Clinical Radiology (2005) **60**, 623–626

REVIEW

Is radiologists' volume of mammography reading related to accuracy? A critical review of the literature

S.M. Moss*, R.G. Blanks, R.L. Bennett

Cancer Screening Evaluation Unit, Institute of Cancer Research, Sutton, UK

Received 15 November 2004; received in revised form 12 January 2005; accepted 20 January 2005

KEYWORDS
Breast radiography;
Cancer screening;
Radiology reporting

The current UK quality assurance guidelines for radiologists in the NHS breast screening programme require those reporting screening mammograms to read a minimum of 5000 cases per year. We aimed to review the evidence for this and to assess whether there was justification for lowering the required level. A literature search was conducted to identify relevant studies where accuracy of reporting mammograms was related to reading volume. Three of the five studies reviewed suggested a positive association between reading volume and sensitivity, but there were few data on volumes above 5000 cases per year. The available evidence did not provide any basis for reducing the threshold volume. Further work is needed, in a UK or European setting, to study the relationship between reading volume and accuracy at higher volume levels and also the separate effects of reading volume and reading experience.

Introduction

Accuracy of reading of mammograms affects both cancer detection rates and false-positive rates. Current recommendations in England are that radiologists who report screening mammograms should read at least 5000 cases per year,[1] and current European guidelines recommend a similar figure.[2] Elsewhere, radiologists may read far fewer than this; for example, in the USA the Mammography Quality Standards Act of 1992 stipulates a minimum annual reading volume of 480 per annum. On average, radiologists in the UK read 5 to 7 times more cases than those in the USA.[3]

If the current UK minimum reading volume could be reduced, this might result in more radiologists becoming eligible to report screening mammograms and provide greater flexibility for film reading. Following discussions with members of the NHSBSP Radiologists Quality Assurance Committee, we reviewed the available literature on the effect of volume of film reading on radiologists' performance.

In the UK, radiologists involved in breast cancer screening are encouraged to participate in a voluntary self-assessment programme (PERFORMS™)[11], which involves reporting on a test set of cases. Such test sets of necessity include a higher proportion of cases than observed in real life. The majority of studies on this subject have used such test sets.

Methods and results

We used the reference lists of studies already known to us to identify other potentially eligible studies. A literature search using PubMed

* Guarantor and correspondent: S.M. Moss, Cancer Screening Evaluation Unit, Institute of Cancer Research, 15 Cotswold Road, Sutton SM2 5NG, UK. Tel.: +44 208 722 4191; fax: +44 208 770 0802.

E-mail address: sue.moss@icr.ac.uk (S.M. Moss).

0009-9260/$ - see front matter
doi:10.1016/j.crad.2005.01.011

(unrestricted on date or language) identified no new eligible work. We only included studies where the accuracy of reporting of mammograms was related to reading volume. We did not include those that looked at factors such as radiologist experience (e.g., years of reading mammograms) but that did not also look at reading volume; nor did we include papers looking at case volume related to other outcome measures (such as treatment outcomes). Only five studies were identified, of which two were carried out in the USA,[4,5] one in the USA and the UK,[6] one in Canada[7] and one in Italy.[8] The main characteristics of the studies are described in Table 1.

Four of the papers were based on PERFORMS-type data sets, including varying proportions of cancers (11% to 43%). For these reports, accuracy in terms of sensitivity and specificity was assessed against known outcomes (non-cancer cases being confirmed by negative follow-up or biopsy). One of these studies[8] was based on completion of a test set, mostly as part of a training course, and volume of clinical, not screening, mammograms read.

One investigation[7] was based on real-life reporting, and in this study standardized referral and cancer detection rates were compared with overall programme performance.

Assessment of reading volume

All papers included mammograms read per year (or month). The number of years over which this was assessed varied between 1,[4] 3[7] and lifetime reading;[5] for two reports[6,8] the number was not stated.

Only the work that included UK radiologists[6] provided much information on reading volumes of >5000 cases per year, with no subdivision of reading volumes above this figure. In one study, radiologists with reading volumes >9000 per year were specifically excluded[7].

Outcome measures

All four PERFORMS-type investigations calculated sensitivity as the proportion of cancers recommended for referral, and specificity (where calculated) as the percentage of non-cancers not referred.

Two reports[4,6] calculated ROC curves for individual radiologists; both of these calculated area under the ROC curves and sensitivity at specificity $=90\%$ as measures of accuracy. The work from Italy,[8] based on training data, looked at the

Table 1 Studies of reading volume and accuracy: main characteristics

Paper	Year	Selection of films	Number of films (% cancers)	Gold standard	No. of radiologists (volume range) per year	Setting	Results/conclusions (readings per year)
Beam et al.[4]	2003	PERFORMS-type	148 (43%)	Cancer/non-cancer confirmed by biopsy or 2-year follow-up	110 (96% <5000)	USA	ROC curves against recent volume showed no linear relationship
Esserman et al.[6]	2002	PERFORMS 2	60 (22%)	PERFORMS reviewed by 5 experienced film readers. Cancer/non-cancer confirmed by biopsy or 3-year follow-up	UK, 194 (≥600); USA, 19 low-volume (≤1200), 22 medium-volume (1200-3600), 18 high-volume (≥3600)	UK and USA	Sensitivity at a 90% specificity: 0.785 for UK readers (>5000) 0.756 for high-vol USA readers (>3600) 0.702 for med-vol USA readers (2400) 0.648 for low-vol USA readers (<1200)
Ciatto et al.[8]	1999	PERFORMS-type	150 (11%)	Cancer/non-cancer confirmed by biospy/follow-up	117 (500–51,000)	Italy	Readings per year/test passes: <1000, 18%; 1000-2000, 45%; >2000, 58%
Kan et al.[7]	2000	Real life	?	N/A	35 (2000–5199)	British Columbia, Canada	SCDR 0.89 (<2000), 0.96 (2000–2999), 0.99 (3000-3999), 1.07 (4000-5199). SAIR 1.15 (<2000), 0.81 (2000–2999), 0.78 (3000-3999), 1.09 (4000-5199)
Elmore et al.[5]	1998	PERFORMS-type	150 (18%)	123 non-cancer cases confirmed by biopsy, mammogram or both at 3 years	10 (200–5000)	Connecticut and New York	Total lifetime mammograms, but not mammograms/year associated with cancer detection rate

SCDR, age-standardized cancer detection rate; SAIR, age-standardized abnormal interpretation (recall) rate; N/A, not applicable.