

National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium

Radiology 2017; 283:49–58

Constance D. Lehman, MD, PhD • Robert F. Arao, MPH • Brian L. Sprague, PhD • Janie M. Lee, MD, MSc
Diana S. M. Buist, PhD, MPH • Karla Kerlikowske, MD • Louise M. Henderson, PhD, MSPH
Tracy Onega, PhD, MS, MA • Anna N. A. Tosteson, ScD • Garth H. Rauscher, PhD • Diana L. Miglioretti, PhD

National Performance Benchmarks for Modern Diagnostic Digital Mammography: Update from the Breast Cancer Surveillance Consortium

Radiology 2017; 283:59–69

Brian L. Sprague, PhD • Robert F. Arao, MPH • Diana L. Miglioretti, PhD • Louise M. Henderson, PhD
Diana S. M. Buist, PhD, MPH • Tracy Onega, PhD • Garth H. Rauscher, PhD, MPH • Janie M. Lee, MD, MS
Anna N. A. Tosteson, ScD • Karla Kerlikowske, MD, MS • Constance D. Lehman, MD, PhD
For the Breast Cancer Surveillance Consortium

2017 Breast Cancer Surveillance Consortium Reports on Interpretive Performance at Screening and Diagnostic Mammography: Welcome New Data, But Not as Benchmarks for Practice

Radiology 2017; 283:7–9

Carl J. D'Orsi, MD • Edward A. Sickles, MD

Deborah Levine, MD Hello. I'm Debbie Levine. I'm the Senior Deputy Editor for *Radiology* and I'm here today having a round table podcast with experts in breast imaging screening. The Breast Cancer Surveillance Consortium has two articles in our April *Radiology* issue. One of these is on national performance benchmarks for modern screening digital mammography, and Dr. Connie Lehman Professor of Radiology at MGH will be discussing that. A second is an update on diagnostic mammography with Dr. Diana Miglioretti who's Professor of Biostatistics at the University of California Davis. And then because benchmarks are such an important topic we have accompanying editorial in this issue of radiology that's co-written by Dr. Ed Sickles and Dr. Carl D'Orsi. They're both Professor Emeritus of Radiology; Dr. Sickles at UCSF and Dr. D'Orsi at Emory. I'd like to thank you all for participating in this round table discussion with me today. Let's start with the update on screening mammography. Dr. Lehman can you please tell us a little bit about what your group did and what you found?

Constance D. Lehman, MD, PhD Sure so the Breast Cancer Surveillance Consortium members we really wanted to address performance of modern digital screening mammography. So we had a database of over 1.7 million all digital screening mammograms from 360 radiologists at almost 100 facilities. And we measured the performance and compared it back in time to film screen and earlier performance measures. And we were excited to see the results. What we found was really great news. More cancers are being detected with modern mammography programs than historically. The vast majority of radiologists in U.S. community practice are doing a fantastic job in finding cancers. The sensitivity

of mammography is higher than previously reported. So this was really exciting to see. At the same time, we found that radiologists were struggling more with keeping false positive rates low. This increased cancer detection and increased sensitivity seemed to be at the cause of more false positives. And we also saw there is a lot of variation. We have a lot of radiologists that keep low recall rates and high cancer detection rates, we have others that struggle more with that. So we thought this could really help inform the future directions that we can take in screening mammography in the U.S.

D.L. Terrific, and Dr. Miglioretti can you tell us a little bit about the study on diagnostic mammography and what those main results were?

Diana L. Miglioretti, PhD Sure. We looked at over 400,000 diagnostic mammograms performed by over 400 radiologists at almost 100 facilities. And similar to the results from the screening benchmarks, we found that compared to our prior benchmarks that included film mammography, that we are now detecting more cancers or the cancer detection rate has increased but that has come at a cost of increasing the abnormal interpretation rate. So it seems that we're now doing more biopsies that may not be necessary in order to detect those extra cancers. We've also found wide variability across radiologists in these performance metrics. So some radiologists are finding a lot more cancers and others may be not meeting current benchmarks.

D.L. So Dr. Sickles in your editorial you had some rather pointed comments regarding using the breast cancer surveillance consortium data as bench marks,

basically saying that many people in other groups are going to have difficulty linking to cancer registries and getting data on sensitivity and specificity of their own screening populations and I was just wondering in this digital age shouldn't we work to solve that problem of getting the data rather than saying that these metrics can't work?

Edward A. Sickles, MD The metrics work for each facility. What's important for people to realize is that the real strength of the BCSC data and why the two articles you're publishing are so important is that because the BCSC is able to link to tumor registries, they get near 100% cancer ascertainment. Cancer ascertainment is really important with outcomes because it helps reliably determine whether a positive is a true positive or a false positive, whether a negative is a true negative or a false negative. So that's important and that's why the BCSC data show what's really going on in the United States. The problem is that benchmarks for individual radiologists and individual mammography facilities have to reflect the data collection that these facilities do; and most facilities the vast majority of facilities in the United States don't link with regional tumor registries, they're not large HMOs that have captive populations so they have sub-optimal, often not very good cancer ascertainment so the data they get from their own audits won't match the completeness of the data in the BCSC and that creates a problem because then those practices are hoping to match BCSC benchmarks when in fact they may not be able to.

D.L. So Dr. D'Orsi moving on in your editorial and just talking about breast cancer screening in general, how do you recommend that we combine the goals for reasonably low recall rate while preserving a high cancer detection rate and how can we use these ideas for screening and diagnostic thresholds that are potentially different from our current recommendations?

Carl J. D'Orsi, MD I think one of the things you have to realize is remember the price you pay to detect cancer basically is going to be the false positive. Obviously if you cut the false positive rate down too low you're going to start missing malignancies. That doesn't mean you should have a false positive rate of 80%, but just keep that in mind. So the biggest thing I think to have radiologists perform better in the false positive arena is education and really what we do at many conferences is a face-to-face with trainees with a set of mammograms. The other thing that's important is you have to find out or you have to know what type of cancer you're finding. You many have a very high or very low false positive rate and a good true positive rate but you may be finding stage II cancers. So the point is that these benchmarks do not work in isolation, they work as a cluster and you have to look at the product you're getting from these metrics to ensure that you're doing what a screening exam should do. So the bottom line is let's be a little bit wary about knocking the head of false positives.

D.L. Great. Thank you. So Dr. Miglioretti in their editorial Dr. Sickles and Dr. D'Orsi suggest that we might want to move to use of a larger database such as the National Mammography Database since it has more accrual than the BCSC does and has cancer ascertainment that could potentially reflect current practice of mammography better than the BCSC. What do you think about that idea and is it possible or helpful for these two databases to somehow combine forces?

D.L.M. Yeah I think the BCSC and the NMD both have unique strengths and they're complementary in many ways. The BCSC has very high quality data collection, work very hard to standardize across the facilities and that's what we've been talking about we linked cancer registries to have complete cancer ascertainment. The NMD is very valuable in its large size and a number of participating facilities. We do need to be careful about the variability in those facilities because some will have very good cancer capture. I know several that link to state tumor registries as well like the BCSC where others might have very poor cancer capture if the women have a lot of opportunities at places to receive their follow-up care. So what we need to do is figure out a way to measure how complete cancer capture is at a facility level so that we can correct for that in the performance measures and comparing apples to apples. I would also suggest that we find ways to make it easier for facilities to link to cancer registries so that everyone can have high quality audit data.

D.L. Great. I'd like to move now not to just talking generically about screening benchmarks, but move to what we can actually do to improve performance. Dr. Lehman this question is for you, in both of the BCSC papers your group suggests that one method to improve performance could be to have programs to support second reviews of mammograms recalled by radiologists who have these high or over call a lot of mammograms. That would entail potentially 11 to 20 percent of mammograms overall and I'm just wondering if that's practical in today's busy environment where most radiologists already feel overworked.

C.D.L. I totally agree. We really need to think of options that are feasible. The idea of double reading has been implemented in lots of different ways across the country. There are some big practices where every mammogram is read by two radiologists. In others none of the mammograms are. What we're proposing is that sites might consider if they have a terrific mammographer who finds cancers but their recall rate is 18% which we all agree is just too high, to have that person review their recalls with a partner that has a much lower recall rate. That could really support the radiologist who's struggling with a high recall rate to bring that recall down. I love when Dr. D'Orsi said also our educational programs. What can we do more at our CMEs to support reducing false positives? We're looking for that area where we can maintain a high cancer yield but also keep our false positives as low as reasonable and as possible.

D.L. Thank you. So one of the issues that was brought up in the editorial was that of the United States FDA regulations requiring interpretation of only 480 examinations per year which is lower than that required by other national screening programs. I'm wondering if any of you have a comment about whether we should require higher minimum volumes in order to better our individualized statistics for assessing performance. So Dr. D'Orsi would you like to comment on that?

C.J.D. Yeah this was an issue that was discussed for many years at the European population based screening centers and much higher qualification rates to read mammography. The big argument here against that was, and I'm not saying was a good argument, was the issue of actually having enough facilities here and people not dropping out and reading mammograms because of that number issue. That was one of the issues that was or one on the problems that was discussed. Personally I think it should be increased. Perhaps not to some of the levels, I think and you may know this, but I think in Europe it's like 5,000 and it's 940 here. I think there's some happy medium over here. And I think Connie's point is excellent that to review your false positives with someone who is more experienced and perhaps has better metrics at the facility. So yes I think it could be increased without altering access to mammo facilities.

D.L. Thank you. And then Dr. Sickles I have a question that I'm not sure you can answer, but when we talk about all of these different problems with false positives and cancer detection rate, how do you think we should best set the performance bar that's a minimum expectation for practicing radiologists?

E.A.S. I think looking national performance, seeing where it is, and trying to get as close as possible to what's averaged there is a good target. I would mention also in addition to the suggestion that Dr. Lehman gave about double reading and mentoring radiologists, there's one thing that all radiologists can do, it's very simple, it's make sure or at least increase the likelihood that when you read a mammogram you have the prior exams available for comparison. We've known for decades that comparison with priors dramatically reduces recall rate without adversely affecting the cancer detection rate. So this is like a win-win. In many practices for a variety of reasons are in a situation where they accept patients who walk in for screening exams and don't have the priors available and they wind up with higher recall rates. Other practices try for patient convenience to read out the screening mammograms while the patient waits. Often in that situation they don't have the prior mammograms available. In our practice where we make a great effort to have priors available, we pretty much insist, there are always a few exceptions but we pretty much insist on having the priors available when we read out which is why our whole practice has a very low recall rate.

D.L. A very positive outcome from all of this discussion regarding metrics is that the BCSC has agreed to post yearly performance metrics and I think that this is a huge amount of work for the consortium but will be very important as we move from digital mammography to tomosynthesis. So Dr. Miglioretti given how involved you are in the BCSC do you have any ideas about how tomosynthesis will change these performance metrics?

D.L.M. Well we're just starting to look at tomosynthesis now in the breast cancer surveillance consortium as part of (inaudible) grants. I don't have data from our study but I'm really looking forward to seeing whether we also find reduced recall rates and possibly improve cancer detection with tomosynthesis. I think it's a very promising technology.

D.L. Terrific, and then just to sum up what we're talking about, obviously breast cancer diagnosis is a hot research topic and I was wondering if each of you could just briefly mention a new project that you're working on in this arena. Let's start with Dr. Lehman what are you working on?

C.D.L. There's so much going on right now that's incredibly exciting. One of these areas we've been talking about is how can we take big data to improve patient care? Dorothy Sippo is a recent recipient of a Giraffe Award and she's using informatics and big data to help radiologists during their regular routine clinical day get feedback about their recalls, about their recommendations for biopsies in a way that really supports them to learn more from their own cases. We're excited about that. We think we live in an exciting time with tomosynthesis. How can we make sure with this advanced imaging technology that we're teaching how to use it in a way that we really benefit from its promised potential of reduced recalls and higher cancer detection? We're excited about this area. We feel like we're just at the brink of the next revolution in early detection through screening and can't wait to see what the future holds.

D.L. Terrific. Dr. D'Orsi what are you working on?

C.J.D. We're currently evaluating a device that an Israeli company is manufacturing. What it does is through various imaging parameters it evaluates a woman's risk of malignancy and doesn't give a score but it suggests which patients should be recalled and/or sent for further imaging. Interestingly we just finished a little leader study on this device and the ROC curve showed a 15% increase in the AUC in the area under the curve which to me was very surprising because I'm very skeptical about everything and I was very, very surprised at that. So that's one device if it proves in a larger study to be fruitful and would help to decrease false positives in increased cancer detection rates. That part is very interesting and we're right now going to approach the NIH if they have enough money in their budget to try to do a larger study.

D.L. Terrific. Dr. Sickles what is your team working on?

E.A.S. One of the things that we've just been working on we just completed is a study looking at prior exams as I spoke before, but not just comparing with one prior but comparing with more than one prior. What we found in that study was that when you compare with more than one prior you actually have a lower recall rate and a higher cancer detection rate than when you compare with one prior. For all of you out there remember don't just look at the one, look at everything that's available if you want to get the best results. The second thing that we've been doing is we've been working with the NMD, the National Mammography Database. Even though their data are not anywhere near – have as anywhere near as complete cancer ascertainment as does the BCSC. They do have valuable data and one of the things that their data has shown recently is that when you look at screening outcomes in elderly women, that is women 75 and older, for whom there is less than complete data from randomized controlled trials because very few women, elderly women, were studied in randomized trials. When you

look at the NMD data it's very strongly indicative that the benefits that have been proved for screening in women up to the age of 75 continue up to at least the age of 90.

D.L. Wow. And then finally Dr. Miglioretti can you tell us something about what you're working on?

D.L.M. Sure. The BCSC is working towards trying to come up with a more personalized approach to both breast cancer screening and surveillance in women with a history of breast cancer. And so we hope to use personal risk factors and breast density to identify women who may need more frequent screening or may benefit from supplemental imaging with breast MRI or breast ultrasound or tomosynthesis or who might want to start screening earlier than recommended by guidelines.

D.L. Well I'd just like to thank all four of you for participating today and for all of the work that you do for breast cancer screening and for imaging our patients. Thank you very much and we'll look forward to getting some of those papers you just discussed submitted to *Radiology*.