available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

Breast screening with ultrasound in women with mammography-negative dense breasts: Evidence on incremental cancer detection and false positives, and associated cost

Vittorio Corsetti[a], *Nehmat Houssami*[b,c], *Aurora Ferrari*[a], *Marco Ghirardi*[a], *Sergio Bellarosa*[a], *Osvaldo Angelini*[a], *Claudio Bani*[a], *Pasquale Sardo*[a], *Giuseppe Remida*[a], *Enzo Galligioni*[a], *Stefano Ciatto*[c,*]

[a]*Ambulatori Oncologici Raphael, Cooperativa Sociale O.N.L.U.S., Calcinato (BS), Italy*
[b]*Screening and Test Evaluation Program (STEP), School of Public Health, University of Sydney, Sydney, Australia*
[c]*CSPO, Istituto Scientifico per la Prevenzione Oncologica, Viale A. Volta., 171 – I-50131 Florence, Italy*

ARTICLE INFO

Article history:
Received 17 August 2007
Received in revised form
9 January 2008
Accepted 11 January 2008
Available online 11 February 2008

Keywords:
Breast cancer
Screening
Ultrasonography
Dense breast
Interval cancers

ABSTRACT

Background: We evaluated the contribution of ultrasound (US) in detecting breast cancer in women with dense breasts and negative mammograms.
Methods: 9157 (35.8%) of 25,572 self-referring women during 2000–2007 had BI-RADS D3–4 negative mammograms – all were screened with bilateral US.
Results: US detected 37 cancers – incremental cancer detection rate (ICDR) was 0.40% (95% CI: 0.39–0.41%); ICDR was 0.33% in women <50 and 0.51% in those 50 years and older. US detected a larger proportion of cancers below age 50 compared to older women. US-only detected cancers had a more favourable stage (pTis–pT1a–pT1b: 64.8% versus 35.5%, $p = 0.001$; pN1: 13.5% versus 31.3%, $p = 0.047$) than cancers detected on mammography. US caused additional investigations in 4.9% of women and benign surgical biopsies in 0.9%. Cost per US-screened woman, and per US-detected cancer ranged between €59–62 and €14,618–15,234, respectively.
Conclusion: US detects early-stage cancers in women with mammography-negative dense breasts, with higher contribution in women younger than 50 years.

1. Introduction

Breast density reduces the sensitivity of both diagnostic[1] and screening mammography[2–5], is associated with a higher rate of interval cancers[2,5], and is one of the reasons for reduced screening efficacy in women younger than 50 years. The sensitivity of ultrasound (US) is less affected by age than mammography[6] and US-detection of mammographically-occult cancers is encountered particularly in dense breasts. The contribution of US in screening women with dense breasts on mammography has been examined in non-randomised studies[4,7–11] that have looked at additional cancers detected with US (but none have documented the associated cost), and the outcome of a prospective randomised study is awaited.[12] A review of the evidence on ultrasound in breast screening concluded that, whilst US may have an adjunct role

* *Corresponding author:* Tel.: +39 055 5012275; fax: +39 055 5001623.
E-mail address: s.ciatto@cspo.it (S. Ciatto).
0959-8049/$ - see front matter
doi:10.1016/j.ejca.2008.01.009

in cancer detection in dense breasts, its main limitation is its high false positive rate.[13]

We report one of the largest studies to date of the detection of breast cancer using ultrasound in self-referring women with dense breasts and a negative report on mammography. We present data on incremental detection and false positive findings, and estimate the cost of adjunct screening US.

2. Materials and methods

The study centre (Raphael Clinic) is a charity-funded multidisciplinary breast service for women self-referring to mammography for screening (including those requesting screening due to perceived risk) or for symptoms. We included all 26,047 consecutive mammograms in women attending the centre between January 2000 and February 2007 (mean subject age 52 years). However, cancers in women attending for symptoms are excluded from our primary analysis of cancer detection rates since these women routinely undergo clinical and US assessment in addition to mammography.

For the timeframe of the study, all mammograms were reported by one of six radiologists, and were classified according to the proportion of breast volume occupied by fibroglandular density in line with BI-RADS[14] (D1 = 0–25, D2 = 26–50, D3 = 51–75, D4 = 76–100%). All subjects with negative BI-RADS D3 and D4 mammography underwent bilateral breast US, using an Aloka Pro Sound SSD–5500 unit and a multifrequency linear probe operated at 7.5–10 MHz. US was performed by one of six operators (2 radiologists, 4 breast physicians), with approximately half of subjects having the US on the same day as the mammography, and about half having the US within 4 weeks of the mammography (due to logistics of the service). While some variation in scanning technique always exists in the setting of several operators, the clinical protocol required US operators to scan in vertical and horizontal parallel stripes covering the breast, axillary tail, and areola region, so that the entire breast volume was scanned twice. The ipsilateral axilla was only scanned if a significant abnormality was identified in the breast.

Breast imaging (mammography and US) in the study centre is reported using a classification system widely used in Europe and Australia (1 – normal, 2 – benign, 3 – indeterminate, 4 – suspicious, 5 – malignant) with published estimates of the likelihood of malignancy.[15,16] US findings classified as 3–5 were further assessed with detailed mammography views, percutaneous needle biopsy (needle cytology or core biopsy) and surgical biopsy where indicated.

To assist judgement of the generalisability of our findings, mammography negative and US-detected cancers (invasive or *in situ* malignancy) were reviewed to reasonably ascertain 'negativity' on mammography. Review consisted of the combination of internal review (performed independently by four radiologists involved in mammography reporting) and an external expert radiologist (who has reported over 200,000 mammograms) using a blinded and 1:4 (cancer:control ratio) case–mix method.[17] Data of the outcome of this review are reported based on the majority opinion for classifying cases as positive (3–5 on mammography).

We report data on the incremental cancer detection rate (ICDR) for US. ICDR was calculated as the rate of US-only detected cancers amongst mammography-negative subjects undergoing systematic US for radiologically dense breast. To quantify false positives (FP), we calculated the frequency and type of tests and procedures (cytology, core biopsy and surgical biopsy) generated by abnormal (3–5) US findings in women with negative or benign outcomes. Based on current tariffs, we calculated the additional cost per woman examined with US, as well as that for each additional cancer detected by US. The cost of surgical biopsy was computed according to two alternatives, namely, (a) excisional biopsy under local anaesthesia on an outpatient basis, and (b) excisional biopsy under general anaesthesia as a hospitalised patient.

3. Results

Overall, 216 cancers were detected in 26,047 mammography examinations (0.82%). Cancer was suspected on the basis of mammography in 475 women, who underwent further assessment and 166 cancers were diagnosed (positive predictive value for recall to assessment = 36.3%). TNM pT category was pTis in 26, pT1mic in 6, pT1a in 6, pT1b in 21, pT1c in 65, pT1 (no substage available) in 3, pT2 in 31, pT3 in 1 or pT4 in 7 cases, respectively. There were 25,572 negative mammography reports, and of those 9157 (35.8%) were classified as D3–4, and underwent US examination. US identified 50 additional cancers, however, of these 50 cancers, 13 women declared symptoms or were considered to have clinical findings (Table 1) and were therefore excluded, leaving 37 US-only detected cancer in asymptomatic subjects. A flow-diagram summarising the outcomes of this study is presented in Fig. 1. The ICDR was 0.40% (95% CI: 0.39–0.41%) for the total group, 0.33% in women aged <50 years and 0.51% in those ⩾50 years. US detected 41.3% of the cancers in women aged

Table 1 – Relative incremental cancer detection for US according to age-group

	Women <50 years	Women >50 years	All cancers
Mammography detected cancers	40	126	166
US-detected cancers in mammography-negative dense breasts in women with symptoms	6	7	13
US-detected cancers in mammography-negative dense breasts in asymptomatic women	19	18	37
Relative incremental cancer detection at US (asymptomatic women) over mammography detected	41.3% (19/46)	13.5% (18/133)	20.6% (37/179)

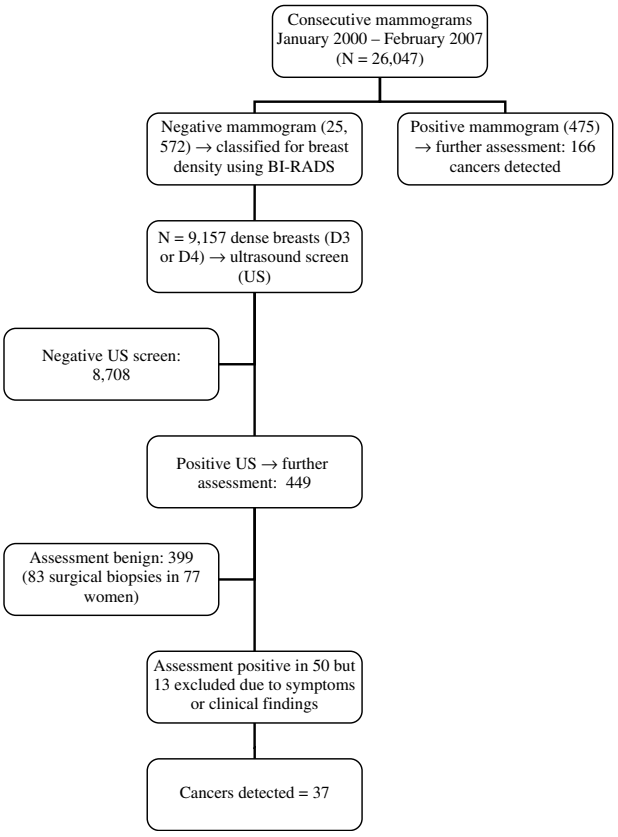

Fig. 1 – Flow-diagram summarising the study design and main outcomes.

<50 and 13.5% in those aged ⩾50 years ($\chi^2_{df1} = 8.41$, $p = 0.003$). Age-related cancer detection is presented for both mammography and US in Table 1, indicating that the largest contribution of US in screening dense breasts is in women younger than 50 years.

Original mammograms were available for the review process described in our methods in 33 of the 37 US-only detected cancers in asymptomatic women. Combining both internal and external reviewers, 8 cases were considered to have a mammographic finding (correlating with the cancer subsequently identified on US) that could have warranted recall. All such cases were classified as asymmetric density or non-specific density. The review process yielded an additional recall in about 10% of normal (control) mammography screens seeded in the set.

Table 2 shows the stage (pT and pN) distribution of cancers detected at mammography and those that were only detected with US. The proportion of early-stage cancer (pTis–pT1a–pT1b) was significantly higher in the US-only detected (64.8% or 24/37) than the mammography detected (35.5% or 59/166) cancers ($\chi^2_{df1} = 9.58$, $p = 0.001$). Axillary node metastases were significantly less frequent in the US-only detected (13.5% or 5/37) than the mammography detected (31.3% or 52/166) cancers ($\chi^2_{df1} = 4.79$, $p = 0.047$).

Overall, US caused additional investigations in 4.9% (449/9157) of subjects. Additional tests or surgery caused by US screening were 490 aspiration cytology examinations and/or 24 core biopsies, and 133 surgical biopsies. Using current average tariffs of different procedures, the cost per each additional cancer detected by US only is estimated and reported in Table 3. From the 133 surgical biopsies attributed to US

Table 2 – Stage (pT and pN) distribution in cancers detected at mammography and additional US-only detected cancers

Stage variable	Mammography detected	US-detected
Tumour size		
pTis	26	1
pT1mic	6	–
pT1a	6	3
pT1b	21	20 (one case cT1b)
pT1c	65	10
pT1	3	–
pT2	31	2 (one case cT2)
pT3	1	1
pT4	7	–
Node status		
pN0	113	30
pN1	52	5
pNx	1	2
All cancers	166	37

Table 3 – Estimates of costs of including US in screening mammography negative women with dense breasts

Variable	Number of procedures	Unitary cost (€)	Total cost (€)
US	9157	36.15	331,025.55
US (factoring the cost of BCE)[A]		54.74	501,254.18
US-guided needle biopsy (procedure)	514	37.18	19,110.52
Cytology (FNAB)	490	33.78	16,552.20
Surgical biopsy as			
(a) Outpatient surgical biopsy (local anaesthesia)	133	29.75	3,956.75[a]
(b) Inpatient (hospital) surgical biopsy (general anaesthesia)	133	201.32	26,775.56[b]
Histopathology (CNB (N = 24) or surgical biopsy)	157	46.48	7,297.36
Total cost (€)	540,873.65[a] or 563,692.46[b]		
Estimated cost per US-only detected cancer (N = 37)	14,618.20[a] or 15,234.93[b]		
Estimated incremental cost of including US in screening per woman examined with US (N = 9157)	59.06[a] or 61.55[b]		

A BCE (Breast clinical examination is a mandatory requirement for US scanning in some heath systems).
a Outpatient surgical biopsy (local anaesthesia).
b Inpatient (hospital) surgical biopsy (general anaesthesia).

screening, 83 were due to false positive findings (benign on surgical biopsy) – this represents 0.9% of subjects (83/9157) examined with US.

4. Discussion

This is a study of the incremental detection of breast cancer by US in asymptomatic women with mammography-negative dense breasts, and cannot tell us about the expected long-term benefit (mortality reduction) of the cancers detected by US. We report an incremental cancer detection rate of 0.40% in women with mammography-negative dense breasts, further confirming earlier studies of this issue. The most important aspect of our work, however, is the balance between additional cancers detected on US and associated false positives (in particular unnecessary surgery in women who do not have breast cancer), the latter being a major limitation of US in the screening context. To allow judgement of the implications of our data, we have summarised key outcome measures of previous studies of US in breast screening in Table 4: these data indicate that we were able to achieve a similar cancer detection rate as reported by other studies, whilst maintaining a very low surgical biopsy rate (and substantially lower than previously reported studies). Furthermore, we were able to achieve this using basic technology and in the hands of 6 dedicated operators (relative to studies that have relied on a single highly expert operator). We feel that this adds to the generalisability of our findings.

Overall US screening of mammography-negative dense breasts contributed an additional cancer detection rate of 20% in asymptomatic women compared to mammography alone. This contribution was substantially higher for younger than older women in terms of the proportion of cancers detected from all cancers in each age-group (an additional 41.3% at age <50 relative to an additional 13.5% at age $\geqslant$50). This suggests that routine adjunct US screening in asymptomatic women might provide the highest relative cancer detection yield if applied to women with dense breasts aged less than 50 years.

It may be argued that, since our population consisted of self-referring women, some of whom had symptoms and some were requesting screening outside of the normal age-range for screening, our subjects had higher risk of breast cancer. This may potentially bias data on cancer detection. Whilst we acknowledge the limitation of having symptomatic subjects in the initial study base of consecutive mammography examinations, we have reported data separately in asymptomatic women and have excluded women with symptoms or clinical abnormalities from our estimates for incremental US-detection. However, since this adjustment only affects the numerator and not the denominator of the calculated incremental cancer detection rates (ICDR), we might have slightly underestimated these rates (i.e. incremental US-detection may be higher). Potential underestimation of ICDR is probably small, as the majority (estimated at 80%) of the self-referred population was asymptomatic. In addition, it should be noted that at least in the context of many European breast services (and also in North America and Australia), self-referral to mammography for screening outside of the recommended age-range (on the basis of perceived risk) is common. In our data, there were 2 cases of cancers detected in asymptomatic women younger than 40, both were 37 years of age and were requesting screening on the basis of family history or perceived increased risk.

It could be argued that pathological stage of mammographically detected cancers is more advanced as to size (pT2+ = 23.5%) and nodal positivity (pN1 = 31.5%) compared to some series of screen detected cancers. This may be related to (a) the generally younger age distribution of the studied population which justifies a lower sensitivity of mammography for early cancer compared to screening in the 50–69 years age range, with reduced diagnostic anticipation due to a shorter detection lead time, and to (b) the fact that a substantial proportion of women were at their first screening mammography, and screen detected cancer stage is known to be

Table 4 – Summary of published studies on US in breast screening

Author (year)	Setting and population (mean age)	No. of US-screened women	US-only cancers (%)	US-generated benign surgical biopsies (%)
Gordon (1995)[7]	Canada, 1989–1994 (age not reported); subjects with clinical/mammography evident lesions, dense breast	12,706	44 (0.34)	235 (1.8)
Kolb (2002)[8]	USA, 1995–2000 (54 years); asymptomatic, self-referred, mammography-negative, dense (D2–4)	12,193	33 (0.27)	287 (2.3)[a]
Buchberger (1999)[9]	Austria, 1996–1998 (49 years); asymptomatic, self-referred, mammography-negative, dense (D2–4)	6113	21 (0.34)	192 (3.1)
Buchberger (2000)[10]	Austria, 1996–2000 (49 years); asymptomatic, self-referred, mammography-negative, dense (D2–4)	8103	32 (0.39)	202 (2.4)
Crystal (2003)[4]	Israel, 2000–2001 (52 years); asymptomatic, self-referred, mammography-negative, dense (D2–4)	1517	7 (0.46)	Not reported
Kaplan (2001)[11]	USA, 1998–2000 (range 35–97 years); self-referred, mammography-negative, dense (D3–4)	1862	6 (0.32)	51 (2.7)
Present study	Italy, 2000–2007 (52 years); self-referred, mammography negative, dense (D3–4)	9157	37 (0.40)	83 (0.9)

a Separate data for needle and surgical biopsies not given.

more advanced at prevalent compared to incident screening. The PPV of assessment was quite high (36.3%), which might suggest that the benefit of US screening was possibly overestimated in this study, as mammography interpretation went for specificity as much as sensitivity. However, one must keep in mind that the study was screening a population with high density categories, and our understanding of the outcome of assessment in such populations is evolving. The review of mammograms in US-only detected cancer cases, with one-fourth of cases (8/33) being retrospectively classified as having abnormalities warranting recall, raises the possibility of sub-optimal mammography sensitivity but should be placed in the context of a consensus review process.

We have already discussed the surgical biopsy rate caused by false positive US findings in this study (0.9%), being substantially higher than currently reported in mammography screening, but lower than other studies of US screening of dense breasts – it would be reasonable to say that this is modest (and may be acceptable) relative to the incremental number of cancers detected due to US. The lower surgical biopsy rate compared to other studies might be explained by a combination of the following factors: well-established ultrasound experience of the clinicians and radiologists, the adoption of a less defensive and more specific approach to assessment, and intensive but judicious use of fine needle biopsy as a determinant of surgical biopsy. Furthermore, we expect that surgical intervention will be reduced as we increasingly use core needle biopsy to resolve equivocal or non-diagnostic outcomes on fine needle biopsy (and some of our surgical biopsies were due to non-adoption of this practise). We expect that consistent adoption of core biopsy in this scenario might reduce surgical intervention to about half of the rates we have reported, considering that core biopsy has almost no 'inadequates' and a very high negative predictive value.[16]

Due to the use of adjunct US, 4.9% (449/9157) of women with mammography-negative dense breasts had additional tests to allow the detection of breast cancer in 0.40%. Since false positives are a major concern in US screening, we have reported on all additional tests and associated cost estimates (see Table 3). Clearly, the cost per additional cancer detected with US only is considerable and higher than the standard cost of detecting one cancer in routine screening (approximately 5000 €), and we are not advocating routine use of adjunct US to screen asymptomatic women. It will be up to individual services and screening programs to decide on the potential value of adopting this approach with consideration of all the evidence we have presented. This should factor the fact that the use of adjunct screening (irrespective of whether it is US or any other test) is essentially searching for the 'difficult' cancers that mammography screening has missed in a group with a (reduced) residual cancer prevalence since mammography has already removed a substantial proportion of cancers.

As indicated earlier, our findings on additional cancer detection in women with dense breasts cannot quantify potential benefit in terms of reducing mortality – to do so would require randomised trials with sufficient follow-up.[18] We have, however, presented three areas of evidence relevant to this issue that may guide consideration (and future research) on the potential benefit of US in adjunct breast screening. First, data on stage distribution of the US-detected cancers are consistent with early-stage cancers (and a possible shift towards an earlier stage relative to mammography detected cancers). Second, evidence from our review of the negative mammograms of women in whom US detected additional cancers – the cancers identified retrospectively were all classified as asymmetric density or non-specific density. Whilst the review process was largely conducted to allow judgement on the generalisability of 'mammography negativity' in our setting, it has provided important information in that these are the type of imaging lesions most often missed with mammography screening, and subsequently emerging as interval cancers. Finally, the largest relative contribution of including US in screening is shown for women aged 40–49 years, the group in whom the largest proportional incidence of interval breast cancers is reported.[19] In combination, such evidence suggests that additional cancer detection with US in asymptomatic women (with mammography-negative dense breasts) has the potential to translate into long-term benefit.

5. Conclusion

US detects cancers in about 0.40% of women with mammography-negative dense breasts, with a higher contribution in women younger than 50 years, but at considerable financial costs. Additional US-detected cancers were smaller than those identified with mammography, and were less frequently associated with node metastases, a pattern consistent with early detection (using US) of cancers otherwise missed by mammography. Our data suggest a maximum relative detection in women younger than 50 years in whom only US found about half of the cancers in that population. Future research in this area might use a randomised design to screen (with versus without US) women with dense breasts aged 40–49 years to identify if this impacts 1–2 year interval cancer rates as a short-term outcome. Whilst our reported incremental cancer detection was associated with additional tests and a considerable estimated cost, the additional false positive surgical biopsy rate attributed to US was lower than other studies and may be acceptable considering women in the highest breast density categories are at increased risk of breast cancer, and are more likely to have an interval cancer in mammography screening.

Conflict of interest statement

None declared.

Acknowledgements

The authors thank the following colleagues for their valuable contribution to implementation of this study: Antonino Amico, Piermaria Boni, Anna Maria Paris and Ileana Tudor.

REFERENCES

1. Leconte I, Feger C, Galant C, et al. Mammography and subsequent whole-breast sonography of nonpalpable breast cancers: the importance of radiologic breast density. *Am J Roentgenol* 2003;**180**:1675–9.
2. Mandelson MT, Oestreicher N, Porter PL, et al. Breast density as a predictor of mammographic detection: comparison of interval and screen detected cancers. *J Natl Cancer Inst* 2000;**92**:1081–7.
3. van Gils CH, Otten JD, Verbeek AL, Hendriks JH, Holland R. Effect of mammographic breast density on breast cancer screening performance: a study in Nijmegen, The Netherlands. *J Epidemiol Commun Health* 1998;**52**:267–71.
4. Crystal P, Strano SD, Shcharynski S, Koretz MJ. Using sonography to screen women with mammographically dense breasts. *Am J Roentgenol* 2003;**181**:177–82.
5. Ciatto S, Visioli C, Paci E, Zappa M. Breast density as a determinant of interval cancer at mammographic screening. *Br J Cancer* 2004;**90**:393–6.
6. 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.
7. Gordon PB, Goldenberg SL. Malignant breast masses detected only by ultrasound. A retrospective review. *Cancer* 1995;**76**:626–30.
8. Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of screening mammography, physical examination and breast US and evaluation of factors that influence them; an analysis of 27,825 patient evaluations. *Radiology* 2002;**225**:165–75.
9. Buchberger W, DeKoekkoek-Doll P, Springer P, Obrist P, Dunser M. Incidental findings on sonography of the breast: clinical significance and diagnostic workup. *Am J Roentgenol* 1999;**173**:921–7.
10. Buchberger W, Niehoff A, Obrist P, DeKoekkoek-Doll P, Dunser M. Clinically and mammographically occult breast lesions: detection and classification with high resolution sonography. *Semin Ultrasound CT MR* 2000;**21**:325–36.
11. Kaplan SS. Clinical utility of bilateral whole breast US in the evaluation of women with dense breast tissue. *Radiology* 2001;**221**:641–9.
12. Berg WA. Rationale for a trial of screening breast ultrasound: American College of Radiology Imaging Network (ACRIN) 6666. *Am J Roentgenol* 2003;**180**:1225–8.
13. Irwig L, Houssami N, van Vliet C. New technologies in screening for breast cancer: a systematic review of their accuracy. *Br J Cancer* 2004;**90**:2118–22.
14. American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADS™). 3rd ed. Reston (VA): © American College of Radiology; 1998.
15. Houssami N, Irwig L. Likelihood ratios for clinical examination, mammography, ultrasound and fine needle biopsy in women with breast problems. *Breast* 1998;**7**:85–9.
16. Ciatto S, Houssami N, Ambrogetti D, et al. Accuracy and underestimation of malignancy of breast core needle biopsy: the Florence experience of over 4,000 consecutive biopsies. *Breast Cancer Res Treat* 2007;**101**:291–7.
17. Ciatto S, Catarzi S, Lamberini MP, et al. Interval breast cancers in screening: The effect of mammography review method on classification. *Breast* 2007;**16**:646–52.
18. Irwig L, Houssami N, Armstrong B, Glasziou P. How should we evaluate new screening tests for breast cancer? *Br Med J* 2006;**332**:678–9.
19. Taylor R, Supramaniam R, Rickard M, et al. Interval breast cancers in New South Wales, Australia, and comparisons with trials and other mammographic screening programmes. *J Med Screen* 2002;**9**:20–5.