

Mammography: Yet Another Challenge¹

Daniel B. Kopans, MD

Mammographic screening is one of the major, unrecognized, medical advances of the past 35 years. Prior to 1990, the death rate from breast cancer had remained unchanged in the United States for at least 50 years. In the middle of the 1980s, the decrease in breast cancer deaths reported from the Swedish Two-County Trial, which was a randomized, controlled trial, led to the onset of widespread screening in the United States that involved enough women to influence national statistics (1). There was a sudden increase in the incidence of breast cancer that began around 1985–1986 that, initially, raised concern that there was an “epidemic” of breast cancer. It was soon recognized that this sudden increase was due to the fact that, with screening, we were finding “future” cancers years earlier. At the same time, the incidence of ductal carcinoma in situ also began to increase. Since ductal carcinoma in situ is, virtually, only found by using mammography, this increase confirmed the fact that mammographic screening had begun on a large scale. Periodic screening is unlikely to affect the fast-growing cancers (length bias), but it is likely to interrupt the moderately growing and slower-growing cancers. Thus, it is not surprising that the death rate from breast cancer began to decrease in 1990, 5–7 years after the onset of screening (2,3).

This decrease in deaths has continued as more and more women participate in screening so that the death rate has decreased by almost 30% since 1990 (4). Some suggest that better therapy is responsible for the decline. In 2005, seven computer models were used to try to determine what percentage of the decline in deaths resulted from better therapy and what percentage resulted from early detection (5). The modeling gave a range for the con-

tribution of early detection from 28% to 65%. Citing the lower estimate, this computer modeling has been used to suggest that therapy is the reason for fewer deaths. It is surprising that so much reliance has been placed on computer modeling when there are actually several direct measures that have been published that aided analysis of actual population-based data that clearly show that mammographic screening is the major reason for the decline in breast cancer deaths. A review of the data from the “two counties” from the Swedish Two-County Trial showed that the death rate declined over time in direct proportion to the number of women participating in screening, while those who did not participate, but had access to the latest therapeutics, had only a very small decline in death rate (6). A subsequent study of seven counties, which included 30% of the Swedish population, confirmed the fact that screening accounted for most of the decline in deaths (7). The benefit from screening mammography has also been demonstrated in the Netherlands, where despite access to modern therapy, the death rate from breast cancer had continued to increase in the various Dutch health care districts. It was not until screening became available that the death rates began to decline (8).

To those of us who have been involved from the very early years of mammographic screening, it would appear that the challenges never end. We are now entering a new era, with a major reassessment of health care. Interventions will be evaluated with even greater scrutiny, and the benefits of mammography will once again be challenged. The articles by Miglioretti et al (9) and Elmore et al (10), in this issue of *Radiology*, provide important information on the sensitivity and specificity of mammographic screening and how they vary with the train-

Published online
10.1148/radiol.2533091517

Radiology 2009; 253:587–589

¹ From the Department of Radiology, Avon Comprehensive Breast Evaluation Center, Massachusetts General Hospital, Harvard Medical School, 15 Parkman St, ACC240, Boston, MA 02114. Received August 13, 2009; final version accepted September 8. Address correspondence to the author (e-mail: dkopans@partners.org).

Author stated no financial relationship to disclose.

See also the articles by Elmore et al and Miglioretti et al in this issue.

© RSNA, 2009

ing of the practitioners. These results can be used to improve our ability to find more cancers earlier, or they can be misused to deny women access to mammographic screening.

Those of us involved in fellowship training will be happy to learn that radiologists with fellowship training had a higher cancer detection rate than did radiologists without additional training and that radiologists with greater experience also had a higher breast cancer detection rate than did those without postresidency training.

When one analyzes the findings, however, one should also be aware that a crucial set of data are missing. Decreased deaths associated with screening are a result of a decrease in the size and stage of the cancers detected, and these data have not been provided in the reviews by Miglioretti et al (9) and Elmore et al (10). Detecting breast cancer at a smaller size and earlier stage is the fundamental value of mammographic screening. Two radiologists can have similar detection rates (apparent sensitivity), but one may be finding cancers a year later than the other, and these cancers are larger and of a later stage than those detected by the other radiologist. Just reviewing sensitivity and specificity may be misleading without information as to the size and stage of the cancers at the time of detection.

What is also missing from these articles is a detailed description of the consequences of the studies with false-positive findings and their relative importance. These women were screened, and something was seen on their screening mammogram that concerned the interpreting radiologist. The patient was asked to return for a diagnostic evaluation, which usually means a few extra mammographic images and/or an ultrasonographic study that quickly help resolve the issues raised with most recalls. Although the authors do not provide any details, most of the questions raised by a screening study are usually resolved with just a few extra x-ray images. Recalls are inconvenient, they may result in time away from work or

family, and they may provoke anxiety among the recalled women. Elmore et al (10) showed that, among the radiologists who detected the most cancers, 83 women were recalled from screening to detect one additional cancer. There is no objective measure to decide whether this is an excessive number or a reasonable one, given that the woman whose breast cancer is detected may not die as a result. The authors provide no frame of reference, such as other interventions with which to compare this ratio. For example, it is my understanding that a similar proportion of women are recalled because of an abnormal cervical cancer screening study, yet four times as many women die of breast cancer each year as women who die from cervical cancer. Why is the breast cancer recall rate considered excessive?

Elmore et al (10) also provide somewhat of a double message. On the one hand, the article seems to be critical of fellowship-trained radiologists because they have a higher false-positive rate, but then they concede that the same radiologists had a greater sensitivity (found more cancers). My experience suggests to me that there is a fundamental difference in the way health planners and physicians look at breast cancer screening. It is my experience that radiologists are trying to find as many early cancers as we can because this action results in saving lives. My experience with health planners is that they seem to be most interested in reducing studies with false-positive findings. I wholeheartedly support improvements in what we do, but we should not lose sight of the fact that the anxiety and inconvenience of being called back because of an abnormal finding on a screening mammogram are not equivalent to dying from breast cancer. In the effort to reduce studies with false-positive findings, we want to try to avoid missing the all-important small cancers. D'Orsi and Swets (11) and others (12) explained why this is difficult. Radiologists who interpret findings from studies by using established criteria place themselves on the

same receiver operating characteristic curve. This curve relates studies with true-positive findings to those with false-positive findings. Unless one of us has some secret that allows us to differentiate benign from malignant lesions that we do not share with our colleagues, which permits us to reduce studies with false-positive findings without allowing some cancers to be missed, the studies with false-positive findings can only be reduced by increasing false-negative interpretations, causing a reduction in cancer detection. A major benefit from the data sets used to compile the articles by Miglioretti et al (9) and Elmore et al (10) would be a review of the cases recalled from screening to try to determine what findings triggered the higher false-positive recall rates that did not trigger recalls among more experienced radiologists and what findings produced higher false-positive rates among fellowship-trained radiologists, along with a higher sensitivity. It is helpful to examine the three groups of radiologists, but the only way to modify behavior would be to define the false-positive findings that did not prove to be cancers so that radiologists can learn not to be concerned about these findings.

We all would like to reduce the false-positive rate, as long as we can keep the sensitivity for small cancers high. It would be helpful if Miglioretti et al (9) and Elmore et al (10) could take advantage of their access to so much potentially useful data. It is fine to suggest that "[f]ellowship training programs in breast imaging should emphasize lowering radiologists false-positive rates to within the recommended U.S. performance goals while maintaining high sensitivity," but the authors offer no way to accomplish this task. I would urge that they try to determine what the radiologists with the highest sensitivity and lowest false-positives were doing that was not being done by the others. Miglioretti et al (9) and Elmore et al (10) could make a major contribution by reviewing the cases of the 18 radiologists who had both sensitivity and false-

positive rates that were in the highest quartile of interpretive performance to determine the criteria that they used so that these criteria could be taught to all of us to benefit those with lower sensitivity and specificity and to continue to drive down the breast cancer death rate.

References

1. Tabár L, Fagerberg CJG, Gad A, et al. Reduction in mortality from breast cancer after mass screening with mammography. *Lancet* 1985;1:829-832.
2. Kopans DB. Beyond randomized, controlled trials: organized mammographic screening substantially reduces breast cancer mortality [letter]. *Cancer* 2002;94(2):580-581.
3. Tabár L, Vitak B, Chen HH, Yen MF, Duffy

- SW, Smith RA. Reply. *Cancer* 2002;94(2):581-583.
4. SEER cancer statistics review 1975-2006. Table 4.8. Cancer of the female breast (invasive): age-adjusted U.S. death rates by year, race and age. National Cancer Institute Web site. http://seer.cancer.gov/csr/1975_2006/browse_csr.php?section=4&page=sect_04_table.08.html. Accessed July 28, 2009.
5. Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med* 2005;353(17):1784-1792.
6. Tabár L, Vitak B, Chen HH, Yen MF, Duffy SW, Smith RA. Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. *Cancer* 2001;91(9):1724-1731.
7. Duffy SW, Tabár L, Chen H, et al. The impact of organized mammography service

- screening on breast carcinoma mortality in seven Swedish counties. *Cancer* 2002;95(3):458-469.
8. Otto SJ, Fracheboud J, Looman CW, et al. Initiation of population based mammography screening in Dutch municipalities and effect on breast-cancer mortality: a systematic review. *Lancet* 2003;361(9367):1411-1417.
9. Miglioretti DL, Gard CC, Carney PA, et al. When radiologists perform best: the learning curve in screening mammographic interpretation. *Radiology* 2009;253(3):632-640.
10. Elmore JG, Jackson SL, Abraham L, et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. *Radiology* 2009;253(3):641-651.
11. D'Orsi CJ, Swets JA. Variability in the interpretation of mammograms [letter]. *N Engl J Med* 1995;332(17):1172.
12. Reply. *N Engl J Med* 1995;332(17):1173.