ELICITING A TERMINOLOGY FOR MAMMOGRAPHY

E. Alberdi[1], P. Taylor[2], R. Lee[2], J. Fox[3], A. Todd-Pokropek[4]

[1]Centre for Software Reliability, City University, Northampton Square, London, UK

[2]Centre for Health Informatics and Multiprofessional Education, University College London, UK

[3]Advanced Computation Laboratory, Imperial Cancer Research Fund, London, UK

[4]Department of Medical Physics, University College London, UK

Shortened title: Eliciting a terminology for mammography

Keywords: breast cancer screening, digital imaging, mammography, computerized decision support

ABSTRACT

AIM

Two studies were carried out to establish, validate and assess descriptors for use in the differential diagnosis of mammographic calcifications.

METHODS

In Study 1, eleven radiologists were asked to 'think out loud' as they interpreted 20 sets of calcifications. Participants used 159 terms to describe calcifications. We used this data to design a scheme with 50 descriptors. In Study 2, ten radiologists used the scheme to describe 40 sets of calcifications. We assessed the capacity of the terms to discriminate between benign and malignant calcifications, testing them against radiologists' assessments of malignancy and follow-up data.

RESULTS

All descriptors were used by at least 5 radiologists. Five additional descriptors were required. With some exceptions, properties that discriminated between benign and malignant outcomes were highly correlated with radiologists' assessment of risk. Many descriptors have a fairly low sensitivity but high specificity.

CONCLUSIONS

Our data suggest that radiologists consider a wider range of features than is included in existing reporting schemes. Our scheme allows a richer characterization of calcifications, potentially improving the reporting and understanding of these abnormalities.

INTRODUCTION

Microcalcifications are one of the most important signs of cancer on screening mammograms. However, although as many as 30% of screening mammograms reveal some form of calcification, 80 to 90% of it is benign [1]. The differential diagnosis of calcification is therefore one of the most important decisions a radiologist makes in assessing a screen mammogram. A number of companies now sell computer aids for mammographic interpretation [2,3]. These systems use image-processing algorithms to detect microcalcifications and masses in digitised mammograms. In this paper we consider an issue relating to the acceptability of such aids. We focus on their use in the detection of microcalcifications. Essentially this involves identifying small regions of high contrast. Most such regions will not be microcalcifications so a further step involves classifying them as either microcalcifications or as one of the various kinds of false positive – including obvious benign calcifications. This classification is often done by measuring a set of properties of the region and then feeding these into a program which has 'learnt' a classification rule having been 'trained' on examples on which radiologists have marked regions containing areas of potentially malignant calcification [see e.g. 4,5].

These computer aids are marketed as 'prompting' systems. The system identifies areas to which the user is encouraged to pay careful attention. The user is also told that the prompts will very often be false positives and that there may also be false negatives (unprompted abnormalities). There is, as yet, little evidence of the impact of such systems on readers' decision-making [6,7]. The success of the systems will depend crucially on the reader's reaction to false positive prompts. A prompt might distract the reader, allowing unprompted abnormalities to go undetected. A prompt could have an undue impact on the reader, diminishing their specificity. It might simply unsettle the reader, increasing the time taken to read a film. One factor that seems to be important in determining the reader's reaction to false positive prompts is the extent to which he or she feels able to explain the behaviour of the system in prompting a particular region [8].

We are developing a computerised aid that will make explicit the criteria used to identify calcifications as potentially malignant [9]. A critical step in the development of such a system is the identification of an appropriate set of criteria. In an earlier prototype we built up a set of statements that related characteristic properties to diagnoses. These statements were obtained from review articles, textbooks and monographs. The resulting knowledge base was considered by radiologists to be confusing, and in some respects erroneous [10]. Some of the statements, shorn of the context in which they were originally presented, were thought odd or perverse. Many concerned diagnostic categories that did not seem relevant to radiologists.

A different approach to the development of such a knowledge base would be to base it on a standard terminology, such as that of the American College of Radiology (ACR), the Breast Imaging Reporting and Data System (BI-RADS) [11]. However, BI-RADS has not been universally accepted and there is no clear evidence that its use improves consistency and accuracy in mammogram interpretation [12-17]. Furthermore we felt that a terminology designed for use in radiological reports might not provide a sufficient level of detail to explain the basis for the assessment of calcifications. We sought to establish a terminology that would capture the reasoning of radiologists making decisions in a, relatively, naturalistic setting.

In the rest of the paper we describe two studies that were carried out in order to inform the development of the new knowledge base. The first involved the analysis of statements made by radiologists asked to think out load while interpreting mammograms. This then led to the development of a draft database which was validated in the course of the second study.

STUDY ONE

The aim of the first study was to obtain a set of 'protocols', by which we mean the transcribed speech of someone thinking aloud while performing a task, from radiologists interpreting mammograms containing calcifications. These protocols were then analysed in order to identify those features that seemed most salient to the radiologists.

Method

Eleven consultant radiologists from 6 different hospitals and screening centres in England volunteered to take part. The radiologists had an average of 13 years of experience in reading mammograms with an average of 9,000 mammograms read per year. The participants were each presented with the 20 sets of mammograms, which were displayed one set at a time on conventional light boxes. All radiologists saw the same mammograms. The participants were asked to read each mammogram as they would in a normal clinical situation and to "think aloud", reporting everything that went through their mind. More specifically, they were instructed to: a) describe what they saw in each image, especially the calcifications, if present; b) provide a tentative diagnosis for each patient; c) suggest a course of action on the basis of what they saw in the image; and d) rate from 1 to 5 their confidence in that decision. They were also instructed to specify the location on the films of the findings they were reporting. Their verbal reports were recorded on audiotape.

Materials

Mammograms from 20 symptomatic patients were selected to provide a set with as many different types of calcification appearances as possible. The final set consisted of: four patients with no reported calcifications or abnormalities, eight with reported malignant appearances, and eight with reported benign appearances. Mediolateral-oblique and craniocaudal views of the left and right breasts were used with two exceptions: a) one patient for whom only three views were available (mediolateral-oblique of the right and left breasts and an extended craniocaudal view of the left breast); and b) one patient for whom nine contemporary films were available.

Results

The resulting think-aloud protocols were transcribed verbatim. Each radiologist's transcript was transformed into a table listing all the findings reported by the radiologist together with: the location, the terms used to describe the finding, any assessment of the level of suspicion associated with the finding. A total of 220 reports (one per radiologist per mammogram) were generated during the study. One report was lost. The data presented here are based on the remaining 219 reports.

A total of 170 findings were reported by at least one of the participants. Reported findings included opacities, distortions, asymmetries, and calcifications. Calcifications accounted for 53% (91) of these, 16% of the calcifications were reported by all radiologists, 25% were reported by only one radiologist, and 49% were reported by at least half of the participants. On average 51% of the radiologists reported each calcification, 89% of the radiologists who reported the same calcification agreed on its diagnostic assessment and 78% agreed on the tentative etiology of the calcification.

The participants used 159 different terms to describe the calcifications in the study. These terms fall into groups based on twelve properties: 1) shape of the flecks; 2) contour of the flecks; 3) size of the flecks; 4) density of the flecks; 5) distribution of the flecks; 6) number of flecks in a group; 7) variability among the flecks in a group; 8) cluster or area size; 9) location of the flecks; 10) orientation; 11) associated findings; and 12) differences in calcification appearances between the mediolateral and craniocaudal views. Six 'composite' descriptors referred to more than one property at the same time. For example, the term "ring" or "ring-like" makes reference to the shape (round), density (lucent centre) and contour (a rim). Other composite terms were coarse, fine, punctate, popcorn, tubular, and needle-like.

The most commonly used properties and property values for each of the main diagnostic categories of calcifications: benign, indeterminate and malignant are summarised in Table 1. It includes those properties and values that were noted by six or more participants. A term is included in each diagnostic category if a radiologist used the term to describe a calcification that she believed belonged to the category. Most of the properties (8/12) were noted by most of the radiologists. The only properties that were noted by fewer than half of the radiologists were location, orientation and differences in appearances between the mediolateral and craniocaudal views. Five of the properties (size, shape, density, distribution and associated findings) were noted for calcifications in all three diagnostic categories by most radiologists. Several of the specific values of those properties overlapped among categories although their relative frequency of use varied from category to category.

STUDY TWO

Removing all the synonyms, redundant and complex terms from the total set of descriptors obtained in Study 1, we were able to produce a descriptive scheme based on the twelve properties described above. This scheme contained a total of 50 descriptors. The aim of the second study was to validate the set of descriptors obtained from the first study and to obtain more data about their capacity to discriminate between benign and malignant interpretations.

Method

Ten consultant radiologists, with an average of 10 years of experience, viewing an average of 8,000 mammograms a year, were each shown 40 cases of calcifications. The radiologists were shown all the mammograms for a case, one set at a time. On each case an area of calcification was highlighted on a transparent overlay. With each pair of mammograms a sheet of paper was presented containing the descriptive scheme with the 50 terms obtained from Study 1. They were asked to tick all those descriptors that they felt applied to the highlighted calcification(s). If they thought that the descriptors on the sheet were not sufficient to characterize the calcification(s), they had the option to add any other terms that they may find appropriate. They were also asked to note the level of suspicion attributed to the highlighted calcifications (on a 5-point scale). All calcifications were from cases for which the diagnosis had been confirmed at biopsy or through follow-up. There were 29 cases of confirmed benign calcification and 11 cases of confirmed malignant calcification.

Materials

An experienced radiologist selected mammograms containing 40 cases of calcification. A mix of screening and symptomatic mammograms was used. Of these, 24 were considered *benign* by the assisting radiologist (14 different types of benign calcification were identified), 7 were considered *malignant* (5 DCIS and 2 associated with a mass) and the remaining 8 were considered *indeterminate* or difficult to assess. The idea was to show to the participants as many different calcification types & morphologies as possible. From the comments of the participants and the assisting radiologist, we can be fairly confident that the sample is comprehensive and representative.

Results

All the descriptors in the original scheme were ticked at least once by at least 50% (5) of the radiologists; 39 of the 50 descriptors were used by at least 80% (8) of the radiologists. The least frequently used descriptors were:

- streaming, very small cluster/area, (by a maximum of 5 radiologists)
- irregular, with a rim, few specks, in axilla, (by a maximum of 6 radiologists)
- adjacent, variable density, (by a maximum of 7 radiologists)

The radiologists suggested many other descriptors not included in the original scheme. However close to half (44%) of these extra descriptors were either minor variations to features already in our descriptive scheme (17%) or descriptors indicating the radiologists' inability to give a satisfactory response (27%; e.g. "can't distinguish", "difficult to see", "not applicable"). Examples of "minor variations" are: a) descriptors already in the scheme but under a different heading (5.5%; e.g. "pleomorphic" was used by some of the radiologists as a value for property "shape" instead of "between-fleck variability"); b) terms which were interpreted as compounds in Study 1 and subsequently split up into components that were included in the scheme (5.5%; e.g. "fine", "coarse", "punctate"); c) quantitative forms of qualitative descriptors (6%; e.g. "2cm cluster" instead of "small" cluster).

Most of the extra descriptors were only used by 1-3 of the radiologists in the study. Only 5 were suggested by more than 5 of the radiologists, namely: blood vessel (as associated finding), widespread (distribution), intermediate/medium/moderate (cluster size), vessel (location), in mass/density/opacity (location). The five new descriptors were added and 'irregular' was moved to appear under the heading 'shape' rather than 'contour'. The resulting terminology is summarised in Table 2.

The following analyses were conducted to ascertain which descriptors served to discriminate between diagnostic categories. We wanted to assess the discriminatory power of the terms against both the radiologists' assessments of risk and against the follow data for these cases. First we measured the correlation between the number of radiologists who applied a descriptor for a given case of calcification and number of radiologists who included the case in each of the risk categories. This is shown in Table 3 under the heading Correlation. The terms are listed with the most highly correlated terms appearing first. In this table we also give figures for the sensitivity and specificity of each descriptors as a test for benign or malignant disease. Comparing correlation with the sensitivity and specificity gives an indication of the extent to which the radiologists' assessments are reflected in the follow-up data. One anomalous result here is that the descriptor '1-5 flecks' (i.e. 1-5 calcifications) seems a much better indicator of benign disease than the radiologists' assessments would suggest. Similarly 'small' seems a less useful indicator of malignancy than the correlation data would suggest.

DISCUSSION

Based on an analysis of 219 transcripts of think-out loud protocols, we designed a descriptive scheme for the characterisation of calcifications. This scheme consisted of 50 descriptors organised under 12 headings. In a second study we attempted to assess the value of this descriptive scheme. Following this study we added five new descriptors and moved one descriptor to a new heading.

An interesting outcome of our work is that the radiologists used a far richer vocabulary for describing calcifications than existing reporting schemes. For example, BI-RADS contains 22 terms to characterize calcifications, whereas the radiologists in our study referred to at least 50 different property values. Certainly, not all of these descriptors were valuable for diagnostic purposes. But a number of descriptors that we found to be highly discriminating in our study are absent from the BI-RADS scheme. For example, this scheme does not include descriptors for properties such as "size" and "density" or property values such as well versus ill defined "contour", and variable density versus "homogeneous" particles.

Interestingly, there are also descriptors in the BI-RADS scheme that were never mentioned by the radiologists in either study; for example, "spherical", "milk of calcium", "suture", "dystrophic" and "regional" distribution. Arguably, the participating radiologists did not consider these properties to be relevant. But another possible explanation is that the data we used in our study did not contain calcifications with those characteristics. Indeed, a major concern in our data selection was how representative our samples were of the wide range of possible calcification appearances. During the first study, more than one of the participating radiologists noted that some specific appearances (e.g., benign "micro-cystic" and some indeterminate types) were missing from our set of mammograms. We decided to overcome this difficulty by including examples of "micro-cystic" and a wider range of "indeterminates" in the second study. As noted, we are quite confident that the appearances in Study 2 are fairly extensive (this was reinforced by the participants' comments). However we are aware that the generalizability of our results is an issue worth exploring as both the data and subject samples we used were fairly limited. We plan further studies with larger number of cases.

It is worth emphasizing that our conclusions are based on radiologists' interpretations of mammograms. Veldkamp used computer measurements of digitised mammograms to classify clusters of calcifications and found that the most useful features were 'relative distance to pectoral edge' and 'relative distance to breast edge' [18]. These features are not ones that the radiologists in our study mentioned.

Whatmough and colleagues followed a similar approach to ours in a study which looked at radiologists' agreement on the predictive value for malignancy of mammographic features; their results are consistent with ours [19]. Furthermore, our results are consistent to a great extent with studies that have looked at the predictive value of mammographic appearances by comparing radiological descriptors with biopsy diagnoses. Table 4 summarizes the results of five such studies featuring those calcification characteristics that were reported to be highly predictive of malignancy or benignity [16, 20-23]. Many features which we found to be of diagnostic value were not used in these studies. It is worth noting that each study tested a different set of radiological features and not all the authors explain the reasoning behind their selection.

Liberman and colleagues looked at the predictive value of BI-RADS features [16]; the rest of the studies used ad hoc descriptive sets that contained from 7 to 17 features. According to most of these studies, a clear predictor of malignancy seems to be "branching" shape. This was used by the radiologists in our study, but only half of the calcifications so described were found to be malignant on follow-up. Another feature that more than one study found to be a useful predictor of malignancy was "small" size. Again, this was frequently used in our study but did not differentiate clearly between benign and malignant calcifications. As regards benign appearances, the only descriptor which more than one study found to be relevant was "round" shape. This was one of the most frequently invoked descriptors for benign calcifications in our study; but was also often used for indeterminate and malignant appearances.

CONCLUSIONS

An empirical study of the descriptive terms that expert radiologists use when making decisions about calcifications has yielded a set of salient features with a potential value for discriminating between malignant and benign mammographic appearances. These features form the basis for the explanation of decision-making in CADMIUM II. An advantage of this approach is that the advice provided by the computer system is presented in terms that are relevant and informative to the user.

In addition, we believe our results have implications for the reporting and investigation of mammographic findings in general. We have seen that the sets of calcification descriptors in existing reporting schemes are fairly limited; this may partly explain inconsistent use of terminology by practitioners [24]. Our data suggest that radiologists consider a wider range of features. It could be argued that an efficient and usable reporting scheme should contain a richer vocabulary than those currently available to allow radiologists to express the subtleties involved in the characterization of calcifications.

The findings of studies which look at the correlation between mammography and diagnosis are still inconclusive with regards to calcifications [25]. Arguably, the mammographic features they consider may not be sufficient to establish differences between benign and malignant calcifications. We hope that our studies, which look at how practitioners describe mammographic appearances with a view to informing computer aids, will also contribute to establishing a sound set of mammographic descriptors for studying calcifications. Analysis of data on the sensitivity and specificity of the different descriptors suggests that no individual descriptors can serve as a sensitive test for benign or malignant interpretation. Instead there a range of descriptors that each quickly sensitive but which a sufficient specificity to allow a combination of measures to be used in discriminating between benign and malignant calcifications.

ACKNOWLEDGMENTS

This work is funded by the Engineering and Physical Sciences Research Council through the Health Informatics programme and by the Imperial Cancer Research Fund. The co-operation of Drs R. Given-

Wilson, J. Davies, A. Schneidau, J. Cooke, S. Rankin, I. Nockler and their colleagues is gratefully acknowledged.

REFERENCES

1. Given-Wilson R. Microcalcifications. in Morgan M., Warren R. and Querci della Rovere G. (eds.). *Early Breast Cancer* Harwood Academic 1999.
2. Castellino RA, Roehig J, Zhang W. Improved computer-aided detection (CAD) algorithms for screening mammography. *Radiology* 2000; **217** (Suppl.);400.
3. Allen B, Worrell S, Collins M, et al. SecondlookTM microcalcification detection and classification. In: Lemke HU, Vannier K, Inamura AG et al. editiors, Computer Assisted Radiology and Surgery, Excerta Medica, 2001: 544-552
4. Gavrielides MA, Lo JY, Vargas-Voracek R and Floyd CE. Segmentation of suspicious clustered microcalcifications in mammograms. Med Phys. 2000; 27(1): 13-22
5. Yu S, Guan L. A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films. IEEE-Trans-Med-Imaging. 2000; 19(2): 115-26
6. Warren-Burhenne LJ, Wood SA, D'Orsi CJ et al. Potential contribution of computer-aided detection to the sensitivity of screening mammography. Radiology. 2000; 215(2): 554-62
7. Brem RF, Schoonjans JM. Radiologist detection of microcalcifications with and without computer-aided detection: a comparative study. Clin Radiol. 2001; 56(2): 150-4
8. Hartswood M, Procter R, Willimas J. Prompting in practice: How can we ensure radiologists make best use of computer-aided detection systems in screening mammography? In: Karssemeijer E, Thijsen M, Hendriks J, editors. Digital Mammography. Dordrecht: Kluwer Publishing, 1998: 363-370.
9. Taylor P, Alberdi E, Lee R et al. Incorporating Image Processing in a Clinical Decision Support System. In Insana MF, Leahy RM, editors, *Information Processing in Medical Imaging, Springer Verlag,* 2001, 134-140.
10. Taylor P, Fox J, Todd-Pokropek A. The development and evaluation of CADMIUM: a prototype system to assist in the interpretation of mammograms. Medical Image Analysis 1999; 3: 321-337.
11. American College of Radiology. Breast imaging reporting and data system (BI-RADS). 2nd ed. reston, VA: ACR, 1995.
12. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. The New England J of Med 1994; 331(22): 1493-1499.
13. Heilbrun KS. The American College of Radiology's mammography lexicon: Barking up the wrong tree. AJR, 1994; 162:593-594.
14. D'Orsi CJ, Kopans DB. The American College of Radiology's mammography lexicon: Barking up the only tree (commentary). AJR, 1994; 162: 595.
15. Kerlikowske K, Grady D, Barclay J, Frankel SD, Ominsky SH, Sickles EA, and V. Ernster V. Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting Data System. Journal of the National Cancer Institute,1998; 90(23): 1801-1809.

16. Liberman L, Abramson AF, Lassman JR, Morris EA, Dershaw DD. The breast imaging reporting and data system: Positive predictive value of mammographic features and final assessment categories. AJR, 1997;171: 35-40.
17. Baker JA, Kornguth PJ, Floyd CE. Breast imaging reporting and data system standardized mammography lexicon: observer variability in lesion description. AJR 1996; 166: 773-8.
18. Veldkamp W.J.H. *Computer aided characterization of microcalcification clusters in mammograms*. PhD Thesis, University of Nijmegen. 2000
19. Whatmough P, Gale A, Wilson ARM. Do radiologists agree on the importance of mammographic features? In: Doi K, Giger ML, Nishikawa RM, Schmidt RA, editors. Digital Mammography'96. Amsterdam: Elsevier Science, 1996: 111-116.
20. Skinner MA, Swain M, Simmons R, McCarty KS, Sullivan DC, Iglehart JD. Nonpalpable breast lesions at biopsy. Ann Surg,1988; 208(2): 203-208.
21. Franceschi D, Crowe J, Zollinger R, Duchesneau R, Shenk R, Stefanek G. Biopsy of the breast for mamographically detected lesions. Surgery, Gynecology and Obstetrics, 1990; 171(6): 449-455.
22. Monostori Z, Herman PG, and Carmody DP,Eacobacci TM, Capece NR, Cruz VM, Fentin S, Vernace FM. Limitations in distinguishing malignant from benign lesions of the breast by systematic review of mamograms. Surgery, Gynecology and Obstetrics, 1991; 173: 438-442.
23. Harkins K, Tartter PI, Hermann G, Squitieri R, Bower ST, Keller RJ. Multivariate analysis of roentgenologic characteristics and risks factors for nonpalpable carcinoma of the breast. Journal of the American College of Surgeons,1994; 178: 149-154.
24. Simpson W, Neilson F, Kelly PJ. Descriptive Terms for mammographic abnormalities: Observer variation in application. Clin Radiol 1996; 51: 709-71.
25. Albertyn, LE. Mammographically indeterminate microcalcifications. Can we do any better? Australas Radiol, 1991; 35: 350-357.

Table 1
Most frequently used properties and values to describe calcifications

Benign Properties and Values	R's	% (N=527)	*Indeterminate* Properties and Values	R's	% (N=107)	*Malignant* Properties and Values	R's	% (N=242)
Size	**11**	**26% (139)**	**Size**	**10**	**28% (30)**	**Size**	**11**	**21. 5% (52)**
large	11	14% (72)	small	10	26% (28)	small	11	21% (51)
small	11	13% (67)	**Distribution**	**11**	**17% (20)**	**Shape**	**10**	**20.5% (50)**
Density	**11**	**23% (120)**	clustered	8	9% (10)	branching	8	11% (28)
high density	9	11% (59)	**Associated findings**	**10**	**18% (19)**	linear	6	7% (17)
low density	10	6% (31)	assoc. with density	8	11% (12)	**Variability**	**10**	**20% (49)**
lucent centre	11	5% (26)	**Density**	**9**	**18% (19)**	pleomorphic	10	12% (30)
Distribution	**11**	**20% (108)**	low density	9	16% (17)	**Associated findings**	**10**	**14% (35)**
scattered	11	9% (49)	**Variability**	**8**	**10% (11)**	malignant mass	10	9% (22)
vascular	11	6.5%(35)	pleomorphic	6	6% (7)	**Density**	**8**	**10% (24)**
Shape	**11**	**19% (100)**	**Shape**	**8**	**7% (8)**	low density	8	9.5% (23)
round	11	15% (77)				**Distribution**	**9**	**10% (24)**
linear	8	3% (14)				**Contour**	**6**	**4.5% (11)**
Number of Flecks	**11**	**10% (52)**				irregular	6	4% 10
1-5 specks	11	8% (44)						
Contour	**8**	**8% (41)**						
well defined	6	5% (28)						
Associated findings	**11**	**7% (36)**						
opacity	10	4% (22)						
malignant mass	6	2% (10)						

NOTE: Properties are shown in bold type and values are shown in regular type.
R's indicates the maximum number of radiologists who referred, at least once, to each property or property value.
N in each category (i.e., benign, indeterminate & malignant) indicates the total number of instances in which any term was used to describe a calcification that any of the radiologists considered belonged to that category.

Table 2

The revised descriptive scheme following the evaluation in Study 2.

SHAPE
branching
linear
streaming
oval
round
curvilinear
irregular

SIZE
big
medium
small

DENSITY
high density
low density
lucent centre

CONTOUR/EDGE
well defined
ill-defined

with a rim

NUMBER OF FLECKS
1 fleck
few flecks (=< 5)
several specks (>5, <X)
multiple specks (>X)

ASSOCIATED FINDINGS
benign opacity/mass/density
malignant opacity/mass/density
distortion
blood vessel
no associated finding

DISTRIBUTION
isolated
scattered
widespread
adjacent to each other
clustered
segmental
ductal/linear
vascular

ORIENTATION
towards nipple

CLUSTER/AREA SIZE
very small
small
medium
big
N/A (single)

BETWEEN-FLECK VARIABILITY
homogeneous
pleomorphic
variable size
variable density
variable contour
similar density
N/A (single)

LOCATION
in axilla
in skin
within fat
within glandular stroma
opacity/mass/density
vessel/artery

VIEW DIFFERENCES
cc:

l-o:

Table 3
The most discriminating terms from the terminology compared to both the radiologists' assessments of malignant disease and the followup data

Benign					Malignant				
N	Descriptor	Correl	Sens	Spec	N	Descriptor	Correl	Sens	Spec
22	well defined	0.73*	0.76	0.92	6	pleomorphic	0.87*	0.55	0.86
7	homogeneous	0.65*	0.24	0.88	6	variable density	0.77*	0.55	0.75
8	big	0.60*	0.28	1.00	2	branching	0.73*	0.18	0.50
8	within fat	0.59*	0.28	1.00	9	variable size	0.68*	0.82	0.60
4	similar density	0.49*	0.14	1.00	3	segmental	0.68*	0.27	1.00
3	curvilinear	0.44*	0.10	1.00	1	variable contour	0.59*	0.09	0.33
1	with a rim	0.42*	0.03	1.00	4	linear	0.58*	0.36	0.57
4	isolated	0.39	0.14	1.00	5	ill-defined	0.55*	0.45	0.63
4	lucent centre	0.39	0.14	1.00	1	ductal/linear	0.54*	0.09	0.50
12	1-5 flecks	0.38	0.41	1.00	7	small	0.54*	0.64	0.41
6	scattered	0.37	0.21	1.00	11	stroma	0.53*	1.00	0.41
2	vascular	0.27	0.07	1.00	6	clustered	0.53*	0.55	0.50
19	high density	0.26	0.66	0.76	1	malig. finding	0.52*	0.09	1.00
13	round	0.24	0.45	0.76	2	towards nipple	0.46*	0.18	0.66
20	no finding	0.24	0.69	0.80	5	low density	0.37	0.45	0.50
1	in skin	0.21	0.03	1.00	1	distortion	0.36	0.09	1.00
7	big (cluster)	0.18	0.24	0.70	3	several specks	0.28	0.27	0.60
3	medium	0.13	0.10	0.60	6	multiple specks	0.27	0.55	0.40
3	adjacent	0.08	0.10	1.00	6	small (cluster)	0.18	0.55	0.29
4	oval	0.07	0.14	1.00	1	few specks	0.14	0.09	1.00

The columns labelled Correlation above give Spearman's Rho for the correlation between radiologists' use of a term and their assessment of the diagnosis (benign or malignant) for the calcification so described. An asterisk is used to indicate a correlation that is significant at the 1% level. The columns labelled N indicate how many of the calcifications that most radiologists described using this descriptor were identified as Benign or Malignant on followup. The columns labelled Sens and Spec give the sensitivity and specificity of the descriptors, calculated from the followup data.

Table 4
Studies of the correspondence between mammographic appearances and diagnosis

	Liberman et al. (12)	Skinner et al. (16)	Franceschi et al. (17)	Monostori et al. (18)	Harkins et al. (19)
MALIGNANT					
branching	✓	✓	✓	✓	
pleomorphic					✓
number		✓	✓	✓	
segmental	✓				
small			✓		✓
scattered					✓
BENIGN					
round				✓	✓
coarse					✓
solid				✓	
packed					✓
curvilinear					
irregular				✓	