

# Peer Review of Mammography Interpretations in a Breast Cancer Screening Program

## ABSTRACT

Mammograms from a statewide screening program were subjected to a blind review by a panel of expert mammographers. Ninety-five percent (173/182) of original normal mammograms and 53% (164/311) of original abnormal mammograms were reread as normal. In comparison with the expert panel, community radiologists were more likely to request a repeat mammogram in 6 months than to interpret a mammogram as normal or address their uncertainty with an immediate diagnostic workup. (*Am J Public Health*. 1995;85:837-839)

Judith Feldman, MD, MPH, Robert A. Smith, PhD, Ruthann Giusti, MD, MPH, Barbara DeBuono, MD, MPH, John P. Fulton, PhD, and H. Denman Scott, MD, MPH

### Introduction

In November 1987, the Rhode Island Department of Health launched the Rhode Island Breast Cancer Screening Program to promote routine screening for breast cancer among Rhode Island women 40 years of age and older. The program included breast cancer surveillance; monitoring of knowledge, attitudes, and screening behavior; promotion and provision of screening mammography; and program evaluation. Nonpregnant, non-breast-feeding women 40 years old and older who had no symptoms of breast cancer and who had not had a mammogram within the previous 12 months could schedule a screening mammogram through the program by calling a toll-free telephone number at the Rhode Island Department of Health. Mammograms were done through community radiology sites that agreed to participate in the program and met the following criteria. First, the sites were required to use dedicated mammography or xerography units. Second, they had to charge no more than \$50 for a screening mammogram. Third, mammograms had to be reviewed by a board-certified radiologist. Finally, the sites were required to participate in the peer review process.

Virtually all (92%) licensed radiology facilities doing mammograms in Rhode Island participated in the program and submitted films for panel review. Bilateral craniocaudal and lateral views were done and processed in the same manner as other mammograms at the site, and they were classified by community radiologists (the original readers) into one of four diagnostic categories: normal, repeat 6 months, special views needed, or suspicious of malignancy.

As a means of assessing the accuracy of mammographic interpretation for women screened in the program and exploring effective ways to perform quality assurance for mammography, a panel review system was developed in which the interpretations of a group of community

radiologists were compared with the interpretations of an expert panel of radiologists (as described below). In comparison with the expert panel, community radiologists were more likely to interpret films as nonnormal.

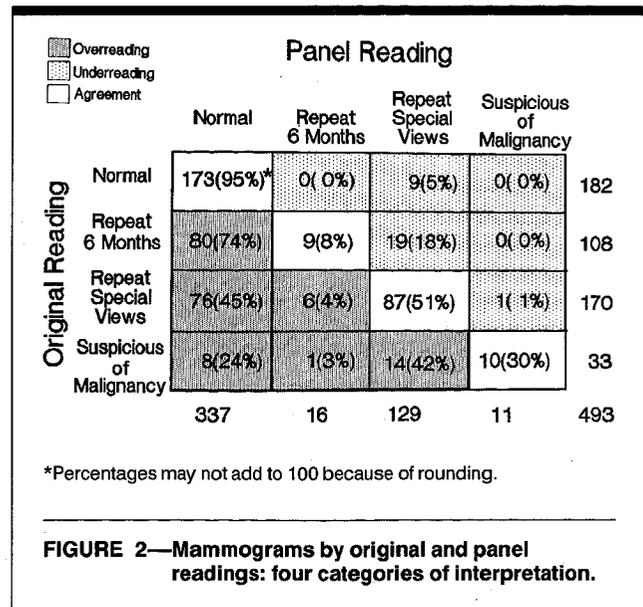
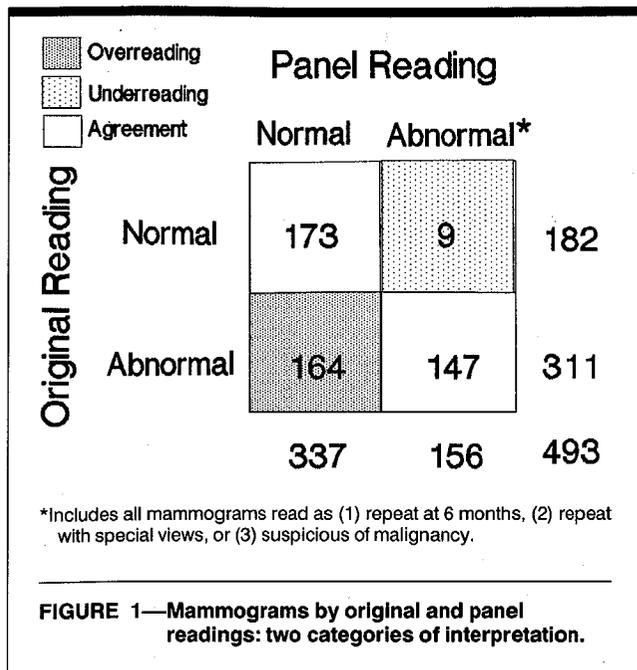
### Methods

All nonnormal screening mammograms performed for the Rhode Island Breast Cancer Screening Program between November 1, 1987, and October 31, 1988, as well as a 10% random sample of normal mammograms, were reinterpreted by a panel of radiologists with special expertise in screening mammography. At monthly sessions, 50 films (mixed normal and abnormal mammograms) were categorized as normal, repeat 6 months, special views needed, or suspicious of malignancy by a radiology panel composed of a nationally prominent diagnostic mammographer (who chaired the panel), two Rhode Island radiologists with advanced mammography training, and a rotating panel member representing one of the participating facilities. The panel was unaware of the original interpretation of the films, and, except for the rotating community radiologist, the panel composition was stable throughout the study. After discussion with panel members, the chair classified each mammogram using one of the four diagnostic categories. Only the final panel result was registered; data on intrapanel disagreements were not recorded. Comparisons were made be-

Judith Feldman, Barbara DeBuono, John P. Fulton, and H. Denman Scott are with the Rhode Island Department of Health, Providence. Robert A. Smith and Ruthann Giusti are with the Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, Ga.

Requests for reprints should be sent to Judith Feldman, MD, MPH, Division of Preventive Health Services, Rhode Island Department of Health, 3 Capitol Hill, Providence, RI 02908.

This paper was accepted November 8, 1994.



tween panel and community interpretations; agreement within categories of community (original) interpretations was described with proportions and assessed with Spearman's rank correlation coefficient. Two-standard-error 95% confidence intervals (CIs) were calculated for the proportions discussed here.

## Results

A total of 2158 screening mammograms were performed for the program between November 1, 1987, and October 31, 1988; 1828 (85%) were interpreted as normal and 330 (15%) as nonnormal by community readers. The expert panel was almost twice as likely to agree with community radiologists that a film was normal (95%; 95% CI = 92, 98) than to agree that a film was nonnormal (47%; 95% CI = 41, 53; Figure 1). Among films read in the community as nonnormal (Figure 2), the panel agreed with 34% (95% CI = 29, 39) of the original recommendations (Spearman's rank correlation = .54). The panel was least likely to agree with original recommendations of repeat 6 months; 74% (95% CI = 62, 82) were reread as normal, and 18% (95% CI = 11, 25) were reread as requiring special views.

Because the mammograms interpreted by the panel as suspicious of malignancy represented a subset of the mammograms interpreted by community radiologists as suspicious of malignancy, the proportion of cases biopsied and

found to be malignant could be compared (Table 1). Seventy-five percent (95% CI = 60, 90) of cases interpreted by community radiologists as suspicious of malignancy were biopsied, and 36% (95% CI = 19, 53) were found to be malignant. In contrast, 91% (95% CI = 74, 100) of cases interpreted by the expert panel as suspicious of malignancy were biopsied, and 91% (95% CI = 74, 100) were found to be malignant.

Two in situ intraductal carcinomas originally interpreted as suspicious of malignancy were reinterpreted by the panel as normal. When the reviewers rescrutinized these two films knowing the original diagnosis, in one case they concluded that they would not have interpreted the film as nonnormal; in the other case, they concluded that they might have recommended repeat 6 months.

## Discussion

The 15% rate of abnormal screening mammograms reported by community radiologists in this program is higher than that of contemporary screening programs.<sup>1-3</sup> In comparison with the expert panel, community radiologists were more likely to interpret films as nonnormal. Fifty-three percent (95% CI = 47, 59) of films originally read by the community radiologists as nonnormal were reread by the panel as normal. In comparison, only 5% (95% CI = 2, 8) of films originally read by the community radiologists as

normal were reread by the panel as nonnormal (requiring special views).

There is some controversy about the use of Spearman's rank correlation statistic in analyzing these data. Suggestions from several expert sources have varied and have included kappa, Kendall's tau, and Spearman's rank correlation. Given the nature of the data and the problems with each of the statistics just mentioned, we believe that Spearman's rank correlation coefficient is as useful as any to summarize the level of agreement between the community and the expert panel.

Biopsy results revealed two intraductal carcinomas for films interpreted as normal by the expert panel. As unsatisfying as these results may be to some, mammography is imperfect, even in the hands of experts, and the results of this study are certainly not unique. For screening mammography to achieve its fullest potential to reduce morbidity and mortality from breast cancer, careful attention must be focused on the quality of the entire screening process, including the technical aspects that contribute to image quality, the interpretation of films, and follow-up in those instances in which further tests are recommended. Factors that lead to underreading or overreading are varied and can include fear of litigation, poor image quality, and lack of mammographer training and experience. Overreading increases the financial costs of breast cancer screening, results in

**TABLE 1—Comparison of Cases Biopsied and Cases Malignant among 33 Cases Originally Considered Suspicious of Malignancy by the Community Radiologists**

	Total Biopsied, No.	Biopsied, No. (%)	Found to Be Malignant, No. (%)
Community radiologists	33	25 (75 <sup>a</sup> )	12 (36)
Expert panel	11	10 (91)	10 (91)

<sup>a</sup>In 25% of the cases categorized as suspicious for malignancy, subsequent workup yielded diagnosis without the need for biopsy.

avoidable anxiety to the women screened, and causes an avoidable number of negative biopsies. As nationally accepted screening guidelines are implemented on a wide scale, these costs can become significant when projected to the state and nation, with the potential that screening may become too costly to be justified. Underreading, on the other hand, results in the tragic cost of delayed detection and missed opportunities for early intervention.

True overreading can be determined only by diagnostic workup and biopsy, and underreading can be determined only by long-term follow-up of screened women. Peer review, however, can serve as an estimate of the extent of these problems and can also help to improve the experi-

ence and expertise of mammographers. Some radiology sites already engage in the peer review process.<sup>2</sup> Double reading is not particularly cumbersome, can provide feedback to radiologists, and can provide a reasonably easy opportunity for colleagues to consult on abnormal or difficult cases. In areas where demand for mammography is still low or where supply outstrips demand, double reading can provide radiologists with the volume of films necessary to develop and maintain their expertise.

In addition, ensuring the technical quality of the images can minimize false negatives and false positives. Although technical quality was not assessed in this study in a manner sufficiently standardized to allow definitive analysis, the panel noted problems with some aspect of technical quality in approximately two thirds of the films reviewed. These problems with technical quality may have led to problems with interpretation.

As with all screening tests, there will always be a "trade-off" between false negatives and false positives. The trade-off is particularly unclear for the slow-growing ductal carcinoma in situ. As protectors of the public's health, public health agencies can play an important role in developing policy and ensuring implementation of standards that improve the quality of screening tests to the fullest extent possible. For mammography, this requires that the technical quality of the films and clinical expertise of the readers be optimized. Results of the Department of Health's peer review study were instrumental in developing regulations for the practice of mammography in Rhode Island. New standards implemented in 1992

now require that all facilities in the state have comprehensive quality assurance programs, including mammographer training, radiology technologist qualifications, and image quality standards. The State-wide Committee for Quality Assurance in Mammography was formed to monitor and promote the quality of films and their interpretation. The state is also exploring the possibility of repeating this study. It will be important to determine whether the improved quality assurance standards help decrease underreading and overreading. □

### Acknowledgments

This study was supported by grant U50-CC0102929-01, Division of Chronic Disease Control and Community Intervention, Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control.

We wish to thank the expert panel members, Gillian Newstead, MD, Norman Sadowsky, MD, and Barbara Schepps, MD, for their insight and their time and patience in reviewing the mammograms for this study and Dolores DiOrto and John Ferruolo of the Rhode Island Department of Health for their invaluable assistance. In addition, this study would not have been possible without the outstanding cooperation of the entire radiologic community in Rhode Island.

### References

1. Tabar L, Gad A. Screening for breast cancer: the Swedish trial. *Radiology*. 1981; 138:219-222.
2. Bird RE. Low-cost screening mammography: report on finances and review of 21,716 consecutive cases. *Radiology*. 1989;171:87-90.
3. Sickles EA, Ominsky SH, Sollitto RA, Galvin HB, Monticciolo DL. Medical audit of a rapid-throughput mammography screening practice: methodology and results of 27,114 examinations. *Radiology*. 1990;175: 323-327.