based on the results of the comparative effectiveness review.
General	It may be difficult to cater to the needs of a consumer and medical professional in the same document.	We agree, dissemination products for clinicians and consumers are being prepared based on the results of the comparative effectiveness review.
General	One could argue that the Italian cohort was informative and does reflect an important population. Rates of hysterectomies were very high a few decades ago and more commonly performed among special populations.	We agree, this trial was included in all relevant sections of the review and is especially applicable to the many women with prior hysterectomies.
General	Prevention research uses chemoprevention as the term to reflect both natural and synthetic agents. I would therefore recommend that the title reflect this use, Comparative Effectiveness of Chemoprevention Agents in the Prevention of Primary Breast Cancer..."	To avoid misunderstandings with some of these terms, the title has been changed to, “Comparative Effectiveness of Medications to Reduce Risk of Primary Breast Cancer in Women.”
General	The primary endpoint for invasive breast cancer in the RTCs is the reduction in incidence of breast cancer. This should be used instead of preventing breast cancer.	To avoid misunderstandings with some of these terms, we are referring to the outcomes as reduction in risk rather than prevention.
Appendix	In general, I did not feel it was necessary for the experts to give reasons for exclusions for each article/study.	The excluded studies section is a required element of the comparative effectiveness review.