Published in final edited form as:

Acad Radiol. 2007 September ; 14(9): 1036–1042.

Computer-Assisted Mammography Feedback Program (CAMFP): An Electronic Tool for Continuing Medical Education

Nicole Urban, ScD^{*}, Gary M Longton, MS^{*}, Andrea D Crowe, BS^ψ, Mariann J Drucker, MD[§],
Constance D Lehman, MD^ψ, Susan Peacock, MS^ψ, Kimberly A Lowe, MS^{*}, Steve B Zeliadt,
PhD^{*}, and Marcia A Gaul, BS^{*}

^{*} *Fred Hutchinson Cancer Research Center*

[§] *Seattle Radiologists*

^ψ *University of Washington*

Abstract

RATIONALE AND OBJECTIVES: Our goal was to develop and evaluate software to support a computer assisted mammography feedback program (CAMFP) to be used for continuing medical education (CME).

MATERIALS AND METHODS: Thirty-five radiologists from our region signed consent to participate in an IRB-approved film-reading study. The radiologists primarily assessed digitized mammograms and received feedback in 5 film interpretation sessions. A bivariate analysis was used to evaluate the joint effects of the training on sensitivity and specificity, and the effects of image quality on reading performance were explored.

RESULTS: Interpretation was influenced by the CAMFP intervention: Sensitivity increased (Δ sensitivity = 0.086, $p < 0.001$) and specificity decreased (Δ specificity = -0.057, $p = 0.04$). Variability in interpretation among radiologists also decreased after the training sessions ($p = 0.035$).

CONCLUSION: The CAMFP intervention improved sensitivity and decreased variability among radiologist's interpretations. Although this improvement was partially offset by decreased specificity, the program is potentially useful as a component of continuing medical education of radiologists. Dissemination via the web may be possible using digital mammography.

Keywords

breast radiography; mammography; continuing medical education; radiologist; computer; women

INTRODUCTION

Mammography is used widely in developed countries to detect breast cancer early enough to treat successfully. Screening accuracy could be enhanced by either improving image quality or by improving the accuracy of radiologists who interpret mammograms [1]. Differences in radiologists' training, experience and reading volume may affect their clinical recommendations and interpretations [2-6]. Regulations set forth by the Food and Drug Administration's (FDA) Mammography Quality Standards Act (MQSA) [7] require continuing

Address of work origin and corresponding author: Fred Hutchinson Cancer Research Center 1100 Fairview Avenue N., M2-B230 Seattle, WA 98109-1024 Phone: 206-667-6774 Fax: 206-667-7264 Email: nurban@fhcrc.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

medical education (CME) for all radiologists who interpret mammograms [8]. Few studies have evaluated the specific effectiveness of CME on improving physician performance in reading mammography since the implementation of MQSA; however, results suggest modest advances in performance among 23 practicing radiologists who attended a one-day lecture on BIRADS assessment [9,10]. In general, CME has proven to enhance overall physician performance in many fields of medicine, with the greatest benefit achieved from learning sessions that allow hands-on practice in contextually relevant or difficult areas [10,11]. This type of learning format is more beneficial than traditional lecture-based programs because it allows the physicians to be interactive and engaged during CME [12].

The FDA does not specify the format in which CME units are obtained and allows radiologists to complete web-based or computer-based programs [8]. Therefore, missed cancers and variations in mammography interpretation may be reduced by providing practice in reading difficult films with feedback on physician accuracy during CME using the computer-assisted mammography feedback program (CAMFP). CAMFP was initially designed to be used as a continuous teaching tool for radiologists in low-volume isolated practice areas. However, since it is a computer-based format that can be easily obtained over the internet [13], it could contribute to CME programs that are used by radiologists with varying experience levels.

To explore the possibility that reading performance could be enhanced by using this program, we conducted a study to test the hypothesis that providing low-volume radiologists with immediate feedback on their interpretations of difficult mammograms would improve their reading skills and result in improved sensitivity and/or specificity of mammography interpretation. To ensure adequate power to detect changes in the most clinically relevant outcomes, the study was designed using sensitivity and specificity as the basis for evaluation [1].

We selected sensitivity and specificity as the primary outcome measures over ROC based measures because they provide a more clinically-relevant and conceptually straightforward method for comparing change in reading accuracy related to intervention. The five levels of the ordinal mammography rating scale have inherent meaning which effectively reduces them to at most three clinical decision categories. It is possible to erroneously conclude an intervention was not clinically beneficial based on similar pre and post-intervention ROC curves which can occur, for example when sensitivity is increased by the intervention without decreasing specificity. This phenomenon would not result in an obvious improvement to the ROC curve but could be a clinically-relevant finding [1]. At the time of this study, digital mammography had not yet been introduced into clinical practice[14,15]. Therefore, we assessed CAMFP by providing both digitized images and screen-films to the radiologists.

MATERIALS AND METHODS

Subjects

Radiologists were recruited to participate in a film-reading study that was approved by our institutional review office. Regional mammography facilities certified by the FDA were surveyed in 1996 to identify low volume mammography radiologists. Responses were received for 80.5% (99/123) of the regional practices. We invited 69 radiologists, representing 37 mammography facilities from communities in our region, to participate in the CAMFP program, of whom 35 agreed to participate and signed informed consent for the study. The current low volume requirement established by MQSA in 1999 requires radiologists to read at least 960 mammograms in a two year period [8]. Since the participants of this study were recruited in 1998 (one year before the enforcement of MQSA), that requirement was not used as a criterion for selection. Instead, subjects were recruited if they reported reading fewer than 300 mammograms per year. Power analysis based on computer simulations suggested that we

would have 95% power to detect a change of 0.10 in sensitivity and/or specificity when using 45 mammograms in a set with a minimum of 30 radiologists. All 35 radiologists participated in the first 4 sessions, which were required for training and to test the hypothesis of interest. We chose not to use a control group, but rather focused on within group comparisons for two reasons [1]. First, CME credit could only be offered to radiologists in the intervention arm; therefore, it is possible that radiologists would not be motivated to continue in the study if they were assigned to the control arm. Second, it is likely that radiologists in the control arm would learn from the baseline assessment, which would make it difficult to maintain a true control group throughout the study.

Software

The CAMFP software was designed to run on a laptop computer (for use in the study) or over the Internet. Mammograms were presented as TIFF (Tagged Image File Format) images, which were the highest quality graphic available at the time; they were 8-bit grayscale images with 0-255 intensity range and 8-bit depth, with the dimension of 768×1134 pixels. The radiologist could review the image in the 4 standard views (Left craniocaudal, Right craniocaudal, Left mediolateral oblique, Right mediolateral oblique) and/or zoom into a single view. He was given a simple case history and asked to assess the case by clicking a single BIRADS assessment (1, 2, 4 or 5). If the radiologist selected BIRADS 4 or 5, he was asked to click on a box to indicate if the suspicious area was a mass, calcification, architectural distortion, density (focal or asymmetric), or a combination of these choices, and on the suspicious area to locate the lesion before moving on to review the next case. For cases that he missed, at the end of the session an arrow on the image pointed to the suspicious area or lesion and the finding was described. Examples of images reviewed by the radiologists are depicted in Figure 1.

Case Selection

Teaching mammograms were selected from the files at four breast care centers. Participating practices were also invited to contribute their own difficult cases for use in the education (not evaluation) sessions. Expert mammographers judged all cases for suitability and film quality including appropriate resolution, optical density and positioning of breast tissue. All mammographic studies were stripped of identifying information. To yield a baseline sensitivity and specificity of 0.7-0.8 for each set of films, a criterion necessary to achieve 95% power, cases were included if they were clearly positive or negative yet judged moderately to extremely difficult to interpret. Cases were excluded if they were coded as BI-RADS 3 [1] or if they could not be copied and digitized successfully. All malignant cases were confirmed by pathology.

Composition of the Film Sets

The case compositions are shown in Table 1, including classification as a mass, calcification, density or architectural distortion.

Sessions

Each radiologist participated in 5 sessions over an 11 month period. Radiologists were given as much time as they needed to review the film sets; on average, sessions lasted for one hour. The time interval between sessions ranged from 6 weeks to 3 months. Radiologists used their own view boxes, roller readers, hot lights, magnifying glasses and other equipment. Study staff hung copies of the films so that the radiologist could review both the digitized images and the films for each case. The radiologist used the CAMFP software to record his BI-RAD assessment for each case, to indicate on the digitized image where he believed the malignancy was located, and to receive feedback after assessments were made for all cases in the set. All of the sessions were presented as an exercise in both training and in testing, but in fact, results from the sessions were used differently to evaluate the effects of training. The first session was

used to obtain baseline sensitivity and specificity estimates for each physician for testing the hypothesis that training improves accuracy. The second session was for training. The third session was used for training and to establish baseline sensitivity and specificity for evaluating the accuracy with which radiologists interpreted the digitized images compared to the screen-films. The fourth session was used as the follow-up evaluation to assess changes in sensitivity and specificity following the intervention. The fifth session was used to obtain follow-up sensitivity and specificity values for assessing the accuracy of interpreting digitized images compared to screen films. Participating radiologists received two CME credits per film-reading session for a total of 10 credits for this study.

Films from 180 mammography subjects were used: 90 to test the hypothesis in session 1 and 4 and 90 for educational purposes in session 1, 3, 5. Each “film” was defined as a screening mammogram film study, including 4 images (2 from each breast). Films used for testing the hypothesis were randomized into 2 reading sets (A & B) of 45 films each, with balance across sets on disease status (20 with and 25 without malignancy) and age (under/over 50 years) achieved through permuted block randomization. The radiologists were informed that prevalence of malignancy in the sets was higher than would be encountered in practice, since this would become apparent during the first session and might affect reading of subsequent sets [1]. Radiologists were randomly assigned to Group I or Group II. A crossover design was used to avoid systematic bias in accuracy assessment due to film set differences: 17 radiologists (Group I) read film set A pre-intervention and set B post-intervention, while the other 18 (Group II) read set B first.

Films from 90 selected mammography subjects were used in sessions 2, 3 and 5 for educational purposes: 41 with breast cancer and 49 without. The 90 films were randomized into 3 reading sets of 30 films each (denoted as sets C, D, and E), with balance across sets on disease status and age (under/over 50 years) achieved through permuted block randomization. The resulting film sets each included images from 16 (2 sets) or 17 (1 set) subjects with breast cancer and 14 or 13 without disease. To evaluate the accuracy with which radiologists interpreted the digitized images compared to the screen-films, 11 of the 34 radiologists read film set C in session 3 and set D in session 5; 11 radiologists read sets D and E respectively; and the remaining 12 radiologists read set E followed by C. None of the images or subjects used to test the hypothesis (sets A and B) were used in sessions 2, 3 or 5.

Mammography accuracy measures

Effects of the intervention on sensitivity, specificity and inter-radiologist variability were assessed to test the hypothesis that the CAMFP improved mammography accuracy [1]. Sensitivity was defined as the proportion of diseased subjects properly classified with a true positive screen. A positive screen was defined as any woman receiving an assessment code of 4 or 5 on either breast. A true positive screen was defined as a woman with a malignancy receiving assessment 4 or 5, with the lesion correctly identified with respect to laterality. Specificity was defined as the proportion of non-diseased subjects correctly classified with a negative screen, defined as assessment code 1 or 2 on both breasts.

Radiologists' Characteristics and Variability in Reading Accuracy

The characteristics of radiologists' training and experience in reading mammograms were assessed for each group (Table 2).

Data Analysis

Intervention effects were measured as the individual radiologist-level differences in sensitivity and specificity, and inter-radiologist variability between assessments at baseline (pre-intervention) and follow-up (post-intervention). Tests for intervention effects on sensitivity

and specificity accounted for the stratified reading group design[1]. The joint effects of the intervention on sensitivity and specificity were also evaluated. Because sensitivity and specificity tend to be inversely related, a bivariate approach was used to account for the lack of independence inherent in these measures. A chi-square test statistic, based on the mean changes in sensitivity and specificity and their covariance, was performed and results were shown graphically[1]. To evaluate change in inter-reader variability between baseline and follow-up sessions, Box's M-test was used to test the equivalence of sensitivity-specificity covariance matrices between these sessions[16]. The effect of image quality on screening performance was evaluated during a fifth session, in which radiologists were required to interpret the digitized images without access to screen films. Sensitivity and specificity values obtained from this session were compared to values obtained in session three.

RESULTS

Characteristics of Participating Radiologists

Table 2 provides the characteristics of participating radiologists, as reported in 1996 in a mailed survey. There are two reading groups, denoted I and II in Table 2. Mammography CME credits, mammograms read per year and mammograms read in the past year are higher in Group II. Other characteristics of radiologists are evenly distributed across the two groups.

Change in Sensitivity and Specificity Attributable to the Intervention

Results of the changes in sensitivity and specificity from baseline to follow-up for each of the two reading groups are given in Table 3. For both groups combined, there was a statistically significant increase in sensitivity (combined Δ sensitivity = 0.086, $p < 0.001$) and a statistically significant decrease in specificity (combined Δ specificity = -0.057, $p = 0.04$).

Figure 2 shows sensitivity and specificity differences for the 35 readers and a 95% confidence region for the joint mean sensitivity and specificity intervention effect (Δ sensitivity, Δ specificity). The figure depicts that improved sensitivity and decreased specificity attributable to the CAMFP were consistent between the two groups.

A test to evaluate the joint effects of the intervention on sensitivity and specificity was used to account for the correlation between these two measures [1]. The results allow us to reject the null hypothesis that both sensitivity and specificity are unchanged post-intervention ($p = 0.001$).

Variability in Sensitivity and Specificity among Radiologists

The observed joint reader-level sensitivities and specificities are shown in Figure 3. The comparison of scatterplots for sessions 1 and 4 illustrates how the improved sensitivity is accompanied by a decrease in specificity post-intervention. The negative correlation between sensitivity and specificity may reflect the variability among radiologists in their thresholds for recommending biopsy.

As depicted by the decreased spread of the plotted points in session 4 relative to session 1, variability among radiologists decreased following the training sessions. A two dimensional likelihood ratio test [16] for the equivalence of the sensitivity-specificity covariance matrices between sessions 1 and 4 supports this suggestion, yielding $p = .035$.

Accuracy was adversely influenced by inferior image quality. Specificity did not change between session three and five, but a statistically significant decrease in sensitivity was observed in the absence of screen-films (Δ sensitivity = -0.22, $p < 0.001$).

DISCUSSION

Research suggests there is an opportunity to improve early cancer detection through increased training of radiologists [17]. A recent effort has been made to provide training and self-assessment workshops for radiologists at national conferences (such as the Radiological Society of North America 2005 annual conference); as well as to develop mammography teaching tools that utilize electronic formats [13]. To determine the effects of practice reading films with immediate feedback on mammography accuracy, we tested the hypothesis that a computer-assisted mammography feedback program (CAMFP) would improve the sensitivity and/or specificity of mammography interpretation. Radiologists were not permitted to use BIRADS category 3 due to its vague clinical implications. When we designed this study, our intention was to develop a novel electronic teaching tool that could be easily disseminated over the internet once the technology for digital mammography improved. After receiving feedback on their interpretations of 105 films, radiologists' sensitivity increased for both reading groups in our study (combined Δ sensitivity = 0.086, $p < 0.001$), but at a cost of decreased specificity (combined Δ specificity = -0.057, $p = 0.04$). The CAMFP affected radiologists' perceptions of images, increasing their ability to detect cancers and improving sensitivity. However, an unequivocal improvement in accuracy requires improvement in both sensitivity and specificity. The CAMFP may have lowered radiologists' thresholds for positivity, thereby increasing sensitivity and decreasing specificity. It is possible that our results reflect both effects.

Since the completion of this study in 1999, digital mammography has become commonplace and radiologists are accustomed to reading digitized images. Dissemination of CAMFP may now be feasible using digital mammography. A recent study of 49,528 asymptomatic women has shown the diagnostic accuracy of digital mammography was better than film mammography for women who were under the age of 50 years, pre-menopausal or peri-menopausal, or had radiologically-dense breasts [18]. Their results showed no difference in sensitivity for the two mammogram modalities ($p = 0.92$); however, there was a statistically significant decline in the specificity of film mammography compared to digital mammography for all women combined (Δ specificity = -0.0003, $p = 0.0006$) [18]. No statistically significant differences were observed when they stratified by age, menopausal status, or breast density. Our assessment of the effects of image quality on performance suggests that efforts to improve image quality may be very fruitful.

A limitation of our study is that we used laterality, rather than exact location of the lesion, to define a true positive for estimating sensitivity. The CAMFP software was designed to give radiologists feedback with respect to the specific region on a breast with a lesion, but this aspect of the program was used only for training to avoid contention regarding whether or not a radiologist had clicked on a lesion correctly. Our results may be of interest in the context of measurement of mammography performance, since it has been shown that the definition of true positives affects the measurements of sensitivity and specificity [19]. There were several instances where diseased subjects were given an ACR BI-RADS 4 or 5 rating but the radiologist incorrectly located the lesion in the contralateral breast. This error was made in 22.5% of malignant cases by at least one radiologist; multiple radiologists made the error in some cases. In many studies, a true positive is defined as an ACR BI-RADS of 4 or 5 followed within a year by a diagnosis of cancer without regard for laterality. Some overestimation of sensitivity may occur in these studies.

Another limitation of the study is that participants were low volume radiologists who may have changed more in response to this exercise than would more experienced mammographers. In addition, although radiologists were informed that the prevalence of cancer in the study sets was higher than would be encountered in practice, radiologists' thresholds may have been lowered after the first session as they became aware of the high number of cases in the reading

sets. This could account for the increase in sensitivity and decrease in specificity through a threshold effect. Although the software supported immediate feedback following each individual case, feedback was withheld until the end of each session in the study. Other limitations of the CAMFP study include the possibility that mammographers may be less accurate when working test setting and viewing difficult cases than in the setting of their clinical practice [2,4]. If this bias was present in the CAMFP, it caused radiologists to be less accurate in their interpretations in all sessions and would not have resulted in an overestimation of the changes in sensitivity and specificity.

Improvement in mammography accuracy requires a combination of efforts. By requiring that mammograms are of extremely high quality, setting training and minimum volume requirements for radiologists, and enforcing audits of radiologists' performance, the MQSA provided a strong foundation for improving the quality of mammography. Advances in the field of radiology, such as digital-based operations, promise further improvement. However, concern over the variable performance of radiologists in low volume categories and the movement towards a culture of continuous quality improvement suggest that another opportunity for improvement may be in providing immediate feedback to radiologists during CME with a computer-aided training system [20]. Although the CAMFP yielded only modest improvements in sensitivity that were partially offset by decreases in specificity, our results demonstrate the feasibility of employing computer tools that allow mammographers to continuously focus on improving the accuracy of interpretation and the quality of their services.

The CAMFP provides an example of a software program that can provide a large number of training films together with immediate feedback regarding mammography interpretations. Although this study was conducted in 1999, the results are timely as it evaluates a teaching modality that may be appropriate for use today. The CAMFP format could contribute a valuable opportunity for radiologists to practice reading difficult films while meeting the CME requirement set forth by the MQSA [8]. Digital images are now of sufficiently high quality that use of digitized images should no longer be an impediment. However, the very modest improvement in accuracy seen in this study suggests that feedback alone is not sufficient to improve performance. The student may need to have the images explained. If the CAMFP were to be used for CME in the future it would be important to provide explanations of the findings, and to incorporate discussion in a group including instructors with access to full information regarding pathological findings. Alternatively, the results of this study could be interpreted as justifying a second, teleradiographic opinion from an experienced radiologist.

Acknowledgements

Grant Information: Department of Defense/CDMRP, DAMD17-02-1-0691 National Institutes of Health/National Cancer Institute, R01 CA63146 US Army Medical Research and Development Command, DAMD17-96-1-6288

REFERENCES

1. Pepe MS, et al. Design of a Study to Improve Accuracy in Reading Mammograms. *J Clin Epidemiol* 1997;50(12):1327–1338. [PubMed: 9449936]
2. Elmore J, Wells C, Howard D. Does diagnostic accuracy in mammography depend on radiologists' experience? *J Womens Health* 1998;7(4):443–449. [PubMed: 9611702]
3. Ciatto S, et al. Proficiency test for screening mammography: results for 117 volunteer Italian radiologists. *J Med Screen* 1999;6:149–151. [PubMed: 10572846]
4. Rutter CM, Taplin S. Assessing mammographers' accuracy: A comparison of clinical and test performance. *J Clin Epidemiol* 2000;53(5):443–450. [PubMed: 10812315]
5. Smith-Bindman R, et al. Physician Predictors of Mammographic Accuracy. *J Natl Cancer Inst* 2005;97(5):358–367. [PubMed: 15741572]

6. Beam CA, Conant EF, Sickles EA. Association of Volume and Volume-Independent Factors With Accuracy in Screening Mammogram Interpretation. *J Natl Cancer Inst* 2003;95(4):282–290. [PubMed: 12591984]
7. Mammography Quality Standards Act Public Law 101-359. 1992
8. Linver M, Newman J. MQSA: the final rule. *Radiol Technol* 1999;70(4):338–353. [PubMed: 10101751]
9. Berg WA, et al. Does Training in Breast Imaging and Reporting System (BI-RADS) Improve Biopsy Recommendations or Feature Analysis Agreement with Experienced Breast Imagers at Mammography? *Radiology* 2002;224(3):871–880. [PubMed: 12202727]
10. Nass, S.; Ball, J., editors. N.C.P.B. Committee on Improving Mammography Quality Standards, Institute of Medicine and National Research Council of the National Academies. National Academies Press; Washington, D.C.: 2005. Improving Breast Imaging Quality Standards; p. 24-71.
11. Thomson O'Brien MA, et al. Audit and feedback versus alternative strategies: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2000;2:CD000260. [PubMed: 10796521]
12. Robertson MK, Umble KE, R.M. C. Impact studies in continuing education for health professions: update. *J Contin Educ Health Prof* 2003;23(3):146–156. [PubMed: 14528785]
13. LoRusso AP, Bassignani MJ, harvey JA. Enhanced teaching of screening mammography using an electronic format. *Acad Radiol* 2006;13(6):782–788. [PubMed: 16679283]
14. Pisano ED, Yaffe MJ. Digital mammography. *Radiology* 2005;234(2):353–62. [PubMed: 15670993]
15. Pisano ED, McLelland R. Implementation of breast cancer screening. *Curr Opin Radiol* 1991;3(3):579–587. [PubMed: 1909543]
16. Box G. A general distribution theory for a class of likelihood criteria. *Biometrika* 1949;36(34):317–346. [PubMed: 15402070]
17. Ikeda D, et al. Analysis of 172 Subtle Findings on Prior Normal Mammograms in Women with Breast Cancer Detected at Follow-Up Screening. *Radiology* 2003;226(2):494–503. [PubMed: 12563145]
18. Pisano ED, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med* 2005;353(17):1773–1783. [PubMed: 16169887]
19. Rosenberg RD, et al. Effect of variations in operational definitions on performance estimates for screening mammography. *Acad Radiol* 2000;7(12):1058–68. [PubMed: 11131050]
20. Institute of Medicine. To Err is Human: Building a Safer Health System. National Academy Press; Washington, DC: 1999.

Figure 1.
Sample of mammogram image reviewed by the radiologists.

Figure 2. Change in sensitivity vs. change in specificity with stratification on reader group (filmset order). Readers from the two groups are distinguished by different symbol types, and joint confidence regions were calculated separately for the two strata.

Figure 3.

Comparison of sensitivity versus specificity by session. The smallest symbol size represents an individual reader; larger symbols denote multiple readers with the same sensitivity/specificity coordinate. The plus sign (+) represents the observed joint mean sensitivity and specificity for each session. The vertical and horizontal lines are provided for reference (at 80 and 90%).

Table 1

Composition of the Film Sets

	Test Sessions (1,4) [*] N (%)	Education Sessions (2,3,5) [*] N (%)
Outcome		
Malignant (Case)	40/90 (44.4%)	49/90 (54.4%)
Control	50/90 (55.6%)	41/90 (45.6%)
Lesion Type		
Mass	25/40 (62.5%)	31/49 (63.3%)
Calcification	14/40 (35.0%)	16/49 (32.7%)
Density ^{**}	0	0
Architectural Distortion	1/40 (2.5%)	2/49 (4.1%)
Patient Age (Mean, SD)	61 (13.3)	58 (11.3)

^{*} Each test session was composed of 45 films and each education session was composed of 30 films.

^{**} Although no cases in the study were densities, this category was included as an option for radiologists to use when describing an abnormality

Table 2
 Characteristics of Participating Radiologists as Reported in 1996

Characteristic	Group I (N=17) Mean (SD)	Group II (N=18) Mean (SD)
Years practicing radiology	15.5(8.3)	16.3(8.5)
Years reading mammograms	15.1(7.8)	13.6(6.3)
Number CME credits for mammography in past 3 years	26.6(18.4)	37.9(39.2)
Year obtained certification	1983(8.3)	1982(7.7)
Average number mammograms read in past year	135.9(93.5)	288.9(336.6)
Average number mammograms read per year	133.5(96.9)	248.3(297.3)

Table 3
Effects of CAMFP: Change (Δ) in Sensitivity and Specificity from Baseline to Follow-up by Reader Group[‡]

	Sensitivity			Specificity		
	Session 1 Mean (SD)	Session 4 Mean (SD)	Δ Mean (SD)	Session 1 Mean (SD)	Session 4 Mean (SD)	Δ Mean (SD)
Group I	0.75 (.112)	0.85 (.094)	0.103 (.139)	0.82 (.127)	0.76 (.073)	-.061 (.159)
Group II	0.77 (.148)	0.84 (.105)	.069 (.134)	0.81 (.117)	0.76 (.096)	-.053 (.166)
Combined	0.76 (.13)	0.84 (.099)	0.086* (.02)	0.81 (.12)	0.76 (.085)	-.057** (.028)

* $\chi^2 = 3.665, p < .001$
** $\chi^2 = -2.049, p = 0.04$
[‡] Group I read set A then set B
Group II read set B then set A