SHORT COMMUNICATION

Comparison of the clinical performance of three digital mammography systems in a breast cancer screening programme

¹E KEAVEY, MSc, ²N PHELAN, MSc, ¹A M O'CONNELL, MB, FFR(RCSI), ²F FLANAGAN, MD, FRCR, ³A O'DOHERTY, MRCPI, FRCR, ¹A LARKE, MB, FFR(RCSI) and ⁴A M CONNORS, MB, FFR(RCSI)

¹BreastCheck, National Cancer Screening Service, Western Unit, Galway, Ireland, ²BreastCheck, National Cancer Screening Service, Eccles Unit, Dublin, Ireland, ³BreastCheck, National Cancer Screening Service, Merrion Unit, Dublin, Ireland, and ⁴BreastCheck, National Cancer Screening Service, Southern Unit, Cork, Ireland

ABSTRACT. This study compares the clinical performance of three digital mammography system types in a breast cancer screening programme. 28 digital mammography systems from three different vendors were included in the study. The retrospective analysis included 238 182 screening examinations of females aged between 50 and 64 years over a 3-year period. All images were double read and assigned a result according to a 5-point rating scale to indicate the probability of cancer. Females with a positive result were recalled for further assessment imaging and biopsy if necessary. Clinical performance in terms of cancer detection rate was analysed and the results presented. No statistically significant difference was found between the three different mammography systems in a population-based screening programme, in terms of the overall cancer detection rate or in the detection of invasive cancer and ductal carcinoma *in situ*. This was shown in both prevalent and subsequent screening examination categories. The results demonstrate comparable cancer detection performance for the three imaging system types operational in the screening programme.

Received 23 November 2010
Revised 14 September 2011
Accepted 19 September 2011

DOI: 10.1259/bjr/29747759

© 2012 The British Institute of Radiology

In recent years there has been a continuing adoption of digital mammography for breast imaging, and a number of comparative studies of screen film mammography (SFM) and full-field digital mammography (FFDM) have been published as screening programmes and practices make the transition to digital imaging [1–7]. In addition, there have been a number of physical evaluation studies, investigating the technical performance of individual FFDM systems [8–12]. These studies have provided evidence for the benefit of digital mammography in respect of objective image quality as well as clinically, through improved cancer detection.

There is, however, a diverse range of digital mammography systems and technologies in the market, and these systems utilise a variety of different X-ray spectra and significantly different detector systems, as well as different approaches to optimisation of image quality and dose.

BreastCheck, the Irish national breast screening programme, offers two-view mammography every 2 years to females aged between 50 and 64 years and now has 3 years' experience of imaging exclusively with digital mammography. Screening is provided nationally through a network of 4 static screening centres and 16 mobile screening units, all of which are subject to centralised

quality assurance in order to ensure optimal imaging performance consistent with the requirements of the European guidelines for quality assurance in breast cancer screening and diagnosis [13]. The installed base of mammography systems at the BreastCheck programme represents three distinctive technologies and thus offers the opportunity to analyse the comparative performance, in terms of cancer detection, of the digital mammography systems used.

The purpose of this research was to carry out a retrospective comparison of the clinical performance of three FFDM systems utilised within a population-based screening programme.

Methods and materials

During a 36-month period from April 2007 to April 2010, 238 182 digital mammography screening examinations were performed in the screening programme. Females were assigned to an available mammography system at time of screening, and there was no preferential selection by system type or by prevalent or subsequent screening round. The screening programme operates across four regions, each of which includes a static centre and four mobile screening units. Each region uses two of the three mammography system types. Consent was obtained from females participating in the screening programme for the collection, storage and analysis of

Address correspondence to: Ms Elizabeth Keavey, BreastCheck, National Cancer Screening Service, Western Unit, Newcastle Rd, Galway, Ireland. E-mail: elizabeth.keavey@nbsp.ie

their screening records for audit and quality assurance purposes.

The screening examinations were performed on one of three different FFDM units; GE Essential (GE Medical Systems, Buc, France), Hologic® Selenia (Hologic, Bedford, MA) and Sectra MDM L30 (Sectra, Solna, Sweden). The three systems differ in respect of the X-ray spectra used through choice of X-ray tube target/filter combination, as well as the digital detector technology, the behaviour of the automatic exposure control (AEC) system and the final image processing and presentation.

The GE Essential system utilises a dual track (molybdenum and rhodium) X-ray tube with two filter materials (molybdenum and rhodium) and an indirect conversion detector, which consists of a flat panel amorphous silicon detector with a caesium iodide (CsI) optimised needle structure scintillator. The image matrix is 2394×3062 with a pixel size of 100 μm for a 24×31 cm field of view. The GE Essential system can operate in three different automatic exposure control modes. The systems working in BreastCheck are all set to operate in the same standard mode, which involves the automatic selection of anode/filter, kV and mA.

The Hologic Selenia consists of an X-ray tube with tungsten anode material and rhodium or silver filter materials, and uses an amorphous selenium direct conversion flat panel detector. The image matrix is 3328×4096 pixels with a pixel size of 70 μm for a 24×29 cm active image area. The system offers four modes of automatic exposure control and four distinct pre-set dose levels. Systems working at BreastCheck are all set to operate in the auto-filter mode (control of exposure duration and optimum selection of kV and filter) and standard screening dose setting. A proportion of examinations at the beginning of the study period were performed on two Hologic Selenia systems with a molybdenum anode tube and molybdenum or rhodium filter. These systems were upgraded to tungsten tube systems during the study period.

The Sectra MDM L30 is a scanning multislit FFDM system based on a tungsten anode X-ray tube with an aluminium filter. It employs a photon-counting detector comprising a linear silicon sensor in an edge-on geometry connected to photon-counting electronics. The pixel size is 50 μm and the field of view is 24×26 cm corresponding to a pixel matrix of 4800×5200 pixels.

For this survey, a total of 11 GE Essential, 10 Hologic Selenia and 7 Sectra MDM L30 systems were investigated. The systems were subject to routine quality assurance during the period of the survey, in accordance with the recommendations of the European guidelines [13].

12 specialist breast radiologists reading an average of 10 000 examinations per year participated in image interpretation. All images were stored in a centralised picture archiving and communication system (PACS) and read on dedicated mammography review workstations (IDS5-MX, Sectra). Each reading workstation included two high-resolution (5 megapixel) monitors. Initially, all four views were presented on two monitors. The images were then presented at full resolution with one medio-lateral oblique view on each monitor followed by one craniocaudal view on each monitor. Any further manipulation such as zooming or windowing was performed at the discretion of the interpreting radiologist. The review

workstation was located in a dedicated reading room under controlled lighting conditions.

All mammograms were double read by two consultant radiologists specialising in breast imaging. The second reader was not blind to the first reader's results. Previous films or digital images were available at the time of reading. Each mammogram was assigned a radiology result on a five-point scale (R1–R5) to indicate the probability of cancer [1, 14]. The R classification system is a rating scale used to define the probability of cancer. The categories were defined as follows: R1, normal; R2, benign finding; R3, intermediate finding requiring further work-up; R4, likely malignant finding; and R5, highly suspicious for malignancy. The classification system is similar to BI-RADS [15], with the main difference being that all patients with R3 findings are recalled for further assessment, whereas a diagnosis of BI-RADS Category 3 usually prompts short-interval follow-up [9]. 6 month recall is not practised in our screening programme.

For each case, if both readers assigned a category of R1 or R2, the female was listed for routine recall for screening in 2 years. If both readers assigned a category of R3–R5, the female was automatically recalled for further assessment work-up.

If there was a discrepancy in result category—that is, if one reader assigned a category R1 or R2 but the other reader assigned a category of R3, R4 or R5—the patient was listed for discussion at a consensus meeting [14]. All radiologists were invited to attend and a minimum of two was required. All cases with a discrepancy in R category from the previous week were reviewed and a consensus was reached as to whether the patient should be recalled for assessment or listed for routine recall for screening.

Further diagnostic workup for all females recalled to assessment was performed at their regional screening centre. Diagnostic work-up involved the acquisition of additional mammographic views (e.g. spot compression, true lateral and magnification views). Ultrasound examination was performed for all females recalled to assessment. Image-guided biopsy with 14 gauge needle biopsy device or 11 gauge vacuum-assisted needle biopsy device was used for lesions requiring tissue sampling. Vacuum-assisted biopsy was used for all cases with microcalcifications. On-site specimen radiographs were obtained to confirm adequate sampling of microcalcifications.

All samples were evaluated by a dedicated breast pathologist. All lesions biopsied were discussed at a multidisciplinary meeting within one week of biopsy [1, 14]. The multidisciplinary meeting was attended by breast surgeons, breast radiologists, breast pathologists, and medical and radiation oncologists, as well as nursing staff and radiographers.

The four regional centres work within a national quality assurance framework, which includes annual programme multidisciplinary meetings and radiology intercentre clinical audit.

Statistical analysis

Clinical outcome data were collected retrospectively, and subdivided by imaging system used, screening category and the number of cancers detected. The cancer

Figure 1. Flow chart representing the study sample.

detection rate, invasive cancer rate and ductal carcinoma *in situ* (DCIS) rate were compared for each of the three FFDM systems within the categories of prevalent and subsequent screening. The cancer detection rate was defined as the number of cancers detected per 1000 females screened.

Statistical analysis was performed using the statistical software program SigmaStat 3.0 (Systat Software, San Jose, CA). A χ^2 test was used to compare cancer detection performance in terms of cancer detection rate, invasive cancer rate and DCIS rate in initial and subsequent examination categories. A p -value of <0.05 was considered to indicate a significant difference in these comparisons.

In order to provide a better appreciation of the uncertainty surrounding the measurements, we also calculated the 95% confidence interval (95% CI) for each measurement according to the standard methodology of confidence limits for proportions as outlined in NHSBSP Report 59 [16].

Results

28 FFDM systems were included in this study (11 GE Essential, 10 Hologic Selenia and 7 Sectra MDM L30). A

total of 238 182 screening examinations were included in the study. Of the total number of examinations, 119 556 (50.2%) were prevalent screening examinations and 118 626 (49.8%) subsequent screening examinations. Of the number of females screened, 104 529 (43.9%) were imaged using a GE Essential system, 83 143 (34.9%) using a Hologic Selenia system and 50 510 (21.2%) using a Sectra MDM L30 system. In more detail, of the 119 556 prevalent screening examinations, 70 027 (58.6%) were imaged using a GE Essential system, 31 601 (26.4%) using a Hologic Selenia system and 17 928 (15.0%) using a Sectra MDM L30 system. Of the 118 626 subsequent screening examinations, 34 502 (29.0%) were imaged using a GE Essential system, 51 542 (43.5%) using a Hologic Selenia system and 32 582 (27.5%) using a Sectra MDM L30 system as shown in Figure 1.

A total of 1632 cancers were diagnosed (of the 238 182 females screened), which represents a cancer detection rate of 6.9 per 1000. The rate was higher for prevalent screening at 8.0 per 1000 than subsequent screening at 5.7 per 1000. Table 1 presents the key data collected.

Comparative data published previously from the BreastCheck programme compared cancer detection rates of FFDM to SFM between 2005 and 2007. The overall cancer detection rate for SFM was found to be 7.0

Table 1. Cancer detection rates for three full-field digital mammography system types

Performance result	Prevalent screening (n=119 556)			p-value
	GE (n=70 027)	Hologic (n=31 601)	Sectra (n=17 928)	
Cancer detection rate ^a	8.1	7.2	8.9	0.13
95% CI ^b	7.4–8.8	6.3–8.1	7.5–10.3	
Invasive cancer detection rate ^a	6.5	5.8	6.6	0.38
95% CI ^b	5.9–7.1	5.0–6.6	5.4–7.8	
DCIS detection rate ^a	1.6	1.4	2.2	0.10
95% CI ^b	1.3–1.9	1.0–1.8	1.5–2.9	
	Subsequent screening (n=118 626)			p-value
	GE (n=34 502)	Hologic (n=51 542)	Sectra (n=32 582)	
Cancer detection rate ^a	5.3	5.7	6.2	0.30
95% CI ^b	4.5–6.1	5.1–6.4	5.4–7.1	
Invasive cancer detection rate ^a	4.3	4.4	4.6	0.88
95% CI ^b	3.6–5.0	3.8–5.0	3.9–5.8	
DCIS detection rate ^a	0.9	1.2	1.6	0.06
95% CI ^b	0.6–1.2	0.9–1.5	1.2–2.0	

CI, confidence interval; DCIS, ductal carcinoma *in situ*.

^aNumber of cancers detected (expressed per 1000 screening examinations).

^b95% confidence interval calculated according to the methodology outlined in NHSBSP Report 59 (expressed per 1000 screening examinations).

per 1000 ($n=44\,156$) for prevalent screening and 4.4 per 1000 ($n=109\,463$) for subsequent screening [1].

No statistically significant difference was found between the three imaging system types for overall cancer detection, for prevalent screening ($p=0.13$) or subsequent screening ($p=0.30$). The 95% confidence interval for each measurement has been included in Table 1, and these results are also presented graphically in Figure 2.

Invasive cancers accounted for 1288 of the cancers detected (78.9%). 759 of these were detected at prevalent screening and the remainder (529) at subsequent screening. No statistically significant differences were found for the detection of invasive cancer between the three system types for prevalent screening ($p=0.38$) or subsequent screening ($p=0.88$).

Of the 1632 cancers detected, 344 were DCIS (21.1%). 198 of these were found at prevalent screening and the remainder (146) were found at subsequent screening. The three system types showed no statistically significant difference in the detection of DCIS for both prevalent screening ($p=0.10$) and subsequent screening ($p=0.06$).

Discussion

The results reported here suggest that the clinical performance, in terms of cancer detection, of these three mammography systems is equivalent. It is important to note that, although large screening numbers were studied, the number of detected cancers were

relatively lower and were insufficient to prove statistical significance.

The graph of DCIS rate for the subsequent screening group in Figure 2(d) does demonstrate a difference between the Sectra and GE systems at the 95% confidence level. However, it is important to recognise the caveat that there are relatively low numbers of DCIS cases ($n=146$) for all three systems in this group, and also that the standard formula for calculation of confidence intervals is not recommended for proportions outside the range 0.1–0.9 [17].

Also, observation of the graphical data presented in Figure 2 clearly illustrates the difference in cancer detection rate that would be required to show significance within the current confidence intervals for each group. In order to narrow the confidence intervals and achieve statistical significance, we continue to collect more screening outcome data.

Mammography imaging equipment used in the BreastCheck screening programme is subject to rigorous physico-technical quality assurance to ensure conformance with the imaging standards of the European guidelines for quality assurance in breast cancer screening and diagnosis [13]. Despite conformance to guidelines, differences in the objective image quality performance, and more particularly in the breast dose performance of each system, have been observed [8, 9, 11, 12, 18]. It is also the case that different mammography systems apply dissimilar approaches to the processing of images for presentation [10], resulting in perceptual differences in image quality.

Figure 2. Comparison graphs of invasive cancer detection rate and DCIS rate for prevalent and subsequent screening. The error bars represent the 95% confidence interval calculated according to the standard methodology outlined in NHSBSP Report 59. DCIS, ductal carcinoma *in situ*.

Overall, the results provide a reassuring audit measure, suggesting that technical differences evident from routine quality assurance measurements do not appear to be reflected clinically in terms of overall cancer detection, at least within the statistical error in the rates observed in this study.

Acknowledgment

We would like to thank Sinead O'Neill for her help with statistical analysis of the data.

References

1. Hambly NM, McNicholas MM, Phelan N, Hargaden GC, O'Doherty A, Flanagan FL. Comparison of digital mammography and screen-film mammography in breast cancer screening: a review in the Irish breast screening program. *AJR Am J Roentgenol* 2009;193:1010–18.
2. Karssemeijer N, Bluekens AM, Beijerinck D, Deurenberg JJ, Beekman M, et al. Breast cancer screening results 5 years after introduction of digital mammography in a population-based screening program. *Radiology* 2009;253:353–8.
3. Vinnicombe S, Pinto Pereira SM, McCormack V, Shiel S, Perry N, dos Santos Silva IM. Full-field digital versus screen-film mammography: comparison within the UK breast screening program and systematic review of published data. *Radiology* 2009;251:347–58.
4. Skaane P, Young K, Skjennald A. Population-based mammography screening: comparison of screen-film and full-field digital mammography with soft-copy reading—Oslo I study. *Radiology* 2003;229:877–84.
5. Skaane P, Skjennald A. Screen-film mammography versus full-field digital mammography with soft-copy reading: randomized trial in a population-based screening program—the Oslo II study. *Radiology* 2004;232:197–204.
6. Skaane P, Hofvind S, Skjennald A. Randomized trial of screen-film versus full-field digital mammography with soft-copy reading in population-based screening program: follow-up and final results of Oslo II study. *Radiology* 2007;244:708–717.
7. Pisano E, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med* 2005;353:1773–83.
8. Lazzari B, Belli G, Gori C, Rosselli Del Turco M. Physical characteristics of five clinical systems for digital mammography. *Med Phys* 2007;34:2730–43.
9. Baldelli P, Phelan N, Egan G. A novel method for contrast-to-noise ratio (CNR) evaluation of digital mammography detectors. *Eur Radiol* 2009;19:2275–85.
10. Zanca F, Jacobs J, Van Ongeval C, Claus F, Celis V, Geniets C, et al. Evaluation of clinical image processing algorithms used in digital mammography. *Med Phys* 2009;36:765–75.
11. Baldelli P, Phelan N, Egan G. Investigation of the effect of anode/filter materials on the dose and image quality of a digital mammography system based on an amorphous selenium flat panel detector. *Br J Radiol* 2010;83:290–5.
12. Hendrick RE, Pisano ED, Averbukh A, Moran C, Berns EA, Yaffe MJ, et al. Comparison of acquisition parameters and breast dose in digital mammography in the American college of radiology imaging network digital mammographic imaging screening trial. *AJR Am J Roentgenol* 2010;194:362–9.
13. Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L, editors. European guidelines for quality assurance in breast cancer screening and diagnosis, 4th edn. Luxembourg City, Luxembourg: European Commission, 2006.
14. Shaw CM, Flanagan FL, Fenlon HM, McNicholas MM. Consensus review of discordant findings maximizes cancer detection rate in double-reader screening mammography: Irish national breast screening program experience. *Radiology* 2009;250:354–62.
15. American College of Radiology. Breast imaging reporting and data system (BI-RADS), 4th edn. Reston, VA: American College of Radiology, 2003.
16. NHSBSP. Report 59: quality assurance guidelines for breast screening radiology, 2nd edn. Sheffield, UK: NHS Cancer Screening Programmes, 2011.
17. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br Med J (Clin Res Ed)* 1986;292:746–50.
18. Baldelli P, McCullagh J, Phelan N, Flanagan F. Comprehensive dose survey of breast screening in Ireland. *Radiat Prot Dosimetry* 2011;145:52–60.