J HK Coll Radiol. 2009;11:149-153

ORIGINAL ARTICLE **CME**

Stratifying Risk for Malignancy Using Microcalcification Descriptors from the Breast Imaging Reporting and Data System 4th Edition: Experience in a Single Centre in Hong Kong

KKK Chan,[1] CY Lui,[2] T Chu,[2] KK Chan,[3] AT Yan,[4] K Wong,[1] PY Fung,[2] LK Chan,[2] HS Lam[2]

[1]Department of Radiology, Tuen Mun Hospital, [2]Department of Radiology, Kwong Wah Hospital, Hong Kong, [3]Department of Systemic Therapy, RS McLaughlin Durham Regional Cancer Center, Oshawa, and [4]Department of Medicine, Division of Cardiology, St Michael's Hospital, Toronto, Ontario, Canada

ABSTRACT

Objective: *To evaluate whether categorisation of microcalcification descriptors based on the Breast Imaging Reporting and Data System 4th edition can stratify the risk for malignancy using pathology results of biopsies as the gold standard.*

Methods: *This retrospective study included 390 biopsies of 399 patients with microcalcifications treated at Kwong Wah Hospital, Hong Kong, from July 2005 to September 2007. Three subspecialty-trained breast radiologists (observers) interpreted copy films of mammograms, and classified them into 1 of the 5 Breast Imaging Reporting and Data System morphological calcification descriptor groups of benign, coarse heterogeneous, amorphous, fine pleomorphic, and fine linear. The number of malignancies in each group was compared and interobserver variability was assessed using the* κ *coefficient.*

Results: *Seventy seven of 390 biopsies (19.7%) demonstrated malignancy. The average percentages of malignancies were 7.8% for the benign group, 19.7% for the coarse heterogeneous group, 15.1% for the amorphous group, 53.0% for the fine pleomorphic group, and 86.0% for the fine linear group. The Breast Imaging Reporting and Data System 4th edition helped to stratify the percentage of malignancies in the high-probability group compared with the intermediate-concern and benign groups for all 3 observers (odds ratio, 15.0; 95% confidence interval, 7.49-29.60; odds ratio, 4.5; 95% confidence interval, 2.60-7.58; and odds ratio, 14.1; 95% confidence interval, 7.00-28.00, respectively). The percentage of malignancies was also significantly higher in the fine pleomorphic group when compared with the coarse heterogeneous group for 2 of the 3 observers. However, interobserver agreement was modest (*κ *coefficient = 0.24).*

Conclusions: *Microcalcification descriptors in the Breast Imaging Reporting and Data System 4th edition can stratify risk for malignancy of highly suspicious microcalcifications for the benign and intermediate-concern groups using local data. For 2 of the 3 observers, further differentiation of pleomorphic calcification into fine pleomorphic and coarse heterogeneous groups helped to stratify risk for malignancy. The quality of the copy films may have accounted for the high interobserver variability.*

Key Words: *Breast neoplasms; Calcification, physiologic; Mammography*

INTRODUCTION

The Breast Imaging Reporting and Data System (BI-RADS) was designed to provide an organised approach to image interpretation and reporting.[1] The descriptive terms and definitions used in the BI-RADS lexicon can help to stratify risk for malignancy by mammography.

For microcalcifications, descriptions are based on the morphology and distribution. The 5 BI-RADS morphological descriptor groups are: benign, coarse heterogeneous, amorphous, fine pleomorphic, and fine linear. In 2003, the BI-RADS 3rd edition[2] was revised, and the pleomorphic group of microcalcifications was divided into 2 classifications of coarse heterogeneous

Correspondence: *Dr KKK Chan, Department of Radiology, Tuen Mun Hospital, Tuen Mun, Hong Kong.*
Tel: (852) 2468 5177; Fax: (852) 2466 3569;
E-mail: karkichan@yahoo.com

Submitted: 11 November 2008; Accepted: 5 February 2009.

(>0.5 mm) and fine pleomorphic (<0.5 mm) for the 4th edition.[1] Morphological descriptors are categorised into benign, intermediate-concern (coarse heterogeneous and amorphous), and high probability of malignancy (fine pleomorphic and fine linear).[1,2]

The purpose of this study was to evaluate whether the categorisation of microcalcification morphologic descriptors based on the BI-RADS 4th edition[1] can stratify the risk for malignancy in a Hong Kong Chinese population. Interobserver variability was determined to assess whether the BI-RADS lexicon of microcalcification morphologic descriptors can be applied consistently. As a secondary endpoint, this study investigated whether the revision to the pleomorphic group of microcalcifications had a better stratification effect on risk for malignancy.

METHODS

Patients

This retrospective study included 399 consecutive patients who underwent stereotactic biopsy of mammographic microcalcifications at Kwong Wah Hospital, Hong Kong, from July 2005 to September 2007. Nine of the 399 patients were excluded due to insufficient specimen size (n = 1), inconspicuous microcalcifications in the mammograms (n = 5), and unavailability of the medical records (n = 3). The age of the patients ranged from 39 to 82 years (mean, 49.8 years).

Design

Three subspecialty-trained breast radiologists (observers) who were blinded to the pathology results independently interpreted the copy films of the diagnostic mammograms (including magnification views), which were obtained prior to biopsy, and the copy films of the biopsy specimen radiographs. The observers classified the microcalcifications of each mammogram into 1 of the 5 BI-RADS morphological descriptor groups of: benign, coarse heterogeneous, amorphous, fine pleomorphic, or fine linear.[1] Invasive carcinoma and ductal carcinoma in situ were considered malignant. High-risk lesions of atypical ductal hyperplasia and lobular carcinoma in situ were considered benign. The percentage of malignant tumours in each group of morphological descriptors for each observer was compared.

Statistical Analysis

Fishers exact test was used to assess whether there was any difference among the morphological descriptors for percentage of malignancies, particularly between the coarse heterogeneous and fine pleomorphic groups, for each observer. A p value of <0.05 was considered to be statistically significant. Odds ratios and 95% confidence intervals for the high probability of malignancy group (fine pleomorphic and fine linear groups) compared with the benign and intermediate-concern groups (benign, coarse heterogeneous, and amorphous groups) were calculated for each observer.

The interobserver variability was calculated using the κ coefficient. The data were analysed using the Statistical Package for the Social Sciences version 15.0 (SPSS Inc, Chicago, USA) and Stata version 10.0 (StataCorp, College Station, USA).

RESULTS

Of the 390 patients, 77 (19.7%) had malignant tumours and 313 (80.3%) had benign tumours. The number of malignant tumours in each group of microcalcification morphologic descriptors is shown in Table 1.

Fishers exact test showed a statistically significant difference between the morphological descriptors for all 3 observers ($p < 0.05$). The percentage of malignancies in the high-probability group compared with the intermediate-concern and benign groups were stratified for all 3 observers (Table 2).

The number of malignancies was significantly higher in the fine pleomorphic group than in the coarse heterogeneous group for 2 of the 3 observers (Table 3). However, the overall interobserver agreement was modest ($\kappa = 0.24$). Table 4 shows the sensitivity and specificity for each of the 3 observers.

DISCUSSION

Burnside et al[3] found that the microcalcification morphologic descriptors and categories described in BI-RADS 4th edition,[1] after reclassifying pleomorphic microcalcifications as coarse heterogeneous and fine pleomorphic microcalcifications, were helpful for predicting risk for malignancy of suspicious microcalcification. This study found that the microcalcifications morphologic descriptors in BI-RADS 4th edition[1] was helpful for stratifying the risk for malignancy of highly suspicious lesions against benign and intermediate-concern lesions. This is helpful for the management of patients with microcalcifications, especially when discussing the risk for malignancy and sharing the decision-making process with patients.

In the study by Burnside et al, the risk for malignancy of the microcalcification morphological descriptors in

Table 1. Number of malignant tumours in each category of the morphologic microcalcification descriptors of the Breast Imaging Reporting and Data System for 3 observers.

Observer	Benign Number* (%†)	Intermediate concern for malignancy Number* (%†)		High probability of malignancy Number* (%†)		p Value‡§
		Coarse heterogeneous	Amorphous	Fine pleomorphic	Fine linear	
1	12/111 (10.8)	8/45 (17.8)	24/186 (12.9)	28/43 (65.1)	5/5 (100)	<0.005
2‖	14/165 (8.5)	16/54 (29.6)	9/76 (11.8)	22/73 (30.1)	16/21 (76.2)	<0.005
3	5/126 (4.0)	6/51 (11.8)	34/166 (20.5)	23/36 (63.9)	9/11 (81.8)	<0.005
Average (%)	7.8	19.7	15.1	53.0	86.0	

* Malignant lesions/number classified.
† Percentage of malignant lesions.
‡ Fisher exact test.
§ Five-group comparisons.
‖ One patient had data missing.

increasing order were: benign, coarse heterogeneous, amorphous, fine pleomorphic, and fine linear.[3] In this study, coarse heterogeneous calcifications had a higher risk for malignancy than amorphous calcifications. This could partly be explained by the use of copy films in this study. Since the quality of the copy films may not have been as high as the original films, some benign calcifications might have been classified as amorphous (due to blurring on the copy films). Therefore, the number of microcalcifications classified as amorphous may have been larger than expected, resulting in a lower percentage of malignant tumours. Similarly, as the number of microcalcifications classified as benign was smaller, the percentage of malignancies in the benign group would be higher.

Burnside et al's study had a number of limitations.[3] Only 1 radiologist analysed the mammograms, so these researchers could not test the extent of interobserver variability, and only a small number of patients (n = 115) were included, limiting the generalisation of their results.

Table 2. Odds ratios and 95% confidence intervals for comparison of the risk for malignancy between the high-probability group and the benign and intermediate-concern groups.

Observer	Odds ratio	95% Confidence interval
1	15.0	7.49-29.60
2	4.5	2.60-7.58
3	14.1	7.00-28.00

Investigating interobserver variability can evaluate the practical implementation of the BI-RADS lexicon. Berg et al,[4] in a study to assess interobserver variability, concluded that microcalcification morphologic descriptors (according to the BI-RADS 2nd[5] and 3rd editions[2]) were one of the most difficult factors to attain consistency among different observers. The overall agreement was fair, with a κ value of 0.36. In this study, the κ value was 0.24, which is comparable with the study by Berg et al.[4] However, in a study by Baker et al[6] to investigate variability in radiologists' descriptions of breast lesion morphology (according to the BI-RADS 1st edition[7]), the interobserver κ value for microcalcification description was 0.5, indicating that substantial agreement was achieved among radiologists.

Variability in mammography interpretation can be attributed both to differences in detection of lesions and to variation in lesion characterisation and subsequent management. This topic received attention after the publication of the results of Elmore et al.[8] Kopans[9] and D'Orsi

Table 3. Comparison of stratification of risk for malignancy of the Breast Imaging Reporting and Data System 3rd edition[2] with the Breast Imaging Reporting and Data System 4th edition.[1]

Observer	4th edition			3rd edition
	Coarse heterogeneous microcalcifications Malignancy (%)	Fine pleomorphic microcalcifications Malignancy (%)	p Value*	Pleomorphic microcalcifications Malignancy (%)
1	17.8	65.1	<0.005	40
2	29.6	30.1	>0.05	29
3	11.8	63.9	<0.005	33

* Fisher exact test.

Table 4. Sensitivity and specificity for each of the 3 observers.

Observer	Sensitivity (%)	Specificity (%)
1	42.8	95.0
2	49.3	82.0
2	41.5	96.0

and Swets[10] suggested that much of this variation was attributable to variation in intervention threshold, with observers being plotted at different points along a similar receiver operating characteristic curve. High sensitivity (high true-positive rate) for some observers was achieved at the expense of low specificity (high false-positive rate).[11] This was also noted in the results of this study, as the sensitivity and specificity for observers 1 and 3 were similar (sensitivity, 42% to 43%; specificity, 95% to 96%), whereas the sensitivity for observer 2 was higher (49%), but the specificity was lower (82%) [Table 4]. Differences in threshold may affect implementation of the BI-RADS descriptors. However, differences in experience among the 3 observers did not appear to affect the results of this study.

The use of copy film in this study was also thought to account for the relatively high interobserver variability, and original film should be used for assessing interobserver variability. This may not be difficult for future studies, as many centres use digital mammography.

In this study, the average percentage of malignancy in the benign group was 7.8% (observer 1, 10.8%; observer 2, 8.5%; observer 3, 4.0%). This was partially attributed to the use of copy film, and the radiologists commented that it was difficult to differentiate punctate microcalcification from amorphous microcalcifications since the margin of microcalcifications in some of the copy films were blurred. Fifteen biopsies (3.8%) with malignant pathological results were regarded as benign or of intermediate-concern by all 3 observers. However, the samples were reviewed and no fine linear or fine pleomorphic microcalcifications were identified. These samples could partly explain the high percentage of malignancy in the benign group.

Although the percentage of malignancy in the benign group was high, it was considered that distribution and stability are also important descriptors. In the presence of suspicious morphology or suspicious distribution, biopsy is appropriate.[12] Therefore, even when the microcalcification morphology appeared benign, if the microcalcifications were in a cluster or if there was an interval increase in the number of microcalcifications, biopsy would be performed. Clustering of microcalcifications is a common finding in patients undergoing biopsy, so it is difficult to assess the ability of distribution descriptors for stratifying the risk for malignancy using biopsies without scattered microcalcifications, which are unusual at the Kwong Wah Hospital.

The microcalcification descriptors in the BI-RADS 3rd edition stratified the risk for malignancy into amorphous (26%), pleomorphic (41%), and fine linear (81%).[2] In this study, if both the coarse heterogeneous group and the fine pleomorphic group were classified as pleomorphic, the percentage of malignancies in the pleomorphic group would be 40.0%, 29.0%, and 33.0% for the 3 observers. Using the description in BI-RADS 4th edition, the percentage of malignancies for the coarse heterogeneous group were 17.8%, 29.6%, and 11.8%, and for the fine pleomorphic group, were 65.1%, 30.1%, and 63.9% (Table 3). Fishers exact test showed that the percentage of malignancies in the coarse heterogeneous and fine pleomorphic groups was significantly different for the first and third observers, but not for the second observer. Therefore, using BI-RADS 4th edition[1] did benefit the stratification of malignancy risk between the coarse heterogeneous and fine pleomorphic group for 2 of the 3 observers, and helped to stratify the intermediate-concern and benign groups from the high-probability group.

The results from this study justify the use of morphological descriptors of microcalcifications in BI-RADS 4th edition[1] for this Chinese population. This approach will be beneficial when communicating with patients about the likelihood of malignancy, and will help in discussion of the risks and benefits of biopsy. Improved estimation of the potential risk for malignancy will enable patients to share decisions, and provide better management of mammographic microcalcifications.

REFERENCES

1. American College of Radiology. Breast Imaging Reporting and Data System (BI-RADS). 4th ed. Reston: American College of Radiology, 2003.
2. American College of Radiology. Breast Imaging Reporting and Data System (BI-RADS). 3rd ed. Reston: American College of Radiology, 1998.
3. Burnside ES, Ochsner JE, Fowler KJ, et al. Use of microcalcification descriptors in BIRADS 4th edition to stratify risk of malignancy. Radiology. 2007;242:388-95.
4. Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment. AJR Am J Roentgenol. 2000;174:1769-77.
5. American College of Radiology. Breast Imaging Reporting and

Data System (BI-RADS). 2nd ed. Reston: American College of Radiology, 1995.
6. Baker JA, Kornguth PJ, Floyd CE Jr. Breast Imaging Reporting and Data System standardized mammography lexicon: observer variability in lesion description. AJR Am J Roentgenol. 1996;166: 773-8.
7. American College of Radiology. Breast Imaging Reporting and Data System (BI-RADS). Reston: American College of Radiology, 1993.
8. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. N Engl J Med. 1994;331:1493-9.
9. Kopans DB. The accuracy of mammographic interpretation. N Engl J Med. 1994;331:1521-2.
10. D'Orsi CJ, Swets JA. Variability in the interpretation of mammograms [letter]. N Engl J Med. 1995;332:1172.
11. Metz CE. Some practical issue of experimental design and data analysis in radiological ROC studies. Invest Radiol. 1989;24: 234-45.
12. Harvey JA, Fajado LL, Inni CA. Previous mammograms in patients with impalpable breast carcinoma: retrospective vs blind interpretation. AJR Am J Roentgenol. 1993;161:1167-72.