Measuring the accuracy of diagnostic imaging in symptomatic breast patients: team and individual performance

[1]P BRITTON, FRCS, FRCR, [2]J WARWICK, BSc, PhD, [3]M G WALLIS, ChB, FRCR, [1]S O'KEEFFE, FRCR, FFRRCSI, [1]K TAYLOR, DCR, MSc, [1]R SINNATAMBY, MRCP, FRCR, [1]S BARTER, MRCP, FRCR, [1]M GASKARTH, FRCS, FRCR, [2]S W DUFFY, MSc, CStat and [4]G C WISHART, MD, FRCS (Gen.)

[1]Department of Radiology, Cambridge Breast Unit, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK, [2]Cancer Research UK Centre for Epidemiology, Mathematics and Statistics, Wolfson Institute of Preventive Medicine, Barts and the London School of Medicine and Dentistry, London, UK, [3]National Institute for Health Research, Cambridge Biomedical Research Centre, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK, and [4]Department of Surgery, Cambridge Breast Unit, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

Objective: The combination of mammography and/or ultrasound remains the mainstay in current breast cancer diagnosis. The aims of this study were to evaluate the reliability of standard breast imaging and individual radiologist performance and to explore ways that this can be improved.
Methods: A total of 16 603 separate assessment episodes were undertaken on 13 958 patients referred to a specialist symptomatic breast clinic over a 6 year period. Each mammogram and ultrasound was reported prospectively using a five-point reporting scale and compared with final outcome.
Results: Mammographic sensitivity, specificity and receiver operating curve (ROC) area were 66.6%, 99.7% and 0.83, respectively. The sensitivity of mammography improved dramatically from 47.6 to 86.7% with increasing age. Overall ultrasound sensitivity, specificity and ROC area was 82.0%, 99.3% and 0.91, respectively. The sensitivity of ultrasound also improved dramatically with increasing age from 66.7 to 97.1%. Breast density also had a profound effect on imaging performance, with mammographic sensitivity falling from 90.1 to 45.9% and ultrasound sensitivity reducing from 95.2 to 72.0% with increasing breast density.
Conclusion: The sensitivity ranges widely between radiologists (53.1–74.1% for mammography and 67.1–87.0% for ultrasound). Reporting sensitivity was strongly correlated with radiologist experience. Those radiologists with less experience (and lower sensitivity) were relatively more likely to report a cancer as indeterminate/uncertain. To improve radiology reporting performance, the sensitivity of cancer reporting should be closely monitored; there should be regular feedback from needle biopsy results and discussion of reporting classification with colleagues.

Received 23 February 2010
Revised 19 May 2010
Accepted 26 May 2010

DOI: 10.1259/bjr/32906819

The standard diagnostic process for patients presenting with breast symptoms is ''triple'' assessment of expert clinical examination, imaging (mammography and/or ultrasound) and, where necessary, needle biopsy [1]. This study concentrated on the performance of the imaging aspects of this assessment with respect to breast cancer detection. As well as evaluating the overall sensitivity of mammography and ultrasound we will demonstrate how patient age and breast density affect imaging performance. Finally, we will examine the differences between the breast radiologists involved in this study with respect to imaging sensitivity and thereby identify factors that adversely affect performance and suggest ways in which this can be improved.

Address correspondence to: Dr Peter Britton, Department of Radiology, Cambridge Breast Unit, Box 97 Cambridge University Hospitals NHS Foundation Trust, Cambridge CB2 0QQ, UK. E-mail: peter.britton@addenbrookes.nhs.uk

Materials and methods

All new patients referred to a specialist breast clinic by their GP from 1 January 2001 to 31 December 2006 were identified using a dedicated database. The investigation and management of patients was carried out according to local department protocols, which have been adapted from national guidelines [2, 3]. Following careful clinical breast examination all patients over the age of 35 years underwent bilateral mammography. The vast majority of mammograms performed were analogue, with only an estimated 12% being digital during the study period. Any patient with a focal clinical or radiological abnormality had a targeted ultrasound examination. The imaging results were recorded on a five-point scale: 1, normal; 2, benign; 3, suspicious but probably benign; 4, suspicious but probably malignant; 5, malignant [4]. The mammograms were interpreted and reported with knowledge of the clinical examination findings. Ultrasound interpretation was performed with the knowledge of both the clinical and mammographic findings. Mammographic breast density was determined by the reporting radiologist using Wolfe's semiquantitative criteria [5]. Parenchymal patterns were therefore assigned to one of four categories, as follows: N1, lowest risk, parenchyma composed primarily of fat; P1, low risk, parenchyma chiefly fat but with prominent ducts occupying no more than one-quarter (in terms of volume) of the breast; P2, high risk, prominent ducts occupying at least one-quarter (in terms of volume) of the breast; DY, breast typically contains extensive regions of homogeneous mammographic densities, which appear as sheet-like regions.

Patients without suspicious clinical or imaging findings were discharged from the clinic. Any patient with a clinically suspicious or focal solid abnormality underwent an ultrasound-guided core biopsy (CB). All patients who underwent breast needle biopsy were subsequently discussed at a multidisciplinary meeting to decide future management. All data were entered onto a dedicated breast database, the Joint Clinical Information System (JCIS). JCIS is an *n*-tier web-based clinical information system supported by an SQL Server database. The system was built in-house in partnership with the breast clinical team using the i5 Web application and Microsoft technologies including Visual Basic. The system is designed to assist in the management and care of cancer patients, by providing tools for clinicians to enter coded data at the point of care and to reuse that information for clinical notes, letters, waiting time performance management, clinical audit and research.

Following clinical examination and imaging with or without biopsy, all patients were allocated to one of two groups: those with a breast cancer diagnosis confirmed on histopathology and those with no breast cancer diagnosis.

Statistical methods

The performance of mammography and ultrasound, by age and mammographic breast density, was assessed by calculating the sensitivity, specificity and area under the receiver operating curve (ROC). An overall score of M4 or M5 for mammographic findings or U4 or U5 for ultrasound findings was taken to mean that breast cancer had been detected on imaging. Sensitivity was defined as the percentage of women with breast cancer who were correctly identified as having malignancy (reporting Category 4 or 5), and specificity as the percentage of women without breast cancer who were correctly identified as not having malignancy (reporting Category 1, 2 or 3). Overall performance, and that of the five consultant radiologists who saw the most patients, was investigated. The joint sensitivity of two or more diagnostic procedures was defined as the proportion of women with breast cancer correctly identified as having malignancy (overall reporting Category 4 or 5) by any one of the diagnostic procedures. The joint specificity of two or more diagnostic procedures was defined as the proportion of women without breast cancer who were identified by each procedure as not having the disease (overall reporting Category 1, 2 or 3). The incremental effect on overall sensitivity of adding a further diagnostic procedure to the patient assessment process (in women who underwent clinical examination, mammography and ultrasound) was assessed by comparing the sensitivity and specificity of a sequence of appropriate combinations (*i.e.* clinical examination alone *vs* clinical examination and mammography *vs* clinical examination and mammography and ultrasound). Using breast cancer patients only, a logistic regression model with cancer detected (yes/no) as the dependent (y) variable and age and mammographic density as the explanatory (x) variables was used to estimate sensitivity after adjusting for age and mammographic breast density. The performance of the five main reporting consultant breast radiologists was also compared, for mammography and ultrasound separately. This study formed part of a registered audit at our institution, and did not require local ethics committee approval or patient consent.

Results

During the 6 years of the study 13 958 patients were referred for evaluation. 86% were referred only once during the study period, 10.6% twice and the remaining 3.4% were referred on 3–9 separate occasions. Consequently there were 16 603 separate assessment episodes, which form the basis of this analysis. Clinical examination was performed at 16 585 (99.9%) episodes, mammography at 11 003 (66.3%) episodes and ultrasound at 10 849 (65.3%) episodes. 2816 CBs were performed and results were recorded in JCIS in all but 124 (4.4%). A cancer diagnosis was made on CB (B5 malignant result) in 1108 and surgical histology in 127 patients, of whom 65 had no prior CB and 62 the prior CB result was not recorded. The total number of cancers diagnosed in the study was therefore 1235 (Table 1). The prevalence of breast cancer increases markedly with age. A woman aged 90+ years attending the symptomatic breast clinic is more likely than not to have breast cancer (prevalence 52.94%)

Based upon a total of 11 003 examinations, the overall mammographic sensitivity, specificity and ROC area were 66.6%, 99.7% and 0.83, respectively. The sensitivity of mammography improved dramatically with increasing age from 47.6% in patients aged 30–39 years to 86.7% in patients over 90 years (Table 2). Specificity fell with increasing age but changed to a lesser extent than sensitivity (99.9% in patients aged 30–39 years to 92.3% in patients over 90 years). Similar trends were observed with respect to ultrasound. Based upon 10 849 ultrasound examinations, overall sensitivity, specificity and ROC area were 82.0%, 99.3% and 0.91, respectively. The sensitivity of ultrasound improved dramatically with increasing age from 66.7% in patients aged 30–39 years to 97.1% in patients over 90 years. Specificity also fell with increasing age from 99.7 to 73.3% over the same patient age range. The effect of breast density on mammographic and ultrasound performance is shown in Table 3. Mammographic sensitivity fell from 90.1, to 45.9% with increasing breast density. Ultrasound performance is similarly, though less markedly, affected with a sensitivity reduction from 95.2 to 72.0% with increasing breast density. The incremental effect on sensitivity and specificity of clinical examination, mammography and ultrasound in the 10 999 patients undergoing all three investigations is shown in Table 4. Mammography increases sensitivity by 18.3% over clinical examination

Table 1. The number of patients attending the symptomatic breast clinic, by age and final diagnosis of either having breast cancer or not having breast cancer

Age group (years)	Number of patients	Number without cancer	Proportion without breast cancer (%)	Number with cancer	Prevalence of breast cancer (%)
<30	2289	2280	99.61	9	0.39
30–39	3693	3597	97.40	96	2.60
40–49	4810	4548	94.55	262	5.45
50–59	2872	2660	92.62	212	7.38
60–69	1424	1243	87.29	181	12.71
70–79	1016	752	74.02	264	25.98
80–89	431	256	59.40	175	40.60
90+	68	32	47.06	36	52.94
Overall	16603	15368	92.56	1235	7.44

with minimal (0.1%) loss of specificity. The addition of ultrasound further raises sensitivity by 9.8%, with only a 0.3% reduction in specificity. The results of the logistic regression analyses with cancer detection as the dependent (y) variable and age and mammographic breast density as the explanatory (x) variables are shown in Table 5. Both age and mammographic density are highly significant in the model for mammography alone ($p<0.001$ for both factors), with increasing age leading to an increased chance of a breast cancer being detected with clinical breast examination and/or imaging and increasing density leading to a reduced chance of cancer detection. The effect of ageing was similar for the other two models (ultrasound alone and clinical assessment plus mammogram and ultrasound) with odds ratios (ORs) of 1.04 and 1.05, respectively. For a parenchymal pattern, however, the effect was still significant ($p=0.02$ and $p=0.05$, respectively), but the reduction in chance of a breast cancer being detected with increasing density was less obvious. The major reduction in chance of breast cancer being detected, for all three models, occurs between the N1 and P1 categories. Since both age and parenchymal pattern are significant in these models, this indicates that the two factors have independent effects, *i.e.* the improved detection rates in older women are not fully explained by the fact that they tend to have less mammographic density. The OR for age in the mammography alone model suggests that the relative risk of a woman with breast cancer having it detected by mammography alone increases by 3% each year of age. The beneficial effect of age on chance of having a breast cancer detected is fairly similar for all three modalities, but greater for ultrasound than for mammography, and greatest for clinical examination, mammography and ultrasound combined. The detrimental effect of having a dense parenchymal pattern is considerable for all modalities but greatest for mammography alone.

Individual radiologist performance of mammographic and ultrasound reporting and experience are shown in Tables 6 and 7 and Figures 1 and 2. The sensitivity ranges widely between the radiologists (53.1–74.1% for mammography and 67.1–87.0% for ultrasound). Experience in breast radiology is shown in Table 6, expressed as age during the study period and years as a consultant breast radiologist. All the radiologists in this study routinely performed NHS Breast Screening Programme as well as symptomatic work, and the average number of screening and symptomatic mammograms reported

Table 2. Mammographic and ultrasound sensitivity, specificity, ROC area and prevalence of breast cancer in patients attending the symptomatic breast clinic by age

Diagnostic procedure	Age group (years)	Number of women	Prevalence of breast cancer (%)	Sensitivity (%)	Specificity (%)	ROC area
Mammography	<30	15[a]	40.0	83.3	88.9	0.86
	30–39	1909	4.4	47.6	99.9	0.74
	40–49	4238	5.7	51.2	99.8	0.76
	50–59	2301	8.6	55.6	99.9	0.78
	60–69	1170	14.0	68.9	99.8	0.84
	70–79	925	27.0	79.8	99.3	0.90
	80–89	389	43.0	86.1	97.3	0.92
	90+	56	54.0	86.7	92.3	0.90
	Overall	11003	10.0	66.6	99.7	0.83
Ultrasound	<30	1500	0.6	44.4	99.9	0.72
	30–39	2504	3.7	66.7	99.7	0.83
	40–49	3277	7.8	73.2	99.5	0.86
	50–59	1860	11.0	74.9	99.5	0.87
	60–69	782	22.0	86.7	98.7	0.93
	70–79	594	43.0	92.5	96.8	0.95
	80–89	282	59.0	90.3	92.3	0.91
	90+	50	70.0	97.1	73.3	0.75
	Overall	10849	11.0	82.0	99.3	0.91

ROC, receiver operating curve.
[a]Data should be interpreted with caution as extremely small numbers of patients.

Table 3. Mammogram and ultrasound sensitivity, specificity, receiver operating ROC area and prevalence of breast cancer in women attending symptomatic breast clinics by breast parenchymal pattern

Diagnostic procedure	Parenchymal pattern	Number of patients	Prevalence of breast cancer (%)	Sensitivity (%)	Specificity (%)	ROC area
Mammography	N1	1310	13	90.1	99.1	0.95
	P1	2532	13	72.5	99.6	0.86
	P2	4556	10	60.6	99.9	0.80
	DY	2125	6.3	45.9	99.7	0.73
Ultrasound	N1	615	27	95.2	97.3	0.96
	P1	1452	22	86.0	98.7	0.92
	P2	3077	15	81.7	99.6	0.91
	DY	1683	7.8	72.0	99.2	0.86

DY, breast typically contains extensive regions of homogeneous mammographic densities, which appear as sheet-like regions; N1, lowest risk, parenchyma composed primarily of fat; P1, low risk, parenchyma chiefly fat but with prominent ducts occupying no more than one-quarter (in terms of volume) of the breast; P2, high risk, prominent ducts occupying at least one-quarter (in terms of volume) of the breast; ROC, receiver operating curve.

Table 4. The incremental effect on sensitivity and specificity of clinical examination, mammography and ultrasound in the 10 999 patients undergoing all 3 investigations

Diagnostic procedure	Prevalence of breast cancer (%)	Sensitivity (%)	Specificity (%)	ROC area
Clinical examination	10	59.6	98.1	0.79
Clinical examination plus mammography	10	77.9	98.0	0.88
Clinical examination plus mammography and ultrasound	10	87.7	97.7	0.93
Additional effect of mammography		+18.3	−0.1	+0.09
Additional effect of ultrasound		+9.8	−0.3	+0.05

ROC, receiver operating curve.

Table 5. Logistic regression analysis to investigate how age and parenchymal pattern affect the chances of a woman's breast cancer being detected when assessed using mammography alone, ultrasound alone and clinical examination, mammography and ultrasound in combination

				Multivariate (all factors together)		
Assessments	Factor	Level	Number (detected/total)	OR	95% CI	*p*-value
Mammography alone	Age	Mean (range)	62 (23–98)	1.03	1.02–1.04	<0.001
	Parenchymal pattern	N1	154/171	1.00	–	
		P1	235/325	0.35	0.20–0.61	
		P2	287/474	0.27	0.16–0.48	<0.001 (linear trend)
		DY	61/133	0.17	0.09–0.33	
Ultrasound alone	Age	Mean (range)	62 (23–98)	1.04	1.03–1.05	<0.001
	Parenchymal pattern	N1	158/166	1.00	–	
		P1	271/315	0.40	0.18–0.89	
		P2	374/458	0.43	0.20–0.95	0.02 (linear trend)
		DY	95/132	0.30	0.13–0.70	
Clinical, mammogram and ultrasound	Age	Mean (range)	62 (23–98)	1.05	1.03–1.06	<0.001
	Parenchymal pattern	N1	166/171	1.00	–	
		P1	293/324	0.40	0.15–1.07	
		P2	402/474	0.38	0.15–1.00	0.05 (linear trend)
		DY	104/133	0.31	0.11–0.87	

CI, confidence interval; OR, odds ratio.
Parenchymal patterns N1, P1, P2 and DY according to Wolfe [5].

Table 6. Table showing mammography reporting performance and experience for five consultant breast radiologists in order of sensitivity

				Mammogram reports in cancer patients								Radiologist experience		
Radiologist	Number of mammogram reports	Number of cancers detected	Prevalence of breast cancers (%)	M1, *n* (%)	M2, *n* (%)	M3, *n* (%)	M4, *n* (%)	M5, *n* (%)	Sensitivity (%)	Specificity (%)	ROC area	Age during study period (years)	Years of breast radiology experience	Average number of mammogram reports per annum (symptomatic and screening)
1	2309	270	12.0	30 (11)	7 (3)	32 (12)	65 (24)	136 (50)	74.1	99.8	0.87	58–64	20	5329
2	1091	104	9.5	13 (12)	5 (5)	10 (10)	48 (47)	27 (26)	72.1	99.9	0.86	55–61	20	4010
3	2564	276	11	35 (13)	8 (3)	45 (16)	62 (22)	127 (46)	68.5	99.7	0.84	45–51	16	4949
4	1150	84	7.3	13 (16)	2 (2)	17 (20)	24 (29)	27 (33)	60.7	99.8	0.80	35–37	1	7508
5	2041	209	10.0	32 (15)	2 (1)	64 (31)	62 (30)	49 (23)	53.1	99.7	0.76	35–41	7	4957

M, mammography; ROC, receiver operating curve.

Table 7. Table showing ultrasound reporting performance for five consultant breast radiologists in order of sensitivity

					Ultrasound reports in cancer patients							
Radiologist	Number of ultrasound examinations	Number of years experience	Number of cancers	Prevalence of breast cancer (%)	U1, *n* (%)	U2, *n* (%)	U3, *n* (%)	U4, *n* (%)	U5, *n* (%)	Sensitivity (%)	Specificity (%)	ROC area
3	2110	16	269	13.0	7 (3)	11 (4)	17 (6)	81 (30)	153 (57)	87.0	99.3	0.93
1	2031	20	268	13.0	7 (3)	11 (4)	17 (6)	48 (18)	185 (69)	86.9	99.2	0.93
2	1048	20	99	9.4	1 (1)	8 (8)	6 (6)	44 (45)	40 (40)	84.8	99.4	0.92
5	1819	7	198	11.0	5 (3)	13 (7)	31 (16)	79 (40)	70 (35)	75.3	99.3	0.87
4	1145	1	76	6.6	8 (11)	5 (6)	12 (16)	21 (28)	30 (39)	67.1	99.5	0.83

ROC, receiver operating curve; U, ultrasound.

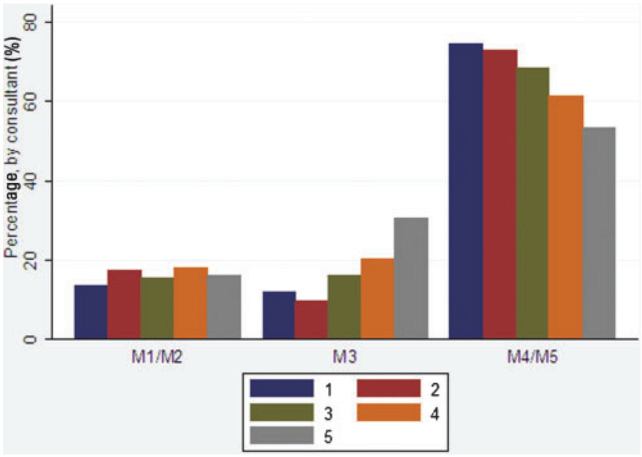

Figure 1. Distribution of overall mammographic assessment categories of patients with breast cancer, by five consultant breast radiologists, combining M1 and M2, and M4 and M5. M, mammograph.

per annum during the study period is also shown in Table 6. Those radiologists with greater experience had higher sensitivity than those with less experience. The breakdown of the five radiologists' M1–M5 and U1–U5 classification in patients with cancer is also shown. Specificity was uniformly high for both mammography and ultrasound for all radiologists.

Discussion

We have shown that for woman referred to a multidisciplinary, symptomatic breast clinic overall mammographic and ultrasound sensitivity is 66.6% and 82% and specificity is 99.7% and 99.3%, respectively. We have used the immediate assessment endpoint of cancer or no cancer when calculating sensitivity and specificity. We have not taken into account those patients with false-negative assessment, *i.e.* patients who were reassured only to be subsequently diagnosed with cancer. Our previous work established that the false-negative rate in such a clinic evaluating over 7000 patients for 3 years following discharge with a non-cancer diagnosis was very low [6]. The overall diagnostic accuracy of ''triple'' assessment was 99.6% and the ''missed'' cancer rate was 1.7 per 1000 women discharged. As the false-negative rate is so low it would not materially affect the sensitivity and specificity calculations at initial assessment. Comparison with published literature sensitivity and specificity is complicated by the wide range of different study populations and reporting protocols. Reported sensitivities for mammography range between 57 and 97% and for ultrasound between 49 and 100%. Specificity for mammography ranges between 36 and 97% and for ultrasound between 29 and 100% [7]. The ultrasound examination in this study was performed with prior knowledge of mammographic findings. In a review of 22 studies Flobbe et al [7] found that ultrasound performed better than mammography in 4 of 10 studies in which prior knowledge of mammography results were available and in 8 of 11 studies in which ultrasound was interpreted independently. Their results indicate a higher relative diagnostic performance when no prior knowledge of mammography was available; however, this has not been corroborated by other investigators [8]. As well as calculating individual test performance we have shown that the addition of mammography increased sensitivity by 18.3% over clinical examination and that the addition of ultrasound further raises sensitivity by 9.8%. Previous studies have shown a similar rise in sensitivity by combining mammography and ultrasound, which ranges from 4 to 22% [9]; however, most of the published data show a concomitant fall in specificity which was only minimal in this series (0.3%). Houssami et al [10] found very similar results with an overall increased sensitivity of 11.6% with a small drop in specificity of 2.2%.

We have shown that mammography performance is directly and powerfully affected by patient age, with a substantial reduction in sensitivity in younger woman. Similar findings were obtained in a study of 3799 breast cancers in which mammographic and ultrasound sensitivity rose from 56.9% and 70.7%, respectively, in woman less than 40 years old up to 91.3% and 88%, respectively, in woman over the age of 75 years [11].

We have shown that age and parenchymal pattern have independent effects on the sensitivity of mammography, ultrasound and the clinical examination/mammography/ultrasound combination. The effect of increasing age on sensitivity is fairly consistent across the three techniques/combinations investigated, with the ORs for the chance of a breast cancer being correctly designated (reporting category 4 or 5) increasing by 3–5% per year of age. The mammographic density was classified according Wolfe's four categories, which have been linked to breast cancer risk [5]. These have similar principles to the four categories of the Breast Imaging Reporting and Data System lexicon of mammographic density (1, almost entirely fat; 2, scattered fibroglandular densities; 3, heterogeneously dense; 4, extremely dense) [12]. The effect of increasing density is also consistent in that it is associated with reduced sensitivity in all the techniques/combinations considered, but the magnitude of the effect is greater for mammography (ORs of 0.35, 0.27 and 0.17 for the P1, P2 and DY categories relative to the N1 category, respec-

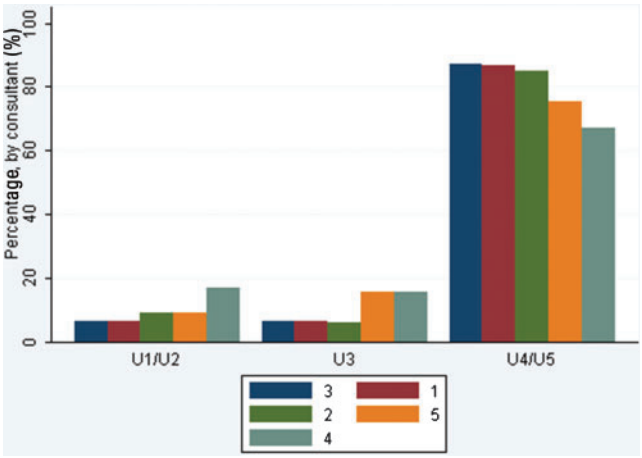

Figure 2. Distribution of overall ultrasound assessment categories of patients with breast cancer, by five consultant breast radiologists, combining U1 and U2, and U4 and U5. U, ultrasound.

tively), whereas for ultrasound the corresponding figures are 0.40, 0.43 and 0.30, respectively, and for clinical examination/mammogram/ultrasound are 0.40, 0.38 and 0.31, respectively. The inherent lack of contrast between breast cancers and dense breast parenchyma makes it readily understandable why breast density should adversely affect mammographic sensitivity. Although a similar effect on ultrasound sensitivity has been previously reported, it has not been convincingly explained. Dense parenchymal patterns are frequently associated with increased shadowing artefacts and coupled with greater benign changes, such as cyst formation, may account for the cumulative reduction in sensitivity. For all three techniques/combinations, therefore, age has an effect on sensitivity that is statistically significant and is not explained by mammographic breast density.

With regard to individual radiologist performance, the data for mammography (Table 6, Figure 1) and ultrasound (Table 7, Figure 2) are markedly similar. The specificity for all radiologists for both mammography and ultrasound is extremely high. This will in part be due to the predominance of non-cancer patients (92.6%) in the symptomatic population examined. On the other hand, there is a marked difference in sensitivities across the five radiologists. Sensitivity is therefore a more useful measure of performance, *i.e.* how many of the breast cancer patients does the examining radiologist assign an M4/M5 or U4/U5 result? The data show that the rate of reporting a mammogram or ultrasound as normal (M1/U1) or benign (M2/U2) in the presence of cancer was low and broadly similar for all five radiologists (best illustrated in Figures 1 and 2). This was not the case for M3 and U3 reports in patients with cancer in whom the range across the five radiologists was 10–31% for mammography and 6–16% for ultrasound. As can again be seen most clearly in Figures 1 and 2, those with the lowest sensitivity tended to assign an excess of patients with breast cancer to M3 or U3. We have also shown that sensitivity, and consequently overall reporting performance, is closely linked to an individual's experience in breast radiology. In general, it was uncommon for any of the radiologists to report a cancer examination as normal or benign. However, with increased experience, the radiologists were more confident in reporting changes as probably or definitely malignant rather than borderline (M3/U3). When measuring a radiologist's reporting performance, their accuracy in classifying cancer patients is more discriminatory than their reporting of non-cancer patients. Consequently, when confronted with a lesion that a radiologist wishes to classify as M3 or U3 he or she should be encouraged to decide whether their level of suspicion really warrants an M4/U4 or conversely an M2/U2. It is hoped that the continuous re-evaluation of imaging classification and the regular feedback from needle biopsy results will lead to improved breast radiologist performance.

Evidence from the literature evaluating the factors affecting reporting accuracy is almost exclusively limited to screening mammography performance. Increased experience not only in years but also in reporting volume has been associated with improved performance [13]. However, this is not universal and it is clear that performance is subject to many variables [14]. The majority of the assessments in this study were performed by radiologists in single-handed clinics; however, when two radiologists are available our data would suggest that pairing a less experienced radiologist with a senior colleague might well be of benefit. Double reading of screening mammograms is routine practice, so should it be extended to symptomatic imaging? As stated above, our published "missed" cancers rate is very low and routine double reading of all symptomatic imaging would probably result in very little positive gain for a large increase in radiology workload [6]. All patients undergoing CB in this study were discussed at twice weekly multidisciplinary meetings, which is an invaluable forum for radiologist feedback on image interpretation and outcome. Ongoing training and feedback is clearly essential to sustain and continue improvement.

Conclusion

We have shown from this large data set covering over 16 000 symptomatic patient assessments the individual contribution of mammography and ultrasound to the diagnostic process. Mammography improves sensitivity by 18.3% over clinical examination and ultrasound an additional 9.8%. We have also shown that the accuracy of mammography is reduced by both decreasing age and, independently, increased breast parenchymal density. When comparing individual reporting accuracy the imaging sensitivity is the most accurate reflector of performance for both mammography and ultrasound. Reporting sensitivity was strongly correlated with radiologist experience, and those radiologists with less experience (and lower sensitivity) were relatively more likely to report a cancer as M3 or U3. To improve radiology reporting performance, the sensitivity of cancer reporting should be closely monitored and there should be regular feedback from needle biopsy results and discussion of reporting classification with more experienced colleagues.

Acknowledgments

The authors were supported with research funding from the National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre (BMRC).

The JCIS team is funded by Cancer Research UK and NIHR Cambridge BMRC.

References

1. Britton P, Sinnatamby R. The investigation of suspected breast cancer. BMJ 2007;335:347–8.
2. Royal College of Radiologists. Guidance on screening and symptomatic breast imaging, 2nd edn. London, UK: The Royal College of Radiologists, 2003.
3. The BASO Guidelines for surgeons in the management of symptomatic breast disease in the UK (1998 revision). Eur J Surg Oncol 1998;24:464–76.
4. Maxwell AJ, Ridley NT, Rubin G, Wallis MG, Gilbert FJ, Michell MJ; Royal College of Radiologists Breast Group. The Royal College of Radiologists Breast Group breast imaging classification. Clin Radiol 2009;64:624–7.
5. Wolfe JN. Breast patterns as an index of risk for developing breast cancer. AJR Am J Roentgenol 1976;126:1130–9.

6. Britton P, Duffy SW, Sinnatamby R, Wallis MG, Barter S, Gaskarth M, et al. One-stop diagnostic breast clinics: how often are breast cancers missed? Br J Cancer 2009;100:1873–8.
7. Flobbe K, Nelemans PJ, Kessels AG, Beets GL, von Meyenfeldt MF, van Engelshoven JM. The role of ultrasonography as an adjunct to mammography in the detection of breast cancer. a systematic review. Eur J Cancer 2002;38:1044–50.
8. Houssami N, Irwig L, Simpson JM, McKessar M, Blome S, Noakes J. The influence of knowledge of mammography findings on the accuracy of breast ultrasound in symptomatic women. Breast J 2005;11:167–72.
9. Flobbe K, van der Linden ES, Kessels AG, van Engelshoven JM. Diagnostic value of radiological breast imaging in a non-screening population. Int J Cancer 2001;92:616–18.
10. Houssami N, Irwig L, Loy C. Accuracy of combined breast imaging in young women. Breast 2002;11:36–40.
11. Houssami N, Ciatto S, Irwig L, Simpson JM, Macaskill P. The comparative sensitivity of mammography and ultrasound in women with breast symptoms: an age-specific analysis. Breast 2002;11:125–30.
12. D'Orsi CJ, Bassett LW, Berg WA. Breast imaging reporting and data system: ACR BI-RADS-Mammography, 4th edn. Reston, VA: American College of Radiology; 2003. Available from: http://www.acr.org
13. Esserman L, Cowley H, Eberle C, Kirkpatrick A, Chang S, Berbaum K, et al. Improving the accuracy of mammography: volume and outcome relationships. J Natl Cancer Inst 2002;94:369–75.
14. Barlow WE, Chi C, Carney PA, Taplin SH, D'Orsi C, Cutter G, et al. Accuracy of screening mammography interpretation by characteristics of radiologists. J Natl Cancer Inst 2004;96:1840–50.