OPERATOR: Good day, ladies and gentlemen, and welcome to AHRQ’s Effective Health Care Program, Applying Existing Evidence to Breast Cancer Prevention and Diagnosis. If you get disconnected at any time from the Web conference, you may dial 888-632-5065 or 201-604-0318 to be reconnected to the audio portion.

At this time, it is my pleasure to turn the floor over to Debbie Rogal.

DEBBIE ROGAL: Good afternoon, ladies and gentlemen. Thank you for standing by.

On behalf of the Agency for Healthcare Research and Quality, also known as AHRQ, welcome to today’s Web conference, Applying Existing Evidence to Breast Cancer Prevention and Diagnosis, held by AHRQ’s Effective Health Care Program. My name is Debbie Rogal, and I am a contractor for AHRQ’s Office of Communications and Knowledge Transfer, and I will be moderating today’s event. This event is part of a series of Web conferences we are holding over the next 2 weeks about Effective Health Care Program research findings on different clinical topics, so we are especially happy you were able to join us today and hope that you will join us for future events.

Before we get started, I want to review some information about the Web conference technology.

If you have questions during the presentation, you may submit them electronically by entering them via the “Ask Question” button. The “Ask Question” button is located at the bottom of your screen. When you click on the button, a box will appear requesting that you enter your question. Once completed, press the “Submit” button. A selection of audience-submitted questions will be addressed during the moderated Q&A session at the end of the Web conference, though I encourage you to ask your questions throughout the event as you think of them.

Also, if you are experiencing technical difficulties, there are a few options. You can click on the “Click Here for Web Conference Help” link and be directed to a troubleshooting Web site to do a system check, or you can open the “Web conference FAQs” document under the “Downloadable Files” button on the bottom of your screen for other troubleshooting ideas. You can also contact technical support by submitting your issue in the “Ask Question” box.

Under the “Downloadable Files” button, you will also find the slides for this event, which may be helpful for reviewing slide details, and a document with speaker biosketches.
Today’s Web conference includes closed-captioning. The captioning appears in a box adjacent to the slides through the Windows Media Player or RealPlayer you selected when entering the event.

Finally, this presentation is being recorded and will be made available on the AHRQ Web site shortly.

Let’s start with the presentations.

During the Web conference, Effective Health Care Program investigators Drs. Heidi Nelson and Karen Schoelles and I will highlight the benefits of using patient-centered outcomes research in clinical decisionmaking. I will kick off the conference by giving you a brief overview of AHRQ’s Effective Health Care Program before I turn it over to our clinical researchers to share their findings on prevention and diagnosis of breast cancer.

Patient-centered outcomes research, also known as comparative effectiveness research, delivers unbiased, practical, evidence-based information to help you and your patients weigh different options to make the most informed health care decisions. It compares drugs, devices, procedures, tests, and methods of health care delivery. Patient-centered outcomes research shows which treatments have been shown to work best in different clinical situations and how they compare when it comes to benefits, harms, and side effects, and also tells us what is known and what is not. Most importantly, patient-centered outcomes research is descriptive, not prescriptive. It does not tell you how to practice medicine. It does not mandate a particular test or treatment for anyone, nor does it prohibit any test or treatment. It gives you tools, not rules, that you and your patients can use to make the best possible decisions.

Here at AHRQ, the investment in patient-centered outcomes research has been built around the framework displayed here. The colored boxes and ovals show the different types of work involved with patient-centered outcomes research. Underneath is the research platform that supports the work, including research infrastructure, methods development, and training of researchers.

The research process starts with scanning the horizon to identify new and emerging clinical interventions that may impact health care in the U.S. That leads to a systematic review and synthesis of current medical research to compare effectiveness. Evidence synthesis often tells us where the gaps lie between existing medical research and the needs of clinical practice. We also promote and generate new scientific evidence and analytic tools to fill those critical gaps.

All of the information gained needs to be communicated in a way that makes sense to the health care decisionmakers. This includes translating the research into plain language and making it accessible and useful to diverse audiences in order to improve health care. We also have made a commitment to reaching out to stakeholders and communities for input to make sure we get the research right, since if it isn’t relevant and applicable, then we can’t expect it to have an impact.

The Effective Health Care Program has a cradle-to-grave research agenda that focuses on the 14 priority conditions listed here. As you can see, one of the priority conditions is cancer. As you can see here, the research focuses on key populations. The goal is to fill in the gaps the traditional
clinical trials have left out to produce pragmatic, evidence-based information to help inform everyday clinical decisions by you and your patients, including women.

Since the inception of the Effective Health Care Program, AHRQ has funded and completed dozens of patient-centered comparative effectiveness research projects. These projects include comprehensive reviews of diagnostic or treatment options for cancer, diabetes, osteoarthritis, depression, and many other conditions. The Effective Health Care Program creates a variety of products that are based on these research reviews and reports. These include executive summaries, summary guides written for clinicians and consumers and policymakers.

We have recently added to our portfolio a number of materials to support clinician education, including continuing education modules, interactive case studies, and faculty slide sets. We’ll soon be adding patient decision aids as well.

I would like to highlight our consumer guides that summarize the evidence in plain language in easy-to-read formats. These guides are paired with our clinician guides to promote shared decisionmaking. Most of our consumer guides have also been translated into Spanish. The consumer guides can be found online or are available in print. We also have audio podcasts of the guides online as well.

Currently, the Effective Health Care Program offers decisionmaking resources related to the primary prevention of breast cancer and to core-needle biopsy. The findings presented in the clinician and consumer guides can be found on the Effective Health Care Program Web site.

Finally, we want to encourage you to get involved in the Effective Health Care Program. Your participation is mutually beneficial. There are multiple points of involvement in our programs before, during, and after the research is completed.

Before, you can nominate topics for research on our Web site. If there is a breast cancer-related or other topic that you feel should be addressed, we will give you instructions on how to nominate the topic at the end of this conference.

During, you can give input on draft key questions and reports. This kind of involvement helps you get the type of research that will really help answer those controversial questions, and it helps us by getting the research right.

After the research is completed, you can disseminate the information to your colleagues and patients. You can implement the findings in your clinical decisions. This helps both you and us by creating opportunities for better, more informed decisionmaking and by making an impact on the quality of health care.

Now I’d like to turn to the main presentations. First, we have the pleasure of hearing from Dr. Heidi Nelson. Dr. Nelson is a research professor in the departments of medical informatics and clinical epidemiology and medicine at Oregon Health Sciences University and medical director of women’s health at Providence Health and Services in Oregon.
At the Oregon Evidence-based Practice Center, Dr. Nelson has led approximately 25 systematic evidence reviews and analyses for the U.S. Preventive Services Task Force, National Institutes of Health, State of Oregon, and other partners. Today, she will present her findings from the Effective Health Care Program review, *Medications to Reduce the Risk of Primary Breast Cancer in Women*.

Dr. Nelson?

DR. HEIDI NELSON: Thank you.

Prevention strategies for breast cancer currently focus on early detection with screening mammography. However, newer approaches target risk reduction, and these include several things, including identification of BRCA mutation carriers, prophylactic mastectomy or oophorectomy for high-risk women, and the use of medications. Our talk today will be focusing on these medications.

Despite the availability of medications that reduce risk of primary breast cancer, they’re not commonly used in the United States. This may be because it’s unclear how to apply results of recently published trials in clinical practice.

I’m sorry; we are out of order here. Let me just get back to our correct slide.

This comparative effectiveness review was commissioned by AHRQ to help inform new U.S. Preventive Services recommendations. The previous recommendations were released in 2002 and have become very much outdated. At that time, the task force gave a B recommendation for high-risk women. A B recommendation means that clinicians should discuss chemoprevention with women at high risk for breast cancer and at low risk for adverse effects of chemoprevention.

They gave a D recommendation for women considered at low risk. A D recommendation means that they recommend against routine use of tamoxifen and raloxifene for the primary prevention of breast cancer in women at low or average risk for breast cancer. As you can see, the decision to use these medications really hinges upon identifying women as high or average risk for breast cancer.

This comparative effectiveness review is a summary of the available evidence around the effectiveness and harms of these drugs. It involves several steps that are laid out in a procedure that all of the comparative effectiveness reviews in this program follow and are detailed in the full report. But, just briefly, to walk through those, there is a step of development and prioritization of the topic; development of the key questions that actually guide the review; and then collection of abstracts and papers by searching for published and unpublished studies in registries and soliciting information from drug companies.

The published review that we conducted went through data collection through January of 2009. We selected studies based on predetermined eligibility criteria, and this included randomized, controlled trials for studies of effectiveness; randomized, controlled trials as well as observational studies and registry data regarding harms; and discriminatory and diagnostic accuracy studies looking at risk models.
We extracted data from the studies that met the inclusion criteria; evaluated studies for quality and applicability using predefined criteria; and statistically combined results of trials in meta-analyses for the major health outcomes. We evaluated the strength of evidence for each comparator and outcome using the GRADE criteria, as further described in our report; interpreted results in the context of the strengths and limitations of the evidence; identified future research needs; and are now completing some of our steps in dissemination.

We published a paper a year ago, November 2009, that reported the main outcomes in the “Annals of Internal Medicine.” There is a full summary available on the AHRQ Web site that describes our methods and all of the results. We have two products, one you saw earlier, a guide for women, and a guide for clinicians. And then the U.S. Task Force will be releasing recommendations, most likely in 2011. So they’re not available yet, but they will be forthcoming.

The key questions, of which we’ll cover four of them in this talk, include primarily key question one. In adult women without pre-existing breast cancer, what is the comparative effectiveness of selective estrogen receptor modulators tamoxifen citrate and raloxifene and the selective tissue estrogenic activity regulator tibolone, when used to reduce risks for primary breast cancer in improving short-term and long-term outcomes, including invasive breast cancer; noninvasive breast cancer, including ductal carcinoma in situ; breast cancer mortality; all-cause mortality; and osteoporotic fractures?

As with all comparative effectiveness evaluations that we do looking at drugs, we have a parallel question relating to the evidence for harms. This is particularly important in prevention because we’re starting with a patient who has no—doesn’t have the health concern existing, is asymptomatic, and so the threshold for causing harm is different than if you’re treating someone who actually has breast cancer.

So, key question two actually takes a lot of effort to find evidence for harms in looking beyond the randomized trials, and we had a long list of potential harms and usually keep these lists open-ended so that if, in the process of our review, we discover things we didn’t know about, we would include them as well. So, these are the main harms that we investigated in the review.

Key question three, looking at how the outcomes, both benefits and harms, vary by heterogeneity in subpopulations. And here are a few of the subpopulations we looked at, whether outcomes varied by age, menopausal status, use of exogenous estrogen, their risk of breast cancer as determined by a number of ways, and ethnicity and race. We also kept this list open-ended as well in that any other subgroups that we could find, we would report on.

And key question four, what methods, such as clinical risk assessment models, have been used to identify women who could benefit from medications to reduce risk of primary breast cancer?

That’s a very important question clinically; as you saw, the recommendations really do depend on whether we can correctly identify women as high or average risk.

So we looked at three drugs in the comparative effectiveness review. In 2010, a new drug came out that is not yet FDA approved, but for the sake of completeness, I have included it in the tables here.
so that you are informed of that as well. But the actual evidence review includes tamoxifen, raloxifene, and tibolone.

I would like to highlight that neither tibolone nor lasofoxifene is FDA approved. They are in the process of being approved, so they may be something you'll deal with in the future. The patient and clinical guides and the U.S. Task Force recommendations are going to be focusing only on tamoxifen and raloxifene because they are FDA approved.

As you see, three of these are SERMs; one is a STEAR, a slightly different mechanism. Tamoxifen was approved to treat breast cancer, was found to reduce cancer in contralateral affected breasts, and had the first trials relating to prevention involving tamoxifen.

The raloxifene trials were designed primarily for osteoporosis outcomes, and, in the course of that evaluation, it was found that those women using raloxifene also had reduced breast cancer. So, in 2007, raloxifene was approved for that use as well.

Tibolone is used in many countries outside of the United States, mostly for treatment of menopausal symptoms and also postmenopausal osteoporosis. However, many American women actually use this medication by obtaining it from outside the U.S., and, indeed, the main tibolone trial that we'll talk about here was primarily conducted in the United States. So it’s something to be aware of, but it’s not currently part of treatment in the United States.

There are nine primary prevention trials if we include the one on lasofoxifene that was just published this year, and these are well-designed, high-powered studies with lots of participants in the various arms. They are very well done. It’s kind of a reviewer’s dream to have such a nice set of data to work with.

We have four trials of tamoxifen compared to the placebo, three of those done outside the United States. We have two trials of raloxifene compared to placebo, both done primarily in United States, but were international as well. One trial of tibolone compared to placebo, and the one lasofoxifene versus placebo. We are lucky to have one well-done head-to-head trial of tamoxifen compared to raloxifene that was primarily done in the United States, the STAR trial.

The one real take-home point from looking at this array of studies is to enforce the idea that we have different women enrolled in these trials, so when we look at—when we compare the placebo-controlled trials against each other, we have to be very cautious on our conclusions because the women in the tamoxifen trials are younger and were enrolled based on their breast cancer risk. The women in the raloxifene, tibolone, and lasofoxifene trials are older and were primarily enrolled based on their osteoporosis or cardiovascular risk. So, they’re quite different in that regard.

Thank goodness for the STAR trial, where the women who were enrolled are sort of intermediate in age—between the tamoxifen trials and the raloxifene other trials—and also had an increased risk for breast cancer, although their threshold was quite low. They used a cut point of the Gail model score of 1.67 percent 5-year risk for breast cancer, so actually kind of a low threshold to get into the trial, but it was at least based on that risk. They don’t have the same populations of women, so we had to be careful when we looked at the outcomes.
As I mentioned in the method section, we did a series of a meta-analyses. We actually have about 25 of these forest plots in that report. I’m going to show you only one because we have much less time to dig into that today.

But this is a forest plot of the main outcome, invasive breast cancer, showing the results of the four medications in the trials that we evaluated. For tamoxifen, we have the four placebo-controlled trials indicated in the white squares. The horizontal lines through those squares are the confidence intervals.

For those of you not familiar with forest plots, you’ll recall that the line down the middle—kind of separating the favors drug/favors placebo sides of this slide—indicates no difference, and anything on the side that says favors drug means that that is a finding where the drug has a benefit or has a significant finding compared to placebo. So, in this case, reducing cases of invasive breast cancer is clearly a benefit. And, as you see, every point on this slide is in the direction of favoring the drug.

The confidence intervals, though, as indicated by those small horizontal lines, don’t all fit on this side of the line; some cross. But when we do a meta-analysis for the tamoxifen trials, which is represented in the white triangle, we see that that result is statistically significant, showing a 30 percent reduction in invasive breast cancer for women using tamoxifen.

If we look at the raloxifene trials, we look at the yellow boxes for the two trials. Again, older women, designed for osteoporosis and cardiovascular outcomes. We see, however, a significant reduction in invasive breast cancer in both of those trials, and the combined estimate is shown in the triangle.

We look at the LIFT trial for tibolone, the green square. One trial, no meta-analysis done, but that shows a similar result. And the same thing with the PEARL trial for lasofoxifene.

So, there’s consistency in all the drugs, all leaning on this side. It’s attractive to try to take those point estimates and say one’s better than the other, but the confidence intervals show us lots of overlap. So we all know they work. But, at this point, it’s hard to say if one is superior to another.

In order to make this a little more real in regard to patient care, we calculated the events prevented per thousand women per year, assuming a 5-year treatment course. And you see the major clinical outcomes for benefits listed on the left side of this slide, and then the four drugs across the top.

Our strength of evidence is stronger for tamoxifen and raloxifene because we have more than one trial, but the results for tibolone and lasofoxifene are very consistent with those. And you’ll see for invasive breast cancer, really, they’re about the same kind of magnitude, 7, 8, 9, or 10 per thousand across the 5-year treatment. So that looks fairly consistent.

If we are able to break that down into estrogen receptor positive breast cancer, some of the studies were able to do that—the tibolone trial did not do that—we see that most of that benefit happens in estrogen receptor positive invasive breast cancer. We don’t see a significant difference in estrogen
receptor negative between placebo and treatment. Similarly, no difference in noninvasive cancer. No difference in all-cause deaths. There are benefits shown for vertebral and non-vertebral fractures for the different drugs.

Again, some caution with those point estimates because, as I said earlier, tibolone and lasofoxifene were selectively enrolling women with low bone density, so you would expect more benefit there. The tamoxifen trial specifically had younger women without a lot of osteoporosis going on, so they did show a significant benefit, but a much smaller magnitude of effect. So thank goodness for our head-to-head trial because that really does help sort out the tamoxifen/raloxifene question.

And, in our evidence report, we reported the results of the STAR trial, and that came out in 2006. Well, we were quickly outdated because new results have come out in 2010. And, in 2006, the drugs showed the same outcomes for invasive breast cancer.

The results in 2010 show that there are more cases with raloxifene. So, tamoxifen has an edge over raloxifene in reducing invasive breast cancer in a head-to-head trial. And whether that is because of a persistent effect after stopping the drug or whether that is because of superiority in the drug is hard to tell, since the treatment duration was 5 years then followed by a couple more years of followup.

The results in 2010 showed no difference in noninvasive cancer and no difference between the drugs in all-cause death. The 2010 results did not report fracture outcomes specifically, but they did state that there were no differences that were consistent with what we saw in 2006, so they were similar in their effects on fractures.

Now, there are–our look at subgroups indicated that it seemed to work in all groups for tamoxifen and raloxifene. Now, tibolone and lasofoxifene didn’t show us outcomes by the subgroups of interest, so I can talk only about the two here.

Tamoxifen reduced breast cancer in groups that were broken down by different ages, menopausal status, whether they’d had prior estrogen use, whether they had a family history of breast cancer, whether they had other abnormal breast tissue findings, it worked for everybody. But, in the NIH trial, the cancer rates were highest and risk reductions were greatest in women who had prior atypical hyperplasia and the highest modified Gail model risk category. So, there may be some improved magnitude of effect in women who are at the highest risk of the spectrum.

For raloxifene, again, breast cancer was reduced in all the groups that they looked at, and these are mostly selected by the investigators, not us. We just are kind of reporting what they reported, that all groups showed benefit when they actually looked at them all.

So, on the flip side, there’s never any clear free lunch with these drugs; of course, there are going to be some harms inherent in taking these, and the biggest one, the most consistent one across the trials, is the risk of thromboembolic events. And here we use the combined DVT, PE, any kind of thromboembolic event. And, again, showing you the per thousand rates over a 5-year treatment duration, it ranged from four per thousand, seven per thousand, eight per thousand and for tamoxifen, raloxifene, and lasofoxifene. There were no significant differences for tibolone.
For coronary heart disease, there were no differences between placebo and drug until you got to lasofoxifene. We actually saw a protective effect, which will be really interesting as we get more data on this drug, a protective effect for coronary heart disease and stroke.

We get to stroke, though, with tibolone increasing the risk of strokes, 11 per thousand cases. And, indeed, this is what stopped the trial early. When you drill down into that data, it was clear that those strokes occurred in women who were over age 70, so it was very age-specific. And, in that trial, most of the women were over age 60, so we still don’t have a good trial of the effect on younger women for these outcomes.

For tamoxifen, we have additional adverse outcomes: endometrial cancer and cataracts. These were also highlighted in the STAR head-to-head trial when we looked at raloxifene versus tamoxifen. Again, 2010 data, more thromboembolic events with tamoxifen, more endometrial cancer, and cataracts compared to raloxifene.

Now, there are also side effects that are not life threatening but led to more discontinuation of the drug. And leading the pack really were hot flashes that were produced through use of tamoxifen, raloxifene, and lasofoxifene. Tibolone’s advantage is that it reduces hot flashes. But these are the other—listed are other common side effects that were more so in the treatment than placebo groups that sometimes led to discontinuation.

Now, to summarize briefly the risk models on selecting candidates for therapy, you see that there’s a discussion that’s needed. There are definite benefits; there are definite harms. Who would be selected for—to consider for this medication took us to our final key question where we looked at all the risk models and how they performed in identifying individuals at high or average risk.

And many of the trial entry criteria, and the Task Force recommendations, and many of the models were based on the Gail model or the Breast Cancer Risk Assessment Tool, which is on the NIH Web site and certainly something that most clinicians are familiar with. The variables included in the Gail model are listed on this slide.

When we looked at a number of studies, 16 studies of 9 models, many of which were variations of this Gail model, we looked at a way of sort of summarizing that by looking at the range of area under the receiver operating characteristic [ROC] curve showing discriminatory accuracy. So, essentially, it tells us how well it can sort out an individual’s risk.

And, if you remember your ROC curve, see that it had it straight, or they—the diagonal line with the—that I’ve mentioned indicated it’s worthless, is sort of the result you get if you had a coin of—a toss of a coin, where it would be sort of a 50/50 chance, so pretty worthless as a clinical indicator. If you looked at the solid line showing a perfect predictor, that would be an excellent tool in clinical practice. And something maybe not perfect, but high up toward that curve, would be maybe a worthwhile tool to use in clinical practice.

Well, unfortunately, when we looked at these nine models, we found AUC curves that were close to the worthless line. So that leaves clinicians in sort of a quandary as to well what do—how do I
really effectively select these patients? So, that’s hopefully going to generate the next level of research in this area.

So, I’ll stop here. Thank you very much.

ROGAL: Thank you, Dr. Nelson.

As a reminder for our participants, we invite you to submit questions electronically by entering them via the “Ask Question” button at the bottom of your screen. Once you enter your question, press the “Submit” button.

We will now hear from Dr. Karen Schoelles. Dr. Schoelles is the director of the Evidence-based Practice Center (EPC) at the ECRI Institute. She is responsible for ensuring the clinical relevance of ECRI Institute’s EPC projects and Health Technology Assessment Group’s work products. She is also the project director for the AHRQ health care Horizon Scanning System. Dr. Schoelles has over 20 years of clinical experience in internal medicine, with particular expertise in geriatrics, and has served as the principal investigator on numerous systematic review projects.

Full biosketches can be found in the supporting materials box for both of our presenters.

Today, Dr. Schoelles will present her findings on the Effective Health Care Program review “Core-Needle Biopsy for Breast Abnormalities.”

Dr. Schoelles?

DR. KAREN SCHOELLES: Thank you.

I’ll ask you to switch gears now to a different population and a different type of systematic review looking at comparative effectiveness.

Slide 34. The objectives of my talk today are to describe the process of creating this comparative effectiveness review, not getting too heavily into the methodology of it, but just to give you a sense, to describe our findings, and to talk a little bit about the implications of this review.

Slide 35. I present here the analytic framework that we used to map out how we were going to approach the topic. And I’m hoping that your screen is larger than mine is because it’s hard to read on the Web presentation. But, at any rate, the—I’d like to emphasize what we refer to in systematic review as the PICO, P-I-C-O; P standing for patient population; I, intervention; C, comparator; and O for outcomes. We typically design one of these visual diagrams to help us think out what types of evidence we’re going to need and what types of questions we think are most important to answer.

Slide 36. In this instance, the population that we’re looking at are the asymptomatic women; women at average risk, not women with genetically defined risk factors, or, if we had a good model for doing what Dr. Nelson was just describing, not the women who would come out at the high-risk end of the scale. So, asymptomatic women who, on routine physical or routine breast self-exam,
have discovered an abnormality, or women who have undergone routine screening mammography and who have an abnormality identified on the mammogram.

Slide 37. I put up for you the American College of Radiology rating system for mammographic abnormalities. Looks like the slide is not advancing, so let me just do that.

The American College of Radiology rating system is intended to help guide clinicians in interpretation of mammographic findings. So, the women in the studies that we’ve looked at primarily had mammograms that were considered suspicious and for which biopsies should be considered. We believe that, in some of the studies, there would be women whose mammograms suggested something more benign, but that perhaps rather than short-interval followup, folks were more comfortable with proceeding to biopsy.

Slide 38. The intervention core-needle biopsy is a variety of different types of equipment, but generally the gauge of the needles used is anywhere from the largest being 11 to the smallest being 16. There are some needles that are as large as 10 gauge that have been used, but none in the studies that we’ve included.

The lesion is generally located either by palpation or by imaging. So, if an abnormality is identified only at mammography or ultrasound, that modality would be used to guide the person performing the procedure during the actual performance of the procedure. So, stereotactic mammography helps to identify the location of the abnormality in more of a 3D way for the—typically the radiologist.

Core needles can be inserted multiple times to get different samples, or if, in the case of core needles that are—that have an adaptation for vacuum assistance, the needles typically rotate and multiple samples are taken as the vacuum is applied.

Slide 39. Open surgical biopsy is often excisional, in other words, intended to completely remove an abnormality, or incisional, in which it’s intended to sample the abnormality. The problem comes when lesions are not palpable. That means that imaging has to be used to create some type of marking that can be utilized by the surgeon. You may have heard of women undergoing stereotactic mammography, in which a wire is placed that the tip of which is in the lesion seen on mammogram, and then the woman is taken to the operating room, and the surgeon follows that wire down to the lesion to remove it.

Slide 40. In this case, many of the women who had core-needle biopsy initially and had a benign finding, as you might expect, would not want to undergo open surgical biopsy to confirm the findings of a core-needle biopsy. So, we allowed studies to follow women clinically and with imaging as long as they did so for a minimum of 6 months.

Slide 41. As with most systematic reviews, part of the initial upfront work is doing extensive searches. We have a team of medical librarians who use a variety of terminology, some of which is standardized, to search for information. Our searches were last updated in September of 2009. Of the articles identified, over 1,200 of them were screened by reading the abstracts to see whether
they were relevant to this particular project. We then had 589 articles that were obtained—we obtained the full article and screened those against our inclusion criteria.

And our inclusion criteria for this study—and I’ve already described to some extent the women either had to undergo open surgical biopsy to confirm the core-needle findings or be followed for at least 6 months—the abnormality had to have been discovered at screening rather than in response to a symptom or a concern. We required that the studies have information on at least 50 percent of the women who entered the study, that the instruments used be currently available, and that they had studied a minimum of 10 patients. Most of the studies that were excluded were excluded because they were case studies that had an increased risk for selection bias, were not relevant, or did not verify the core-needle diagnoses that were benign.

Slide 42. We then proceed to look at each individual study and try to determine the risk of bias. We’ve often used the term quality of the study, but most of us in the EPC program are favoring the term risk of bias, which, again, is not something necessarily intentional. It’s a matter of the design and conduct of a study and how that might influence how much faith you can have in the results in terms of how they answer our particular questions. Not necessarily the questions the original authors were addressing, but the questions for this systematic review.

The major problem in the studies that we evaluated was the lack of good reporting of what happened to the women and the fact that so many of the core-needle biopsies and open surgical biopsies, the studies did not report sufficient followup to make us comfortable. We allowed them into this study set if they had at least 6 months of followup, but obviously we’d like more than that. Many of the studies did not provide information on as many patients in the study as we would like to have seen.

Slide 43. In the assessment of diagnostic tests, we’d much prefer to be able to look at patient outcomes to know what happened after the diagnostic test was applied, what treatments were chosen, what decisions were made, and how did the patient fare over the long term. For most diagnostic tests, we don’t have that type of information. We’re usually having to look at test characteristics and comparing that to other things that we’re familiar with.

I’ve put up here a typical two-by-two table that those of you who’ve taken epidemiology would be very familiar with that shows you how the new test compares against the presence or absence of disease as determined by some other test or some other way of knowing. In this case, core-needle biopsy is the new test, and the reference standard is either open surgical biopsy or clinical followup. The terms that most of you are familiar with would be sensitivity and specificity, predictive values. Likelihood ratios may be less familiar, but, in terms of comparing tests, the likelihood ratios are particularly helpful to those of us with a clinical perspective.

Slide 44. I’ve put up the formula for the likelihood ratio, and it’s a bit of a mouthful and a bit—it takes a little bit of doing to calculate. So, I’m not going to read that out but just wanted you to have it for reference.

For this study, we thought that the most important thing to consider was that we don’t want to be missing cancers. So, consequently, the diagnostic test characteristics that we focused on are the
sensitivity, the negative predictive value, and the negative likelihood ratio. Core-needle biopsy is considered to be 100 percent specific, and what that means is if a malignancy is found on core-needle biopsy and the subsequent surgery does not identify the malignancy, the assumption is made that the malignancy was actually removed completely by the core-needle biopsy. And that seems to have been borne out with more recent studies that that, indeed, is the case.

Slide 45. This is something called Fagan’s nomogram, and I’m hoping that you can see the diagonal lines superimposed on the figure. This way of thinking about diagnostic tests was proposed by Dr. Fagan in 1975. I’ve put up a link to a Web site where you can play with this tool in thinking about different diagnostic tests.

What I have on the screen, superimposed on the nomogram, is what would be typical for a woman who has had a mammogram done and—had a screening mammogram done that has shown a BI-RADS for abnormalities. That woman typically would have about a 30 percent chance of actually eventually being found to have breast cancer. So, on the left-hand side of that diagram is the 30 percent. That’s what we refer to as the pretest probability of having cancer in this case.

The likelihood ratio demonstrated here is for one of the core-needle biopsy types, and, in this instance, the positive likelihood ratio was 54. That number, in and of itself, won’t mean much to you. But if you see the solid red line drawn between the 30 percent pretest probability through that central line of likelihood ratio, so crossing through the 54, on the far right we have the posttest probability. That’s the likelihood that a woman with a positive test for cancer using this method, if she initially had a 30 percent risk of cancer, a positive test changes your impression to about a 96 percent chance of breast cancer, so there you’re relatively certain that cancer is present.

The diagonal line going in the downward direction that—a dashed line crosses through the likelihood ratio point at 0.04. That’s the negative likelihood ratio. In other words, if a woman with a pretest probability of 30 percent has a negative breast biopsy using this technique, you cross that line at 0.04, and you now can conclude that her posttest likelihood of having breast cancer is down around 2 percent.

So, this test is quite useful. A positive test changes your assessment quite a bit, and a negative test changes your assessment quite a bit. One of my colleagues referred to this as the alligator method of looking at diagnostic tests, in that the wider the alligator’s jaws are open, the better this test is in helping out the clinician.

OK, slide 46. I’ve presented here our meta-analyses for the sensitivity of the different forms of biopsy. Surgical—open surgical biopsy is assumed to be fairly close to 100 percent. Recent studies have not been done to confirm that, but we think that the best estimate is that it’s highly sensitive, close to 100 percent.

The various types of core-needle biopsy included the stereotactic, vacuum-assisted biopsy, a stereotactic procedure without the vacuum, ultrasound, that’s the U/S, either with or without the vacuum, and a freehand technique. So, the stereotactic methods, in this instance, were fairly close to the sensitivity of the open surgical biopsy.
The negative likelihood ratios I’ve demonstrated here, again, show that for the stereotactic techniques, the results were fairly close to those for the open surgical biopsy. And, in this case, again, the smaller this number, the better the test is in ruling out breast cancer.

Slide 48 and 49 I’m going to skip over and go down to slide 50. We estimate the overall strength of evidence in our comparative effectiveness reviews, looking at each question we’re asking along with each outcome. We compile several things in making the judgment about strength of evidence.

First is that risk of bias in the individual studies we’re using to answer the question. The amount of evidence we have; how many studies and how many patients. In this case, overall, we had 107 studies and over 57,000 breast biopsies included. The consistency of the findings across and within studies, and the robustness of our calculated results. In other words, if we do some sensitivity analyses where we alter some of the decisions we make or some of the assumptions we have to make, how well does the result stand up?

In the end, we come up with a grade for the strength of evidence. In the EPC Program, we use the language high, moderate, low, or insufficient. These do not consist of recommendations; this is different from the work that Dr. Nelson described previously in that we are not making any recommendations with this report. We are simply describing the evidence and asking clinicians and professional societies and policymakers to consider, based on a host of factors, that evidence in their decisionmaking.

Slide 51 is an attempt to translate some of this in a way that’s maybe a little easier to think about. I’m just going to focus on the second column on this slide. The number of missed cancers expected for every thousand biopsies performed.

For open surgical biopsy, we think that would be in the range of three to six missed cancers. You can see for the freehand technique it would be greater. So, in other words, that technique is not using any imaging. The ultrasound-guided procedure, six to nine missed cancers. The stereotactic non-vacuum procedure, three to 13 missed cancers.

On slide 52 we have similar findings here. Again, I’ve repeated the results for the open surgical biopsy, three to six missed cancers. For the ultrasound-guided vacuum-assisted procedures, there was a much broader range, and that reflects that some—potentially could point to some differences in population or it could simply be differences in technique. The stereotactically guided, vacuum-assisted procedure was, we thought, quite close to open surgical biopsy and might even, in some instances, be slightly more accurate. The range is somewhere between one and six missed cancers out of every thousand biopsies.
The strength of evidence for all these findings, as you can see from the slides, was generally either insufficient or low and that was because we were—we remained concerned about risk of bias in the studies, that they were quite consistent, and there were quite a lot of biopsies. So, we were confident enough to rate those at least low and not insufficient.

Slide 53 just demonstrates some sensitivity analyses that we did after the fact, in which we said, all right, supposing because the studies were not as good as what we’d like to see in terms of their risk of bias, suppose we overestimated the sensitivity of these procedures. How much would that be likely to affect the woman’s posttest probability of having cancer when the test is negative?

And that’s what these—the numbers in those columns represent. The first—the second column of analysis results is what we actually found in the studies. The next column is if we had overestimated the sensitivity by 1 percent, these would be the posttest probability of still having cancer, even though the test has been negative.

So, just to wind up, the other things that we looked at were the harms of the procedures, and they were generally quite minimal, although it was pretty clear that the core-needle biopsy techniques had fewer harms than did open surgical biopsy. Women who underwent core-needle biopsy ended up having fewer surgeries overall than the women who had open surgical biopsy.

So, how do you use this? Well, you have to take each woman as an individual and apply the evidence. You have to consider her values, you know, for a woman knowing that she has a 30 percent chance of having breast cancer; different women are going to consider the findings of this in different ways. And I think, for some women, the fact that it’s less—potentially less harmful and nearly as accurate as open surgical biopsy—is going to persuade them to undergo core-needle biopsy. In other instances, some women are going to feel more comfortable having an open surgical biopsy.

So, again, this is descriptive, not prescriptive. I refer you to the clinician guide that’s on the AHRQ Web site, the full report, and the manuscript that was published in “Annals of Internal Medicine” earlier this year for more details.

And we’ll be happy to take your questions. Thank you.

ROGAL: Thank you, Dr. Schoelles.

We appreciate both you and Dr. Nelson for making the information and research salient to the clinician audience.

Now, we would like to start the Q&A session. If you have not already submitted a question to either of our speakers, please type your question into the “Ask Question” box at the bottom of your screen.

Dr. Nelson, the first question is for you. It says you mentioned that you examined the discriminatory accuracy curves of nine risk assessment models. Which models did you examine, and how did you choose those models?
NELSON: OK. Great question. We all want to have something we can work with in a clinical setting.

We used inclusion criteria for those risk stratification models and chose ones that could be used in a primary care setting. There are a lot of other more sophisticated models that are used in genetic counseling sessions, but we did not include those. We wanted to have them be ones that a clinician sitting down with a patient to make this decision could start with. They had to be studies that were designed to identify women at higher-than-average risk for breast cancer, and they had to report outcomes in a way that we could analyze them. So we’re looking at discriminatory accuracy studies that specifically reported SC statistic.

The nine types of models that we included were many variations on the Gail model, including ones adjusted for different races, ones that included additional risk factors. So, there were several of those types. There was also a model derived from the Breast Cancer Surveillance Consortium that involved ages, races, Gail scores. One that we called the Chlebowski model, modeled after the Women’s Health Initiative. The Colditz/Rosner model we’re calling mostly after the authors. And the Tyrer/Cuzick model. So, our report describes them all in a lot of detail, and we do have a paper coming later this year—or next year—that will lay that out a little bit for you too.

ROGAL: Thank you.

Dr. Schoelles, should all women have a stereotactic, vacuum-assisted, core-needle biopsy as their first diagnostic test?

SCHOELLES: That’s going to depend. Some lesions are not visible on mammography, and, in that case, ultrasound guidance is going to be more helpful. So, it’s also going to depend on what’s available in the region where the woman lives and whether there are people available who are skilled in performing the various types of biopsy.

ROGAL: Thank you.

We actually have a question that we’re not sure which one of you would be most appropriate to answer, so I’m going to ask it, and then maybe you’d both like to address it or you can choose. Should chemoprophylaxis for reduction of breast cancer risk be offered to premenopausal women with ductal hyperplasia without atypia found on core-needle biopsy?

NELSON: Sure. Unfortunately, none of the trials clearly lays out who should be getting the medication. It’s sort of left for physicians and the patients to discuss.

This is somebody who might—it would be useful to bring up the conversation and do some shared decisionmaking, maybe do some family risk factor analysis as well and get sort of a comprehensive risk evaluation. Somebody who has a finding, but not the atypia that we saw in the nice clean boxes that the NIH trial showed us results for. So, it’s going to have to hinge more on kind of how
you put these risk factors and risk assessment together with that finding and the patient, again assessing benefits and harms.

The only drug that would work for this patient would be tamoxifen, since raloxifene and the others are for postmenopausal women only. So, the choice between tamoxifen and raloxifene for this patient, premenopausal, would be tamoxifen. So, it would be worth starting a conversation, definitely. This person may end up being a good candidate.

ROGAL: OK. And one more question for you, Dr. Nelson. Did you look at the effects of these different medications on women with BRCA mutations?

NELSON: We looked for data on it for that subgroup, and many of the tamoxifen trials started before there was a lot of testing for BRCA mutations. So, there have been a couple of publications that describe them, and it’s sort of retrospective testing of women in the NIH trial. But they aren’t—they don’t have enough women to really power significant findings.

But anyone with the BRCA mutation would be in those highest risk groups and would certainly be someone who should consider chemoprevention. But we don’t have good numbers to pin that onto. So, kind of like the prior question, we know that there’s—they’re out of the average-risk group and may well be a good candidate, but, again, it’s a one-on-one discussion.

ROGAL: Thank you.

Our time is almost up. If we did not get to your question today, please e-mail us at ehc_clinicians@ahrq.hhs.gov. All the resources mentioned today can be accessed and printed from the Effective Health Care Program Web site listed here, or you can order them in larger quantities, for free, through the AHRQ publications clearinghouse. The various reviews and reports on breast cancer can be easily accessed as well as the related educational materials.

Also on the Web site, you can become involved in the topic nomination and refinement process that I described as well as comment on draft reviews and reports. All of these features, in addition to signing up for e-mail updates, can be easily navigated in the panel on the left-hand side of the Effective Health Care Program Web site.

As I noted earlier, we are holding three more similar events within the next 2 weeks. They are listed here. Please visit the Effective Health Care Program Web site to learn more.

I would like to thank Dr. Nelson and Dr. Schoelles for sharing their research findings with us today. It was very helpful.

I would also like to thank our many participants for joining us today. We hope that the information presented here informed you about how you can implement patient-centered outcomes research into everyday practice and how the Effective Health Care Program resources are available to assist you and your patients with decisionmaking. As we conclude this Web conference, let me remind you that this event will be archived and available shortly on the Effective Health Care Program Web site.
Finally, as you leave the event, please answer the one feedback question posed. Your feedback is very important to us as we develop more resources and plan similar events. Have a nice day.

OPERATOR: Thank you. Once again, this does conclude today’s Web conference. We thank you for your participation. You may disconnect at this time, and have a great day.

END