Screening Mammography: Clinical Image Quality and the Risk of Interval Breast Cancer

Stephen H. Taplin[1,2]
Carolyn M. Rutter[1]
Charles Finder[3]
Margaret T. Mandelson[1,4]
Florence Houn[5]
Emily White[4]

OBJECTIVE. We evaluated the association between clinical image quality and breast cancer occurrence within 24 months of a negative mammogram.

MATERIALS AND METHODS. We identified women with breast cancer who were younger than 40 years old and older and screened from January 1, 1988, through December 31, 1993. We retrospectively assigned Breast Imaging Reporting and Data System (BI-RADS) assessments to their screening mammogram. We classified cancers (invasive or ductal in situ) as "screen-detected" when found after positive assessments (BI-RADS codes 3, 4, and 5) and "interval-detected" when found after negative assessments (BI-RADS codes 1 and 2). One reviewer evaluated mediolateral oblique and craniocaudal views for all cancer cases using a 3-point scale (failure, borderline, pass) for each measure of clinical image quality (positioning, breast compression, contrast, exposure, noise, sharpness, artifacts, overall quality). We used separate logistic regression models and evaluated the odds of interval invasive cancer or invasive plus in situ cancer as a function of each measure of quality using "pass" as the referent group.

RESULTS. We found 492 screen-detected and 164 interval-detected cancers that met study criteria. Cancer detection (sensitivity) was highest (84%) among patients with proper breast positioning, but when images failed this measure (33.4%), sensitivity fell to 66.3%. After adjustment for age, film date, and breast density, interval-detected invasive cancers were more likely after images failing positioning (odds ratio, 2.57; 95% confidence interval, 1.28–5.52%). Failures in overall quality were also associated with interval cancers when cases of ductal carcinoma in situ ($p = 0.037$) were included.

CONCLUSION. Invasive breast cancer detection by mammography may be improved through attention to correct positioning.

A chieving mortality reductions through screening mammography requires high-quality examinations. Both technical and clinical aspects of the examination affect mammographic quality [1]. Technical quality assessment includes evaluation of imaging equipment with a standardized test object (phantom), evaluation of the processor to be sure it is appropriately set for the film used, and measurement of the radiation dose to the breast [2]. Clinical image quality assessment involves review of the films produced by a facility and consideration of positioning, breast compression, contrast, exposure, noise, sharpness, artifacts, and labeling [1, 3]. Our study evaluates the effect of failures in clinical image quality on the likelihood of cancer occurring within 24 months of negative findings on a screening mammogram.

The United States Congress passed the Mammography Quality Standards Act of 1992 to establish national standards [3]. The Mammography Quality Standards Act established a mammography certification program that includes evaluation of facility personnel, procedures, and technical image quality through annual on-site inspections and clinical quality review at least every 3 years through an accreditation body [2, 3]. This certification program is designed to ensure that all facilities in the United States achieve or exceed minimum quality standards. Results of analyses in Minnesota and Colorado suggest that the legislation has led to technical image quality improvements [4, 5].

Despite the intuitive appeal of high-quality clinical images, we are not aware of work that directly assesses the link between all of its dimensions and cancer detection at the time of screening (sensitivity). Some randomized trials of mammographic efficacy assured a minimum standard for mammography, and one trial had a systematic evaluation of clinical image quality

[6]. Another study showed that increased optical density is associated with a higher rate of detecting tumors less than 10 mm in diameter [7]. Although authors have described the parameters of clinical quality, little is known about which ones affect cancer detection [1]. In addition to quality, factors that may reduce the sensitivity of mammography include technical and interpretive errors [8–11], rapid tumor growth [11–16], age [17], and breast density [18–21]. Our study used a cohort design to evaluate the association between mammographic clinical image quality and the detection of cancer at the time of a screening visit.

Materials and Methods

Setting

Patients included women enrolled with Group Health Cooperative of Puget Sound, a consumer-controlled health maintenance organization with 400,000 members serving western Washington state. The study was reviewed and approved by the appropriate institutional and human subjects review boards.

All female members of Group Health Cooperative who are 40 years old or older are invited to enroll in a breast cancer screening program [22, 23]. Program enrollment begins by completing a risk factor questionnaire. Once enrolled, women receive regular reminders that they are due for screening. Screening occurs through breast cancer screening program centers in which women receive a mammogram and clinical breast examination [23, 24]. Eighty-five percent of women members complete the questionnaire and enroll in the program, so they are eligible for regular reminders [23]. Physicians may also order mammograms in the course of usual care for screening or to evaluate a symptomatic woman. All examinations occur through accredited and certified radiology facilities. Results of all screening examinations that occur in the breast cancer screening program are recorded in an electronic file on a mainframe database.

Eligibility and Inclusion

We identified women enrolled in the breast cancer screening program who were screened between January 1, 1988, and December 31, 1993. We designated the "index examination" as a woman's last screening mammogram. Among screened women, we identified those who met the following criteria: did not have a breast cancer history before the index examination, remained enrolled in Group Health Cooperative, or died of any cause during the succeeding 24 months after the examination and were diagnosed with ductal carcinoma in situ or invasive breast cancer (codes 174.0 through 174.9; International Classification of Diseases for Oncology [25]) within 24 months of their index examination and before their next screening mammogram. All cancers were identified and validated through our local Surveillance, Epidemiology, and End Results cancer registry.

Mammographic Assessments

The interpretations for all mammograms occurred before implementation of current American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) terminology, so we retrospectively assigned BI-RADS codes [26]. To make the assignment using a common standard, we gathered information from the women's paper medical records, their recorded mammogram assessments (negative, benign, indeterminate, positive), and the radiologists' recommendations (follow-up in 6 months, follow-up ≥1 year, surgical evaluation, or biopsy). We used this information to systematically classify the interpretations in accordance with American College of Radiology BI-RADS terminology [18, 27]. Any negative or benign screening assessment associated with a recommendation for a follow-up mammogram after 1 or more years was classified respectively as BI-RADS 1 (negative) or BI-RADS 2 (benign finding). Any "indeterminate" assessment associated with a recommendation for a follow-up mammogram after 1 or more years was classified as BI-RADS 2 (benign finding). Any "indeterminate" associated with a 6-month follow-up recommendation was classified as a BI-RADS 3 (probably benign). Any "indeterminate" or "positive" assessment associated with a recommendation for a surgical evaluation, biopsy, or both within 90 days was classified as BI-RADS 4 or 5 (suspicious or highly suggestive). The original mammography assessments were recorded after the evaluation of additional images if any were required.

Classification of Cases

We classified screen-detected cancers as those diagnosed among women within 24 months of a positive assessment of their screening mammogram. A positive assessment included American College of Radiology BI-RADS codes 3, 4, and 5.

We classified interval-detected cancers as those that occurred among women within 24 months of a negative assessment of their screening mammogram. A negative assessment included American College of Radiology BI-RADS codes 1 or 2. We also counted any assessment as negative if abnormalities noted by the radiologist were in the opposite breast from that in which the cancer was detected.

We audited the charts of all interval-detected cancers, all screen-detected cancers diagnosed after 3 months of their index mammogram, and 34 of the screen-detected cancers diagnosed within 90 days of their index mammogram to ensure that our classification of screen-detected and interval-detected cancer occurrence was accurate.

Sensitivity

We included all breast cancer cases (all screen-detected and interval-detected cases) in our study and calculated the sensitivity of screening mammography as the ratio of screen-detected cancers to the sum of all screen-detected and interval-detected cases. We calculated sensitivity for pass, borderline, and failed cases for each measure of quality.

Mammography Quality Review

A single radiologist reviewed all mammographic examinations to evaluate breast-tissue density and image quality. The quality reviewer was unaware of the age, detection status (screen vs interval cancer), or year of the mammogram for any women. The reviewer is the associate director for policy and clinical affairs of the Division of Mammography Quality and Radiation Programs of the Food and Drug Administration. This division administers the Mammography Quality Standards Act. He is a board-certified Mammography Quality Standards Act–qualified interpreting physician.

The reviewer rated the images for density and quality. He evaluated density on the basis of the four standard BI-RADS categories (almost entirely fat, scattered fibroglandular densities, heterogeneously dense, extremely dense). He evaluated the seven image quality attributes described in the Food and Drug Administration's regulations [3] (positioning, compression, exposure, noise, sharpness, contrast, and artifacts) using a grading scale developed specifically for this study based on previously published articles addressing image quality [1, 28, 29].

Each clinical quality attribute (positioning, compression, exposure, sharpness, noise, artifacts, and contrast), as well as an overall assessment, was rated on a 5-point ordinal scale based on the following: positioning, graded on the basis of the amount of pectoral muscle visualized, presence of retroglandular fat, no lateral glandular tissue cutoff film, craniocaudal nipple line within 1 cm of mediolateral oblique nipple line, and nipple in midline and profile on craniocaudal view; compression, graded on the basis of the separation of anatomic structures, uniform tissue exposure, no motion unsharpness, and good penetration of thicker areas without overpenetration of thin areas; exposure, graded on the basis of visualization of detail in dense glandular and fatty tissues, underlying tissue seen through pectoral muscle on mediolateral oblique view, and skin line visualized; sharpness, graded on the basis of the amount of blurring of edges of linear structures, tissue borders, and calcifications—no blurring, minimal blurring identified with a magnifying glass, moderate blurring identified with a magnifying glass, or blurring identified capable of obscuring or creating lesions; noise, graded on the basis of the amount seen—no noise, minimal noise identified with a magnifying glass, moderate noise identified with a magnifying glass, noise identified without a magnifying glass, or noise causing artifacts capable of obscuring or creating lesions; artifacts, graded on number seen—none, 1–5 seen but none simulate a lesion or significantly obscure anatomy, 1–5 artifacts seen and some simulate a lesion or significantly obscure anatomy, 5–10 artifacts seen and some simulate a lesion or significantly obscure anatomy, or greater than 10 artifacts seen and some simulate a lesion or significantly obscure anatomy; contrast, graded on good differentiation of fat, glandular, dense glandular, and calcific densities. After rating each of the clinical quality attributes of the mammograms, the radiologist provided a single subjective overall evaluation of their quality. The specific

grading scales have not been used by accreditation bodies although the quality attributes are part of routine review. Because few films were scored at the very upper and very lower ends of the scale, we collapsed the rating to three points: pass (clinical quality score = 1, 2), borderline (clinical quality score = 3), and fail (clinical quality scale = 4, 5). Quality was rated separately for two mediolateral oblique and two craniocaudal views, and the worst quality rating for either view was used in the analyses.

Analysis

We compared the clinical quality of mammograms among women with interval-detected cancers with that of screen-detected cancers using logistic regression. Quality was included as an independent categoric variable in the logistic regression models with passing quality as the referent category for each measure and interval detection versus screen detection as the dependent variable. Models were estimated separately for each of the seven attributes and for one overall quality measure. For each model, we present estimated odds ratios and 95% confidence intervals of interval detection for borderline versus passing quality and failing versus passing quality. We also include p values for the overall effect of the quality measure, based on the change in log-likelihood. Tables present both unadjusted and covariate-adjusted odds ratios. We estimated adjusted odds ratios from logistic regression models that simultaneously included the main effects of age, the year of the index examination, and the parenchymal density of the contralateral breast. Adjustments for age and density were made because they are known to have an effect on mammographic accuracy [18, 27], so we expected that they might modify the effect of quality on interval detection. We made an adjustment for the year of the mammogram to account for changes in technology over time. We did not adjust for body mass index because of its close correlation with density and because density is a stronger predictor of mammographic accuracy [18]. We used Spearman's rank correlation to explore bivariate relationships among the seven individual and one overall mammographic quality measures.

Our primary analyses focus on the probability of interval detection among women with invasive cancer using BI-RADS 1 or 2 as negative assessments and a follow-up period of 2 years. We explored the robustness of these results via three additional sets of analyses. These subanalyses used different subgroups of women and different definitions of screen and interval detection. The first set of subanalyses expanded the sample used for primary analyses to include women with ductal carcinoma in situ. The second set of subanalyses was restricted to invasive cases diagnosed within 13 months of their index mammogram. The third set of subanalyses used the same sample as the primary analyses but changed the definition of negative assessments to include BI-RADS 1, 2, or 3. This change meant that women given a BI-RADS 3 assessment were classified as interval-detected, whereas they were counted as screen-detected in the primary analysis.

TABLE 1	Proportion of Screen-Detected and Interval-Detected Invasive Breast Cancers and Specific Characteristics of the Woman, Tumor Size, and Year of Examination			
Characteristics	Screen-Detected[a]		Interval-Detected[b]	
	$n = 396$	%	$n = 152$	%
Age of woman				
40–49 yr	34	8.6	34	22.4
50–59 yr	85	21.5	38	25.0
60–69 yr	134	33.8	38	25.0
≥70 yr	143	36.1	42	27.6
Breast density[c]				
Low	300	75.8	73	48.0
High	95	24.0	78	51.3
Body mass index[d]				
Lean (<25)	189	48.6	86	58.5
Preobese (25–<30)	119	30.6	44	29.9
Obese (≥30)	81	20.8	17	11.6
Size of tumor				
<1 cm	110	27.8	17	11.2
1–<2 cm	174	43.9	62	40.8
2–<3 cm	55	13.9	33	21.7
3–<4 cm	17	4.3	9	5.9
4–<5 cm	7	1.8	9	5.9
≥5 cm	5	1.3	9	5.9
Unknown	28	7.1	13	8.6
Year of index examination				
1988	43	10.9	15	9.9
1989	41	10.4	12	7.9
1990	75	18.9	21	13.8
1991	55	13.9	28	18.4
1992	91	22.9	30	19.7
1993	91	22.9	46	30.3

[a]Cancer found within 24 months of a positive (Breast Imaging Reporting and Data System [BI-RADS] [26] codes 3, 4, and 5) assessment of the screening mammogram.

[b]Cancer found within 24 months of a negative (BI-RADS codes 1 and 2) assessment of the screening mammogram.

[c]Based on BI-RADS (almost entirely fat or scattered fibroglandular densities = low; heterogeneously dense or extremely dense = high). Missing for one interval-detected and one screen-detected case.

[d]Weight in kilograms / square of height in meters; missing for five interval-detected and seven screen-detected cases.

Results

Between January 1, 1988, and December 31, 1993, 676 women developed carcinoma in situ or invasive breast cancer subsequent to a screening program visit and met the enrollment criteria. We excluded eight woman for the following reasons: one requested exclusion from breast cancer studies; two women were diagnosed without a screening program visit, so their mammographic interpretation was not in the automated data; one woman was referred for surgical evaluation on the basis of clinical, rather than mammographic, findings; and four women had breast implants and were not eligible for usual mammographic screening. After excluding these eight women, 668 women remained. We excluded 12 women with lobular carcinoma in situ from this sample, leaving 656 women who contributed to final analyses. This sample included 548 women with invasive breast cancer and 108 women with ductal carcinoma in situ. Our first classification system resulted in 164 interval-detected cancers and 492 screen-detected cancer (12 interval- and 96 screen-detected cancers among women diagnosed with ductal carcinoma in situ and 152 interval- and 396 screen-detected cancers among women with invasive cancer). When we changed the definition of a positive mammographic assessment, 58 women given a BI-RADS 3 assessment moved from the screen-detected to interval-detected category so

that there was a total of 222 interval- and 434 screen-detected cases. Eighty-two percent of all the in situ and invasive cancers were found within 1 year, and the remaining 18% of cases were diagnosed from 13 to 24 months of the date of the screening mammogram.

Table 1 displays the characteristics of the women with screen-detected ($n = 396$) and interval-detected ($n = 152$) invasive cancers. As described elsewhere, the interval cancers were more likely to be among younger women, and by the time they were detected, the cancers were somewhat larger than those found at screening [27]. Half the interval cancers occurred in 1992 and 1993, reflecting the growth in total screening volume during that time period.

Table 2 shows the distribution of clinical quality measures for mammograms among women with screen-detected versus interval-detected invasive cancers. Three parameters show distributions that vary significantly among the screen-detected and interval-detected cancers: positioning, noise, and overall quality. Among the cases that failed to have good positioning, 35% failed only the mediolateral oblique view, 48% failed only the craniocaudal view, and 17% failed for both views. Among the cases that failed because of noise, 17% failed only the mediolateral oblique view, 4% failed only the craniocaudal view, and 78% failed both views. Among the cases that failed overall, 30% failed only the mediolateral oblique view, 14% failed only the craniocaudal view, and 56% failed both views.

Failures were rare (i.e., <3%) for compression, exposure, contrast, and artifacts, and these failures appeared to occur at similar rates for screen-detected and interval-detected cancers. Failures for sharpness were somewhat more common, and although interval-detected cases failed the subjective evaluation of image sharpness more often than screen-detected cases (10.5% vs 6.3%), this difference was not statistically significant ($p = 0.167$).

Overall quality was most strongly correlated with positioning (0.59), and only moderately correlated with other components of quality (ranging from 0.23 to 0.36).

Table 3 shows the proportion of cancers that were screen-detected (sensitivity) for each level of quality (pass, borderline, fail) within each of the measures. The proportion of interval-detected cancers is shown as 1-sensitivity. The model estimates the odds ratio and adjusted odds ratio of interval cancer for poor and borderline quality relative to passing quality. The model also estimates the relationship of the measure as a whole to interval cancer occurrence; p values for the

TABLE 2	Distribution of Mammographic Quality Scores Among Images for Women with Screen-Detected and Interval-Detected Invasive Breast Cancers			
Clinical Image Quality Measure	Screen-Detected[a]		Interval-Detected[b]	
	$n = 396$	%	$n = 152$	%
Position[c]				
Pass	65	16.4	12	7.9
Borderline	201	50.8	74	48.7
Fail	130	32.8	66	43.4
Compression				
Pass	300	75.8	110	72.4
Borderline	87	22.0	39	25.7
Fail	8	2.0	3	2.0
Exposure				
Pass	318	80.3	133	87.5
Borderline	70	17.7	19	12.5
Fail	8	2.0	0	0.0
Noise[d]				
Pass	295	74.5	99	65.1
Borderline	84	21.2	35	23.0
Fail	17	4.3	18	11.8
Sharpness				
Pass	258	65.2	89	58.6
Borderline	113	28.5	47	30.9
Fail	25	6.3	16	10.5
Contrast				
Pass	337	85.1	136	89.5
Borderline	56	14.1	15	9.9
Fail	3	0.8	1	0.7
Artifacts				
Pass	337	85.1	123	80.9
Borderline	50	12.6	25	16.5
Fail	9	2.3	4	2.6
Overall quality[c]				
Pass	37	9.3	9	5.9
Borderline	170	42.9	52	34.2
Fail	189	47.7	91	59.9

[a]Cancer found within 24 months of a positive assessment (Breast Imaging Reporting and Data System [BI-RADS] [26] codes 3, 4, and 5) of the screening mammogram.

[b]Cancer found within 24 months of a negative assessment (BI-RADS codes 1 and 2) codes of the screening mammogram.

[c]$p < 0.05$.

[d]$p < 0.005$.

measure reflect the evaluation of this association. Failing or borderline positioning was associated with the occurrence of interval invasive cancers. Sensitivity dropped from 84.4% among cases with passing positioning to 66.3% among cases with failed positioning. Borderline exposure was also associated with interval cancer occurrence but not in the expected direction. Borderline exposure appears to be associated with increased sensitivity in this analysis. Because the number of cases was small and no interval cancers were

detected among those cases that failed exposure, confirmation of the finding regarding exposure must occur in other settings. Noise and overall quality were also associated with interval cancer occurrence.

Table 3 shows that adjusting for patient age, film date, and breast density had little effect on associations between film quality and interval cancer occurrence within 2 years. After adjustment, failing or borderline positioning remained significantly associated with the occurrence of interval cancers by 24 months.

Clinical Image Quality Measure	Screen-Detected Proportion[b] (Sensitivity) (%)	Interval-Detected Proportion[c] (1-Sensitivity) (%)	Unadjusted		Adjusted	
			Odds of Interval Detection	95% CI	Odds of Interval Detection	95% CI
Position						
Pass	84.4	15.6	1		1	
Borderline	73.1	26.9	1.99	1.05–4.07	2.04	1.04–4.30
Fail	66.3	33.7	2.75	1.43–5.67	2.57	1.28–5.52
p			< 0.01[d]		<0.05[e]	
Compression						
Pass	73.2	26.8	1		1	
Borderline	69.1	30.9	1.22	0.78–1.88	0.74	0.45–1.20
Fail	72.7	27.3	1.02	0.22–3.61	0.43	0.06–1.90
Exposure						
Pass	70.5	29.5	1		1	
Borderline	78.7	21.3	0.58	0.33–0.98	0.51	0.26–0.93
Fail	100.0	0	NE		NE	
p			< 0.05[d]		<0.05[c]	
Noise						
Pass	74.9	25.1	1		1	
Borderline	70.6	29.4	1.24	0.78–1.95	0.83	0.49–1.36
Fail	48.6	51.4	3.16	1.56–6.41	1.93	0.88–4.19
p			< 0.01[d]			
Sharpness						
Pass	74.4	25.6	1		1	
Borderline	70.6	29.4	1.21	0.79–1.82	1.01	0.64–1.58
Fail	61.0	39.0	1.86	0.94–3.61	1.87	0.88–3.89
Contrast						
Pass	71.3	28.7	1		1	
Borderline	78.9	21.1	0.66	0.35–1.18	0.57	0.28–1.10
Fail	75.0	25.0	0.83	0.04–6.51	1.02	0.05–8.88
Artifact						
Pass	73.3	26.7	1		1	
Borderline	66.7	33.3	1.37	0.80–2.29	1.52	0.85–2.67
Fail	69.2	30.8	1.22	0.32–3.81	1.65	0.38–5.89
Overall quality						
Pass	80.4	19.6	1		1	
Borderline	76.6	23.4	1.26	0.59–2.93	1.10	0.49–2.70
Fail	67.5	55.9	1.98	0.95–4.53	1.74	0.80–4.15
p			<0.05[d]			

TABLE 3 Unadjusted and Adjusted Odds Ratios[a] for Interval-Detected Versus Screen-Detected Invasive Cancer

Note.—Compression is missing for one case. Each column represents results of separate logistic regression models: one set of unadjusted models that includes only ratings for a single measure of quality (failure, borderline, pass) and one set of models that includes each measure after adjusting for age, film date, and breast density. p values are included for statistically significant findings. CI = confidence interval, NE = parameter could not be estimated.

[a]Odds of cancer among women with borderline- or failed-quality films relative to women with films that passed the quality assessment.

[b]Cancer found within 24 months of a positive assessment (Breast Imaging Reporting and Data System [BI-RADS] [26] codes 3, 4, and 5) of the screening mammogram.

[c]Cancer found within 24 months of a negative assessment (BI-RADS [26] codes 1 and 2) of the screening mammogram.

[d]p value for overall effect of the quality measure based on the change in log-likelihood, unadjusted model.

[e]p value for overall effect in model, adjusted for age, film data, and density.

Borderline exposure also remained associated with interval cancer occurrence. Covariate adjustment reduced the effect of noise and overall quality so that the associations were no longer statistically significant, but the point estimates continued to suggest an effect that needs further evaluation in a larger study.

Not shown in Table 3 are results based on the three subanalyses using alternate samples and definitions of interval detection. The results for these subanalyses using the fully adjusted model differed slightly from the analyses presented in Table 3, although the direction of effects was the same as that in the primary analysis. In all three subanalyses, positioning remained associated with interval cancer occurrence. When women with ductal carcinoma in situ were included, overall quality was also associated with interval cancer occurrence ($p = 0.04$). When the sample was restricted to interval cancer within 13 months of the screening mammogram, exposure was no longer associated with interval cancer occurrence. When a negative assessment included BI-RADS codes 1, 2, and 3, exposure was no longer associated with interval cancer, although both noise and overall quality were also significantly associated with interval cancer ($p = 0.04$ and 0.02, respectively).

Discussion

To our knowledge, ours is the first study to assess the association between all clinical image quality parameters and interval cancer occurrence. We found an association between images that failed our positioning measure and interval cancer occurrence. The association persists under a variety of definitions of interval cancers, such as whether we include ductal carcinoma in situ in the cancers, whether we identify cancers within 1 or 2 years of the screening mammogram, and whether we include BI-RADS code 3 (probably benign) as a negative assessment. The association between positioning and interval cancer occurrence by 2 years also persists after adjustment for women's ages at the index mammogram, year of the index mammogram, and breast density.

Failures in overall quality appeared to be associated with interval cancers when we included ductal carcinoma in situ cancers and invasive cancers in the dependent variable. This association persisted after adjustment for body mass index and density. Overall quality is highly correlated with positioning quality, but consideration of the other elements such as sharpness and noise may also be important be-

cause failures in these measures also were more common among interval cancers (Table 2). Whether failures in overall quality are important depends on the clinical significance of ductal carcinoma in situ. Despite the controversies surrounding the natural history of ductal carcinoma in situ, belief is widespread that its detection and treatment are important [30, 31]. Attention to noise, sharpness, and positioning as important aspects of overall clinical image quality may, therefore, also improve screening detection.

Mammography that fails in sharpness, noise, or positioning is amenable to improvements. Sharpness could improve by reducing patient motion, reducing mammography unit vibration, using a smaller focal point, and using high-detail film cassettes [32, 33]. Optimizing background optical density, increasing radiation dose, or using high-detail film could decrease failures in noise [34]. Failures in positioning are particularly operator-dependent and can be improved by training technologists [35, 36]. Optimal positioning maximizes the amount of breast tissue seen on the image, uses a fully rotational C-arm, and takes into account the woman's habitus. Consistent with work by Bassett et al. [36], the measure of quality used in our study emphasized visualization of the pectoral muscle, the presence of retroglandular fat, and the nipple profile. However, compression is noted as a condition in Bassett's work, and we assessed that separately in our study.

Despite the robustness of our findings, some limitations in this study might be addressed in future studies of clinical image quality. Our series included interval cancers subsequent to films taken in 1988–1993 and required retrospective assignment of BI-RADS codes. Those codes were systematically assigned before the cancers were classified as screen-detected or interval-detected cancers, so any bias is expected to be minimal. However, a stronger design would use prospective interpretations. Mammographic quality has improved since 1993, and failures in sharpness and noise may be less common now. Therefore, those failures now could be associated with interval cancers. Increased attention has been paid to the importance of positioning [35]. Our study reinforces the importance of positioning, and although the frequency of failure may have changed, it seems unlikely that failures in positioning would have lost their association with interval cancer occurrence. Finally, we drafted measures of clinical quality on the basis of experience and knowledge of the literature, but they do not represent a gold standard. They are a

reasonable standard, and improvements in these measures might be considered in a future work. Furthermore, accreditation bodies assess these same quality attributes when evaluating films submitted for accreditation purposes, but they use different film selection criteria. Films submitted for accreditation are expected to represent the highest quality and are evaluated at that level. Therefore, the specific grading scale described in this article is not the one used by accreditation bodies.

Another limitation of the study deserves separate attention. We used a single rater to evaluate breast density and quality. That person was also the person who developed the quality scale. Although our method provided a consistent standard, it also may have introduced some bias because the application of the standard could have been influenced by the rater's commitment to its use and whether the rater could see the cancer on the films. However, at the time of the rating, the rater did not know whether the cancer was screen-detected or occurred in the subsequent interval. We know that about two thirds of cancers found in the interval subsequent to the screening could actually be seen at the time of the screening [27]. Therefore, we do not think that the rater's ability to see the cancer would introduce a large bias in the evaluation of the images when he did not know if they would be classified as screen-detected or interval-detected cancers. However, some persistent bias may exist if the rater was stricter in the application of the quality scale when a subtle cancer was identified. An alternative design would be to rate density and quality independently and use multiple reviewers with an average of their evaluations rather than the single rating. When others implement our quality measure and when multiple reviewers independently evaluate density and quality, the replication of our findings will be tested. Such a test is even more important given the association we found.

Although the associations between positioning and occurrence of invasive interval cancer appear robust, the lack of association with other measures does not mean those measures are not important. Failures of other quality measures, including noise and overall quality, were associated with interval cancer occurrence under some conditions. Failures in compression, contrast, and artifact were relatively rare. The lack of association between failures in these measures and interval cancer occurrence may simply reflect this rarity. Consideration of more sensitive scales for these measures may also be important. Young et al. [7] showed that achiev-

ing optimal optical density as measured by the automatic exposure control on the mammography unit is associated with the higher observer-rated image quality and smaller tumor detection. In our study, however, we had a single observer, whereas Young et al. had four, and we did not include a measure of optical density in our analysis. We also included all interval cancers with a full range of sizes. Further study of the influence of mammographic quality on interval cancer occurrence might consider the development of more sensitive measurement scales, using multiple reviewers, and the evaluation of the effects of quality on small-tumor detection.

In summary, we found that failures in positioning were associated with subsequent interval cancer occurrence under all conditions considered. Failures in the measures of sharpness were associated with interval cancer occurrence depending on the cancers included; follow-up periods; and adjustments for age, year of study, and breast density. Our results emphasize the importance of attention to positioning and indicate that accepting even borderline positioning that reduces the visualization of the pectoralis muscle or the nipple may increase the likelihood of missing an invasive breast cancer and reduce the sensitivity of mammography.

Acknowledgments

References

1. Bassett LW. Clinical image evaluation. *Radiol Clin North Am* **1995**;33:1027–1039
2. Hendrick RE, Botsco M, Plott CM. Quality control in mammography. *Radiol Clin North Am* **1995**; 33:1041–1057
3. Food and Drug Administration. *Quality mammography standards.* Final rule-21 CFR parts 16 and 900 [Docket no. 95N-0192]. RIN 0910-AA24 ed. Washington, DC: Department of Health and Human Services, Oct. 28, **1997**
4. Fintor L, Brown M, Fischer R, et al. The impact of mammography quality improvement legislation in Michigan: implication for the National Quality Standards Act. *Am J Public Health* **1998**;88:667–671
5. Hendrick RE, Chrvala CA, Plott CM, Cutter GR, Jessop NW, Wilcox-Buchalla P. Improvement in mammography quality control: 1987–1995. *Radiology* **1998**;207:663–668
6. Baines CJ, Miller AB, Kopans DB, et al. Canadian National Breast Screening Study: assessment of technical quality by external review. *AJR* **1990**; 155:743–747

7. Young KC, Wallis MG, Ramsdale ML. Mammographic film density and detection of small breast cancers. *Clin Radiol* **1994**;49:461–465

8. Martin JE, Moskowitz M, Milbrath JR. Breast cancer missed by mammography. *AJR* **1979**;132:737–739

9. Frisell J, Eklund G, Hellstrom L, Somell A. Analysis of interval breast carcinomas in a randomized screening trial in Stockholm. *Breast Cancer Res Treat* **1987**;9:219–225

10. van Dijck JA, Verbeek AL, Hendricks JH, Holland R. The current detectability of breast cancer in a mammographic screening program: a review of the previous mammograms of interval and screen detected cancers. *Cancer* **1993**;72:1933–1938

11. Ikeda DM, Andersson I, Wattsgard C, Janzon L, Linell F. Interval carcinomas in the Malmo Mammography Screening Trial: radiographic appearance and prognostic considerations. *AJR* **1992**;159:287–294

12. Fournier D, Weber E, Hoeffner W, Kubi F, Barth V. Growth rate of 147 mammary carcinomas. *Cancer* **1980**;45:2198–2207

13. DeGroote R, Rush B, Milazzo J, Warden M, Rocho J. Interval cancer: a more aggressive subset of breast neoplasias. *Surgery* **1983**;94:543–547

14. Heuser L, Spratt J, Kuhns J, Chang A, Polk H, Buchanan J. The association of pathologic and mammographic characteristics of primary breast cancers with "slow" and "fast" growth rates and axillary lymph node metastases. *Cancer* **1984**;53:96–98

15. Brekelmans CT, van Gorp JM, Peeters PH, Collette HJ. Histopathology and growth rate of interval breast carcinoma: characterization of different subgroups. *Cancer* **1996**;78:1220–1228

16. O'Malley AS, Mandelblatt J, Gold K, Cagney KA, Kerner J. Continuity of care and the use of breast and cervical cancer screening services in a multiethnic community. *Arch Intern Med* **1997**;157:1462–1470

17. Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Effect of age, breast density, and family history on the sensitivity of first screening mammography. *JAMA* **1996**;276:33–38

18. Mandelson MT, Oestreicher N, Porter PL, et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. *J Natl Cancer Inst* **2000**;92:1081–1087

19. Holland R, Hendriks JH Mravunac M. Mammographically occult breast cancer. *Cancer* **1983**;52:1810–1819

20. Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. *Radiology* **1992**;184:613–617

21. Ma L, Fishell E, Wright B, Hanna W, Allan S, Boyd NF. Case-control study of factors associated with failure to detect breast cancer by mammography. *J Natl Cancer Inst* **1992**;84:781–785

22. Carter AP, Thompson RS, Bourdeau RV, Andenes J, Mustin H, Straley H. A clinically effective breast cancer screening program can be cost-effective, too. *Prev Med* **1987**;16:19–34

23. Taplin SH, Mandelson MT, Anderman C, et al. Mammography diffusion and trends in late-stage breast cancer: evaluating outcomes in a population. *Cancer Epidemiol Biomarkers Prev* **1997**;6:625–631

24. Thompson RS, Taplin SH, Carter AP, Schnitzer F. Cost-effectiveness in program delivery. *Cancer* **1989**;64:2682–2689

25. National Cancer Institute. CID-O International Classification of Diseases for Oncology Field Trial Edition. Bethesda, MD: Working Party Coordinated by the International Agency for Research on Cancer, March, **1988**

26. American College of Radiology. *Breast imaging reporting and data system (BI-RADS),* 3rd ed. Reston, VA: American College of Radiology, **1998**

27. Porter PL, El-Bastawissi AY, Mandelson MT, et al. Breast tumor characteristics as predictors of mammographic detection: comparison of interval- and screen-detected cancers. *J Natl Cancer Inst* **1999**;91:2020–2028

28. Eklund GW, Cardenosa G, Parsons W. Assessing adequacy of mammographic image quality. *Radiology* **1994**;190:297–307

29. Kimme-Smith C, Bassett LW. Evaluation of clinical images. In: Vaughn VM, ed. *Workbook for quality mammography,* 2nd ed. Baltimore: Williams & Wilkins, **1997**:29–56

30. Fonseca R, Hartmann LC, Petersen IA, Donohue JH, Crotty TB, Gisvold JJ. Ductal carcinoma in situ of the breast. *Ann Intern Med* **1997**;127:1013–1022

31. Sakorafas GH, Tsiotou AG. Ductal carcinoma in situ (DCIS) of the breast: evolving perspectives. *Cancer Treat Rev* **2000**;26:103–125

32. Haus AG. State-of-the-art screen-film mammography: a technical overview. In: Barnes GT, Frey GD, eds. *Screen film mammography: imaging considerations and medical physics responsibilities.* Madison, WI: Medical Physics, **1991**:1–46

33. Webb S. *Medical science series: the physics of medical imaging.* Philadelphia: Institute of Physics and Institute of Physics Publishing, **1992**:48–51

34. Haus AG. Screen-film image receptors and film processing. In: Haus AG, Yaffe MJ, eds. *Syllabus: a categorical course in physics—technical aspects of breast imaging,* 3rd ed. Oak Brook, IL: Radiological Society of North America, **1994**:85–101

35. Eklund GW, Cardenosa G. The art of mammographic positioning. *Radiol Clin North Am* **1992**;30:21–53

36. Bassett LW, Hirbawi IA, DeBruhl N, et al. Mammographic positioning: evaluation from the view box. *Radiology* **1993**;188:803–806

The reader's attention is directed to the commentary on this article, which appears on the following pages.