

Measuring performance in clinical breast examination

G. C. Wishart^{1,2}, J. Warwick³, V. Pitsinis¹, S. Duffy³ and P. D. Britton^{1,2}

¹Cambridge Breast Unit, Cambridge University Hospitals NHS Foundation Trust, and ²National Institute for Health Research Cambridge Biomedical Research Centre, Cambridge, and ³Cancer Research UK Centre for Epidemiology, Mathematics and Statistics, Wolfson Institute of Preventive Medicine, London, UK

Correspondence to: Professor G. C. Wishart, Cambridge Breast Unit, Box 97, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, UK (e-mail: gordon.wishart@addenbrookes.nhs.uk)

Background: Clinical breast examination (CBE) remains an essential part of triple assessment of breast lumps, but to date there are no performance measures for clinicians using this technique. The aim of this retrospective audit was to compare the performance and accuracy of CBE to identify key indicators that could be used to monitor performance prospectively.

Methods: Clinical examination findings (E1, normal, to E5, malignant) for 16 585 patients who had CBE as part of triple assessment were obtained from electronic medical records. The performance of CBE, by age group, mammographic density and clinician, was assessed by calculating the sensitivity, specificity and area under the receiver operating characteristic (ROC) curve.

Results: There was marked variation in sensitivity between clinicians (range 44.6–65.9 per cent). There was a strong downward trend in the percentage classified as E5 as sensitivity for breast cancer detection decreased, and a corresponding strong downward trend in the proportion of E4 and E5 cancers classified as E5. Both of these measures could be used as indicators to monitor CBE performance.

Conclusion: The performance measures outlined here could help to identify clinicians who have a lower sensitivity for CBE and who may therefore require feedback and further training.

Paper accepted 10 March 2010

Published online 21 May 2010 in Wiley InterScience (www.bjs.co.uk). DOI: 10.1002/bjs.7108

Introduction

Clinical breast examination (CBE) remains an essential part of the triple assessment of all breast lumps¹. To date, there has been no standardization of examination technique and no performance measures exist for either surgical trainees or consultant breast specialists. Previous studies in the UK have shown that a concordance model, where trainees' examination findings are checked by a consultant surgeon, can be effective for training both junior doctors² and nurse practitioners³. However, this would not be an appropriate model to monitor specialists who are already experienced in CBE. At present there is no benchmark for an acceptable minimum standard for competency in CBE.

The aim of this study was to compare the performance and accuracy of CBE in eight clinicians who regularly examined patients in the Cambridge Breast Unit, to identify key indicators that could be used to monitor performance prospectively.

Methods

In the Cambridge Breast Unit all new symptomatic general practitioner (GP) referrals are seen in a one-stop diagnostic breast clinic with access to same-day imaging and biopsy, as required. Between January 2001 and December 2006, a total of 16 603 new patients were referred to the breast clinic, of whom 16 585 had CBE as part of their diagnostic triple assessment. CBE was performed in all patients before any breast imaging and/or biopsy. All clinic visits were recorded on Joint Clinical Information System, an electronic medical record for all breast referrals. Triple assessment included CBE, mammography, breast ultrasonography and core biopsy, where indicated. CBE findings were recorded for all patients as follows: E1, normal; E2, benign; E3, suspicious but probably benign; E4, suspicious but probably malignant; E5, malignant. Patients were allocated to one of two groups after imaging and biopsy: those with breast cancer confirmed on histopathology and those with no breast cancer.

Women who underwent mammography had the parenchymal density pattern assessed prospectively by one of six consultant radiologists, who assigned them to one of four categories using the Wolfe criteria⁴: 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, highest risk, severe dysplasia.

Statistical analysis

The performance of CBE by age group, parenchymal pattern and clinician was assessed by calculating the sensitivity, specificity and area under the receiver operating characteristic (ROC) curve. Sensitivity was defined as the percentage of women with breast cancer identified correctly (category E4 or E5) at CBE. Specificity was defined as the percentage of women without breast cancer identified correctly (category E1, E2 or E3) at CBE. The area under the ROC curve is the average of the sensitivity and specificity: (sensitivity + specificity)/2.

Clinicians who saw fewer than 450 women and/or 40 cancers during the study interval were excluded from the by-clinician subanalysis. Bar charts were used to show how the distribution of clinical assessment (E1–E5) varied by clinician, in the group as a whole and in women with breast cancer only, in particular to investigate whether a clinician ‘sitting on the fence’ (overuse of category E3) was indicative of a particularly high or low sensitivity for detection of breast cancer.

Logistic regression analysis in women with breast cancer only was used to investigate the relationship between age and parenchymal pattern and the chance of cancer being detected by CBE. This was essentially a multivariable model with sensitivity as the endpoint, and age and parenchymal pattern as the explanatory (prognostic) variables. The model was further expanded to include each clinician (an eight-level indicator variable representing the eight clinicians who each examined more than 450 women and/or 40 cancers during the study) to test for a difference between the clinicians, with regard to their likelihood of detecting breast cancer with CBE, after adjusting for age and parenchymal pattern.

Results

A total of 16 603 women were referred during the study interval, of whom 16 585 had CBE as part of triple assessment, including 1232 (7.4 per cent) with a first

diagnosis of breast cancer. Details of tumour size were available for 739 of these patients: median 21 (interquartile range 15) mm. Seventeen tumours (2.3 per cent) were 4 mm or less, 49 (6.6 per cent) were 5–9 mm, 259 (35.0 per cent) were 10–19 mm, 231 (31.3 per cent) were 20–29 mm and 183 (24.8 per cent) were 30 mm or more. The remaining 15 353 women underwent clinical examination but were found not to have breast cancer: 1829 (11.9 per cent) had CBE alone, 3672 (23.9 per cent) had CBE and breast ultrasonography, and 9852 (64.2 per cent) had CBE, breast ultrasonography and mammography. The core biopsy rate for the study interval was 10.7 per cent.

The sensitivity, specificity, area under the ROC curve and prevalence of breast cancer are shown in *Table 1* for all 16 585 women undergoing CBE, by age. As expected, the prevalence of breast cancer, sensitivity of CBE and area under the ROC curve increased with increasing age, as specificity decreased. The reduction in specificity was particularly marked between groups aged 80–89 and 90 years or more, where it fell from 83.2 to 59.4 per cent. However, the number of women in the oldest age group was relatively small (68 patients) and the 95 per cent confidence interval fairly wide (40.6 to 76.3 per cent). The difference between the age groups, in terms of area under the ROC curve, was highly significant ($\chi^2 = 254.43$, 7 d.f., $P < 0.001$).

The sensitivity, specificity, area under the ROC curve and prevalence of breast cancer in 10 520 women examined using both CBE and mammography are shown by breast parenchymal pattern in *Table 2*. Perhaps surprisingly, the prevalence of breast cancer decreased as breast density increased, with a highly significant difference between the parenchymal patterns with respect to the area under the ROC curve ($\chi^2 = 91.37$, 3 d.f., $P < 0.001$). The sensitivity

Table 1 Sensitivity, specificity, area under the receiver operating characteristic (ROC) curve and prevalence of breast cancer in women undergoing clinical breast examination at the symptomatic breast clinic, by age

Age group (years)	Prevalence of breast cancer (%)	Sensitivity (%)	Specificity (%)	Area under ROC curve	No. of women
< 30	0.4	11.1	100.0	0.56	2286
30–39	2.6	25.0	99.7	0.62	3687
40–49	5.4	32.7	99.3	0.66	4804
50–59	7.4	43.4	98.5	0.71	2870
60–69	13.0	62.2	97.6	0.80	1423
70–79	26.0	85.2	93.2	0.89	1016
80–89	41.0	85.7	83.2	0.85	431
≥ 90	53.0	88.9	59.4	0.74	68
Overall	7.4	58.5	98.6	0.79	16 585

ROC, receiver operating characteristic.

Table 2 Sensitivity, specificity, area under the receiver operating characteristic (ROC) curve and prevalence of breast cancer in women undergoing clinical breast examination at the symptomatic breast clinic, by parenchymal pattern on mammography

Parenchymal pattern	Prevalence of breast cancer (%)	Sensitivity (%)	Specificity (%)	Area under ROC curve	No. of women	Mean age (years)
N1	13.0	80.8	97.3	0.89	1310	59
P1	13.0	68.5	97.6	0.83	2531	55
P2	10.0	51.3	98.3	0.75	4554	49
DY	6.3	37.6	99.0	0.68	2125	46
Overall	10.0	59.3	98.2	0.79	10 520	46

ROC, receiver operating characteristic; 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, highest risk, severe dysplasia.

and area under the ROC curve also decreased with increasing density. To investigate whether this decrease in breast cancer prevalence with increasing density was merely a reflection of the age distribution of clinic attendees (most Wolfe mammographic DY patterns would be expected in the younger age groups and thus with a lower prevalence of breast cancer), a logistic regression analysis was performed with breast cancer diagnosis (yes/no) as the dependent variable, and age and parenchymal pattern as explanatory (independent) variables. Both age and parenchymal pattern were significant in the model, and the relative risk of developing breast cancer was greatest in category P2 (data not shown), suggesting that the observed (negative) association between parenchymal pattern and breast cancer risk was not explained by patient age. The distribution of parenchymal pattern by age group is shown in *Table 3*.

An analysis of CBE performance, by clinician, is shown in *Table 4* for the eight clinicians – four consultant breast surgeons, one registrar, one nurse consultant, one nurse practitioner and one associate (GP) specialist. Between them, a total of 13 184 women, including 969 (7.3 per cent) who were found to have breast cancer, were examined.

Table 4 Sensitivity, specificity, area under the receiver operating characteristic (ROC) curve and prevalence of breast cancer, by clinician

Clinician	Prevalence of breast cancer (%)	Sensitivity (%)	Specificity (%)	Area under ROC curve	No. of women
1	8.4	65.9	98.0	0.82	486
2	7.9	62.1	99.2	0.81	1940
3	7.3	58.5	98.8	0.79	6076
4	6.1	53.0	98.2	0.76	1089
5	7.1	52.6	98.5	0.76	1617
6	9.3	52.4	99.3	0.76	450
7	8.0	51.1	97.4	0.74	584
8	6.9	44.6	99.0	0.72	942
Overall	7.3	56.8	98.7	0.78	13 184

ROC, receiver operating characteristic.

There was marked variation in sensitivity between the clinicians (range 44.6–65.9 per cent), but relatively little difference in specificity (range 97.4–99.3 per cent). For some clinicians the confidence intervals were very wide, however, reflecting the fact that they examined fewer

Table 3 Parenchymal patterns in patients undergoing clinical breast examination and mammography, by age

Age group (years)	N1	P1	P2	DY	Not done	Total
< 30	3 (0.1)	2 (0.1)	6 (0.3)	3 (0.1)	2272 (99.4)	2286 (100)
30–39	128 (3.5)	342 (9.3)	864 (23.4)	516 (14.0)	1837 (49.8)	3687 (100)
40–49	271 (5.6)	760 (15.8)	1963 (40.9)	1086 (22.6)	724 (15.1)	4804 (100)
50–59	275 (9.6)	532 (18.5)	1000 (34.8)	383 (13.3)	680 (23.7)	2870 (100)
60–69	235 (16.5)	375 (26.4)	415 (29.2)	85 (6.0)	313 (22.0)	1423 (100)
70–79	270 (26.6)	341 (33.6)	217 (21.4)	34 (3.3)	154 (15.2)	1016 (100)
80–89	113 (26.2)	157 (36.4)	77 (17.9)	15 (3.5)	69 (16.0)	431 (100)
≥ 90	15 (22.1)	22 (32.4)	12 (17.6)	3 (4.4)	16 (23.5)	68 (100)
Overall	1310 (7.9)	2531 (15.3)	4554 (27.5)	2125 (12.8)	6065 (36.6)	16 585 (100)

Values in parentheses are percentages. 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, highest risk, severe dysplasia.

Table 5 Odds ratios from univariable and multivariable logistic regression models for chance of breast cancer being detected by clinical breast examination (category E4 or E5) in 859 patients with breast cancer, by age, parenchymal pattern and consultant

	No. in category E4 + E5 (of total cancers)	Univariable		Multivariable	
		Odds ratio	P*	Odds ratio	P*
Age (years)			< 0.001‡		< 0.001‡
< 39	15 of 65	1.00		1.00	
40–49	54 of 189	1.33 (0.69, 2.57)		1.32 (0.68, 2.56)	
50–59	68 of 153	2.67 (1.38, 5.16)		2.66 (1.36, 5.22)	
60–69	74 of 121	5.25 (2.65, 10.39)		4.93 (2.43, 9.99)	
70–79	163 of 192	18.75 (9.31, 37.70)		16.94 (8.09, 35.46)	
80–89	103 of 117	24.52 (10.99, 54.74)		21.86 (9.42, 50.73)	
≥ 90	19 of 22	21.11 (5.49, 81.23)		19.77 (4.99, 78.27)	
Parenchymal pattern			< 0.001‡		0.237‡
N1	104 of 128	1.00		1.00	
P1	172 of 258	0.46 (0.28, 0.77)		0.80 (0.45, 1.42)	
P2	179 of 367	0.22 (0.13, 0.36)		0.74 (0.42, 1.31)	
DY	41 of 106	0.15 (0.08, 0.26)		0.66 (0.33, 1.32)	
Clinician			0.362§		0.338§
1	21 of 33	1.24 (0.60, 2.60)		1.64 (0.71, 3.79)	
2	89 of 138	1.29 (0.86, 1.93)		1.43 (0.90, 2.28)	
3	232 of 397	1.00		1.00	
4	29 of 53	0.86 (0.48, 1.53)		0.86 (0.44, 1.67)	
5	57 of 105	0.84 (0.55, 1.30)		0.88 (0.53, 1.47)	
6	20 of 38	0.79 (0.41, 1.54)		0.89 (0.41, 1.94)	
7	22 of 38	0.98 (0.50, 1.92)		1.03 (0.47, 2.25)	
8	26 of 57	0.60 (0.34, 1.04)		0.61 (0.32, 1.15)	

Values in parentheses are 95 per cent confidence intervals. 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, highest risk, severe dysplasia. *Likelihood ratio test; ‡linear trend; §global.

women. The difference between clinicians with respect to the area under the ROC curve was not significant ($\chi^2 = 10.09$, 7 d.f., $P = 0.18$). This analysis did not account for possible imbalances (with regard to age and parenchymal pattern) between the women examined by each clinician.

To investigate the independent effects of these factors, a logistic regression analysis was carried out with the chance of a cancer being assessed as E4 or E5 (having breast cancer detected at CBE) as the dependent variable, and age, parenchymal pattern and consultant as the explanatory variables. The univariable and multivariable results are shown in Table 5. On univariable analysis, both age and parenchymal pattern were highly significant ($P < 0.001$), but, after adjusting for age, parenchymal pattern was not significant in the multivariable model. The linear trend for decreasing sensitivity with increasing density was nevertheless clear, suggesting that this result may be due to lack of power. Refitting using a partial proportional odds model, which is more powerful because it allows clinical assessment to be included as a five-level variable, confirmed this. The odds ratios associated with P1, P2 and DY from this more complicated model were similar to those obtained for the multivariable model in Table 5 (0.95, 0.71 and 0.66 respectively) and the P value for the

test of linear trend was 0.040. Although the sensitivity of CBE varied considerably between clinicians, this was not statistically significant (the observed variation was no greater than would be expected by chance if the clinicians were equally competent).

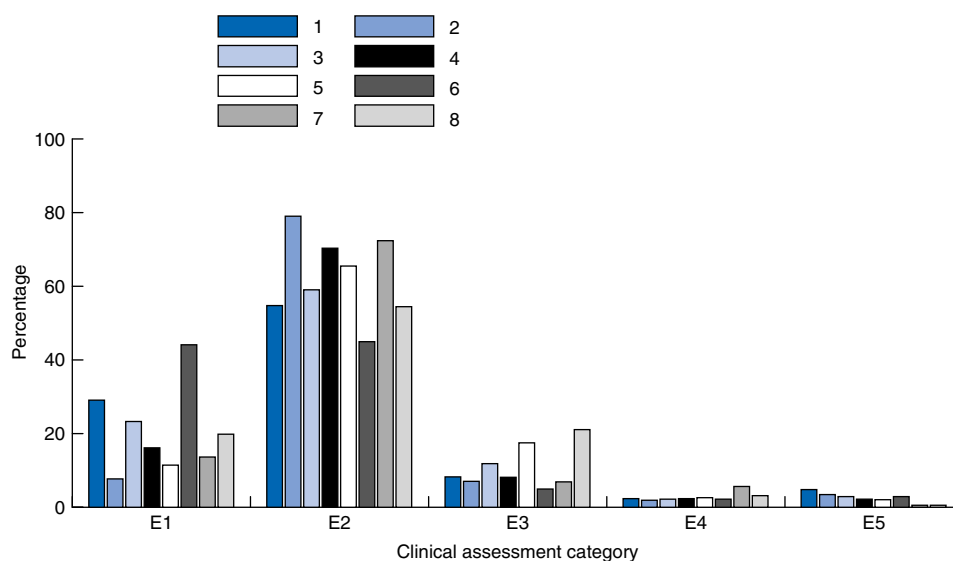
The above multivariable logistic regression analysis was repeated for low- and high-density mammographic patterns separately, N1 and P1 comprising the 'low mammographic density' group, and P2 and DY comprising the 'high mammographic density' group (Table 6). The numbers in each group were therefore smaller, and the power was correspondingly reduced, but the trend for increasing chance of breast cancer detection (with CBE) with increasing age was nevertheless significant in both low- and high-density groups. In all but the oldest age group, the odds ratios were highest in women with high mammographic density. The variation between clinicians was not statistically significant (Table 6).

The percentage of women in each clinical assessment category, for each of the eight clinicians separately, is shown in Fig. 1. Clinicians were numbered 1–8 in order of decreasing CBE sensitivity. Although two of the poorer performers did appear to allocate an excess number to category E3, there was no consistent trend and thus no evidence

Table 6 Odds ratios from multivariable logistic regression models for chance of breast cancer being detected by clinical breast examination (category E4 or E5) in patients with breast cancer, by age and clinician, for low and high mammographic densities

	Low mammographic density*			High mammographic density†		
	No. in category E4 + E5 (of total cancers)	Odds ratio	P‡	No. in category E4 + E5 (of total cancers)	Odds ratio	P‡
Age (years)			< 0.001§			< 0.001§
< 39	3 of 8	1.00		12 of 57	1.00	
40–49	14 of 37	0.76 (0.15, 3.85)		40 of 152	1.37 (0.65, 2.88)	
50–59	24 of 51	1.22 (0.26, 5.78)		44 of 102	3.02 (1.41, 6.47)	
60–69	36 of 56	2.24 (0.47, 10.75)		38 of 65	5.62 (2.48, 12.72)	
70–79	109 of 131	6.91 (1.50, 31.80)		54 of 61	29.44 (10.56, 82.05)	
80–89	75 of 87	8.82 (1.81, 42.94)		28 of 30	53.69 (11.11, 259.44)	
≥ 90	15 of 16	23.54 (1.91, 289.44)		4 of 6	8.47 (1.34, 53.49)	
Clinician			0.490¶			0.623¶
1	14 of 17	2.78 (0.70, 11.02)		7 of 16	1.06 (0.32, 3.54)	
2	51 of 66	1.32 (0.64, 2.72)		38 of 72	1.57 (0.85, 2.88)	
3	128 of 179	1.00		104 of 218	1.00	
4	19 of 26	1.01 (0.36, 2.76)		10 of 27	0.85 (0.34, 2.11)	
5	25 of 39	0.54 (0.24, 1.26)		32 of 66	1.11 (0.58, 2.11)	
6	12 of 17	0.90 (0.27, 2.99)		8 of 21	0.93 (0.33, 2.58)	
7	9 of 13	0.76 (0.20, 2.90)		13 of 25	1.20 (0.47, 3.10)	
8	18 of 29	0.72 (0.29, 1.78)		8 of 28	0.52 (0.20, 1.34)	

Values in parentheses are 95 per cent confidence intervals. *Parenchymal pattern N1 or P1; †parenchymal pattern P2 or DY (see Methods for definitions). ‡Likelihood ratio test; §linear trend; ¶global.

**Fig. 1** Percentage of women in each clinical assessment category for each of the eight clinicians. Clinician 1 had the highest sensitivity for clinical breast examination and clinician 8 the lowest

to suggest that poor performance was due to clinicians 'sitting on the fence'. In general, the percentage of women categorized as E5 fell with decreasing CBE sensitivity.

The distribution of findings by clinical assessment, by clinician and in order of decreasing CBE sensitivity, is

shown in *Table 7*. As sensitivity for breast cancer detection decreased, there was a strong downward trend in the percentage classified as E5, and a corresponding strong downward trend in the percentage of E4 and E5 cancers classified as E5. These two features are potential measures

Table 7 Distribution of findings per clinical assessment in patients with breast cancer, by clinician and sensitivity

Clinician	E1 (%)	E2 (%)	E3 (%)	E4 (%)	E5 (%)	E5/(E4 + E5) (%)	Sensitivity (%)	No. of cancers seen
1	2.4	14.6	17.1	17.1	48.8	74.1	65.9	41
2	1.3	13.7	22.9	20.3	41.8	67.4	62.1	153
3	2.9	14.3	24.3	22.7	35.8	61.2	58.5	441
4	1.5	24.2	21.2	18.2	34.9	65.7	53.0	66
5	2.6	10.5	34.2	23.7	28.9	55.0	52.6	114
6	4.8	19.1	23.8	23.8	28.6	54.5	52.4	42
7	2.1	23.4	23.4	42.6	8.5	16.7	51.1	47
8	1.5	12.3	41.5	33.9	10.8	24.1	44.6	65

E1, normal; E2, benign; E3, suspicious but probably benign; E4, suspicious but probably malignant; E5, malignant.

of clinical competence with CBE and could be used as quality control indicators for monitoring CBE performance.

Discussion

This study demonstrated wide variability (44.6–65.9 per cent) in the sensitivity of CBE in breast clinicians working in a busy one-stop diagnostic breast clinic. Although clinicians with a higher sensitivity tended to have a lower specificity, specificity rates were similar, reflecting the large number of patients in a symptomatic cohort who have a normal finding, as well as the ability of clinicians to categorize a patient without cancer as E1, E2 or E3 correctly. Higher sensitivity for CBE was associated with classifying a higher percentage of patients with clinical findings suspicious of malignancy (E4 and E5) as having E5 disease. This is likely to reflect confidence or experience in the clinical diagnosis of breast cancer. After adjusting for age and parenchymal pattern, the proportion of cases, E5/(E4 + E5), was not so strongly correlated with CBE sensitivity, but this still appears to remain a useful way to monitor performance and provide feedback to breast examiners.

It should be noted that, even in this relatively large cohort, the number of breast cancers was relatively small (7.4 per cent of referrals). As each clinician examined relatively few patients with breast cancer, there was little power to detect small to moderate (and possibly even large) differences in performance. Thus, it is not practical to rely on statistical testing to determine whether or not a clinician is performing as well as he or she should.

Most of the previously published data on CBE sensitivity relate to the use of CBE in the breast screening setting. In a review of North American clinical trials using CBE as a screening tool, Barton and colleagues⁵ reported a sensitivity and specificity of 54 and 94 per cent respectively. A previous study of 468 women with screen-detected breast cancer showed that CBE sensitivity increased with

larger tumour size and decreased with higher bodyweight and extremes of age; age groups 40–49 and 80 years or more had the lowest sensitivities⁶. Although the findings from the present study concur with a lower sensitivity in younger age groups, CBE sensitivity continued to rise with increasing age, and the findings of Oestreicher and co-workers⁶ can probably be explained by the fact that their study included only screen-detected breast cancers, many of which would have been impalpable in the older age groups. A finding from the Oestreicher study that is more difficult to explain is that increased CBE sensitivity was also associated with increased radiographic density, increasing from 23 per cent in women with fatty breasts to 43 per cent in dense breasts. This appears counterintuitive, and conflicts with the present findings where increased breast density was strongly associated with decreasing CBE sensitivity, but may again reflect differences in screen-detected and symptomatic populations.

The introduction of a scoring system for CBE (E1–E5)³ has made it possible to monitor performance of this clinical skill during training or, as in this study, to compare performance between individual clinicians performing CBE in breast clinics for symptomatic patients. Triple assessment remains the standard of care for the investigation and diagnosis of patients with a lump in the breast. A recent audit of over 7000 patients who had core biopsy-based triple assessment in a one-stop diagnostic breast clinic setting documented a diagnostic accuracy of 99.6 per cent with a missed cancer rate of 1.7 per 1000 women discharged⁷. This study also highlighted that the lowest sensitivity of triple assessment was in younger women aged 40–49 years, in keeping with the present finding that younger age was associated with a lower sensitivity for CBE.

Although CBE is an important clinical skill that clearly plays an essential part of the triple assessment of breast lumps, it is surprising that there is currently

no structured training programme for junior doctors in the UK, in contrast to North America where several approaches have been tried^{8–13}. Interestingly, another North American study by Wiecha and Gann¹⁴ documented that 43 per cent of doctors felt they had a poor CBE technique and only 50 per cent felt moderately confident. Although not previously identified¹⁵, breast examination has now been listed as a core clinical skill for breast surgical trainees in the general surgery syllabus produced by the Intercollegiate Surgical Curriculum Programme¹⁶. Unfortunately this document contains no guidelines for training in CBE, or any minimum standards to be attained for completion of training in breast surgery. Two more recent publications from North America¹⁷ and Australia¹⁸ have at least attempted to describe a methodology for standardization of CBE, but failed to provide any evidence of sensitivity or performance targets in the setting of patients with symptomatic breasts.

Based on the data presented here, a cautious and self-regulatory approach to quality control for CBE is advised. Together with sensitivity for breast cancer detection, use of the percentage of E4 plus E5 cancers classified as E5 allows an additional measure of clinical competence with a greater range of results (74.1 to 16.7 per cent), which clearly identified the two outliers in the present study who had the lowest overall sensitivity. This simple value, together with the percentage of cancers scored as E5, should be calculated after clinicians have seen a minimum of 100 cancers, with adjustment for age and parenchymal pattern. With the approach of revalidation for all doctors in the UK, measurement of clinical outcomes and performance will play an essential role in that process; clinicians performing CBE should ensure that they are able to document their own outcomes for this important clinical skill.

Acknowledgements

G.W. and P.B. receive research funding from the National Institute for Health Research, Cambridge Biomedical Research Centre. The authors declare no conflict of interest.

References

- 1 Dixon JM, Sainsbury JRC. *Handbook of Diseases of the Breast* (2nd edn). Churchill Livingstone: Edinburgh, 1998.
- 2 Cunnick GH, Upponi S, Wishart GC. Junior doctor training in breast and axillary examination: a comparative study. *Breast* 2000; **9**: 312–314.
- 3 Chapman D, Purushotham A, Wishart G. Nurse practitioner training in breast examination. *Nurs Stand* 2002; **17**: 33–36.
- 4 Wolfe JN. Breast patterns as an index of risk for developing breast cancer. *AJR Am J Roentgenol* 1976; **126**: 1130–1137.
- 5 Barton MB, Harris R, Fletcher SW. The rational clinical examination. Does this patient have breast cancer? The screening clinical breast examination: should it be done? *JAMA* 1999; **282**: 1270–1280.
- 6 Oestreicher N, White E, Lehman CD, Mandelson MT, Poeter PL, Taplin SH. Predictors of sensitivity of clinical breast examination (CBE). *Breast Cancer Res Treat* 2002; **76**: 73–81.
- 7 Britton PD, Duffy S, Sinnatamby R, Wallis M, Barter S, O'Neill A *et al.* One-stop diagnostic breast clinics: how often are breast cancers missed? *Br J Cancer* 2009; **100**: 1873–1878.
- 8 Lee KC, Dunlop D, Dolan N. Do clinical breast examination skills improve during medical school? *Acad Med* 1998; **73**: 1013–1019.
- 9 Sloan DA, Donnelly MB, Plymale MA, Schwartz RW, Edwards MJ, Singletary SE *et al.* Improving residents' clinical skills with the structured clinical instruction module for breast cancer: results of a multiinstitutional study. Breast Cancer Education Working Group. *Surgery* 1997; **122**: 324–333.
- 10 Chalabian J, Garman K, Wallace P, Dunnington G. Clinical breast evaluation skills of house officers and students. *Am Surg* 1996; **62**: 840–845.
- 11 Benincasa TA, King ES, Rimer BK, Bloom HS, Balshem A, James J *et al.* Results of an office-based training program in clinical breast examination for primary care physicians. *J Cancer Educ* 1996; **11**: 25–31.
- 12 Costanza ME, Greene HL, McManus D, Hoople NE, Barth R. Can practicing physicians improve their counseling and physical examination skills in breast cancer screening? A feasibility study. *J Cancer Educ* 1995; **10**: 14–21.
- 13 Pilgrim C, Lannon C, Harris RP, Cogburn W, Fletcher SW. Improving clinical breast examination training in a medical school: a randomized controlled trial. *J Gen Intern Med* 1993; **8**: 685–688.
- 14 Wiecha JM, Gann P. Provider confidence in breast examination. *Fam Pract Res J* 1993; **13**: 37–41.
- 15 Specialist Advisory Committee in General Surgery, Society of Academic and Research Surgery (SARS). A curriculum, organisation and syllabus for higher surgical training in general surgery and its sub-specialties. SARS: London, 1998.
- 16 Intercollegiate Surgical Curriculum Programme (ISCP). *Syllabus in General Surgery*. Joint Committee on Surgical Training: London, 2007.
- 17 Saslow D, Hannan J, Osuch J, Alciati MH, Baines C, Barton M *et al.* Clinical breast examination: practical recommendations for optimizing performance and reporting. *CA Cancer J Clin* 2004; **54**: 327–344.
- 18 Thistlethwaite J, Stewart RA. Clinical breast examination for asymptomatic women – exploring the evidence. *Aust Fam Physician* 2007; **36**: 145–150.