

Mammography screening: an incremental cost effectiveness analysis of double versus single reading of mammograms

Jackie Brown, Stirling Bryan, Ruth Warren

Abstract

Objective—To compare mammography reading by one radiologist with independent reading by two radiologists.

Design—An observational non-randomised trial at St Margaret's Hospital, Epping.

Subjects—33 734 consecutive attenders for breast screening in the main trial and a sample of 132 attenders for assessment who provided data on private costs.

Interventions—Three reporting policies were compared: single reading, consensus double reading, and non-consensus double reading.

Main outcome measures—Numbers of cancers detected, recall rates, screening and assessment costs, and cost effectiveness ratios.

Results—A policy of double reading followed by consensus detected an additional nine cancers per 10 000 women screened (95% confidence interval 5 to 13) compared with single reading. A non-consensus double reading policy detected an additional 10 cancers per 10 000 women screened (95% confidence interval 6 to 14). The difference in numbers of cancers detected between the consensus and non-consensus double reading policies was not significant (95% confidence interval -0.2 to 2.2). The proportion of women recalled for assessment after consensus double reading was significantly lower than after single reading (difference 2.7%; 95% confidence interval 2.4% to 3.0%). The recall rate with the non-consensus policy was significantly higher than with single reading (difference 3.0%; 2.5% to 3.5%). Consensus double reading cost less than single reading (saving £4853 per 10 000 women screened). Non-consensus double reading cost more than single reading (difference £19 259 per 10 000 women screened).

Conclusions—In the screening unit studied a consensus double reading policy was more effective and less costly than a single reading policy.

Introduction

There has been much discussion recently about the unacceptable incidence of interval cancers—that is, cancers diagnosed between screening examinations within the NHS breast screening programme.¹ In addition, variation in the results of breast cancer screening trials has been a subject of debate.² For example, the Swedish two counties trial³ and the Nijmegen study⁴ showed clear reductions in mortality whereas some other studies whose results emerged later did not show the same success.⁵⁻⁷ One factor which may account for the different results is the quality of screening, in terms of both the technology used and the ability of radiologists to interpret the mammograms. To date, published work lacks discussion of the reporting practice within these trials and the implications for the effectiveness of breast screening.

Viewing of screening mammograms by a second reader was not recommended in the Forrest report.⁸ However, double reading is increasingly being practised by screening units within the NHS breast

screening programme as a consequence of reports suggesting that double reading may increase the number of cancers detected by some 9-15%.⁹⁻¹⁰ A second reading may, however, add to the cost of screening.

Under the NHS breast screening programme, if an abnormality is detected in a woman's screening mammograms she is recalled for further assessment by a combination of further mammography, clinical examination, ultrasonography, and cytology, which will lead to a decision to discharge to routine screening or biopsy. Double reading is likely to lead to one reader recommending recall for further assessment and the other not. Consensus may then be reached by discussion between the readers or by review by a senior radiologist. An alternative double reading policy when the first and second readings differ would be to recall all the women recommended for recall. This, however, would add to the cost of a double reading policy. Clarke and Fraser estimated that up to 13% of the health service costs of screening are incurred at the assessment stage.¹¹

We investigated whether a second, independent reading of mammograms by a radiologist is a cost effective means of detecting additional cancers. So far as we know this is the first investigation of the cost effectiveness of double reading of mammograms in a breast screening programme. The cost effectiveness of double reading and consensus and the cost effectiveness of double reading and recalling all women recommended for recall, regardless of whether there was agreement between the first and second readings, was compared with the cost effectiveness of a single reading. The research was undertaken within the breast screening programme at St Margaret's Hospital, Epping.

Methods

We compared three strategies for reporting within a breast screening programme.

Strategy A: single reading—With strategy A mammography reporting is undertaken by one radiologist on a single reading. Recall for assessment is decided on the basis of the one reading.

Strategy B: consensus double reading—With strategy B mammography reporting is undertaken independently by two radiologists. If one recommends recall and the other does not, then either the senior consultant radiologist decides or (if the other reader is available) the final decision for recall is reached by discussion between the two readers.

Strategy C: non-consensus double reading—With strategy C mammography reporting is undertaken independently by two radiologists but the woman is recalled even if recall is recommended by only one.

An incremental economic analysis was undertaken which sought to identify the additional costs and additional effectiveness of changing from strategy A to either strategy B or strategy C. Many mammography screening centres in the United Kingdom use strategy A.

We adopted an observational research design. The study was undertaken in the context of a prevalence screening round of the NHS breast screening pro-

Health Economics
Research Group, Brunel
University, Uxbridge,
Middlesex UB8 3PH
Jackie Brown, research fellow
Stirling Bryan, research
fellow

Breast Screening Service,
St Margaret's Hospital,
Epping CM16 6TN
Ruth Warren, consultant
radiologist

Correspondence to:
Dr Brown.

BMJ 1996;312:809-12

gramme, which routinely invites women aged 50-64. We studied a consecutive series of 33 734 women attending St Margaret's Hospital for breast screening between November 1987 and March 1991. Most of these women (78%) had two view mammography (that is, mediolateral-oblique and craniocaudal views of each breast) and the remainder single mediolateral-oblique views. Six radiologists took part and read first or second at random. The senior radiologist (RW) read all the mammograms and the other radiologists read variable numbers. All the readers were still within their first 10 000 mammograms. The mammography machines were Siemens, CGR, and Philips and the film-screen combination, chemicals, and processors were Kodak.

EFFECTIVENESS

All mammograms were read independently by two radiologists, and each judged whether the woman should be recalled for further assessment. The actual recall decision in the service setting was made according to strategy B. The recall rate for the first radiologist to report provided data for the recall rate under strategy A. Data on the recall rate for strategy C were obtained by assuming that all women would have been recalled had recall been recommended by at least one of the radiologists.

The principal measure of effectiveness was the number of cancers detected. The actual cancers detected by assessment provided data on the number of detections associated with strategy B. Data on the numbers of cancers that would have been detected under strategies A and C were estimated from the actual cancer findings by assessment under the service setting of strategy B and the monitoring of interval cancers. All the women in the study were offered at least one further invitation to screening after three years and cancers were identified through the Thames Cancer Registry. Interval cancer information should therefore be complete. Differences in cancer detection rates were estimated per 10 000 women screened, as this denominator is broadly equivalent to the number of women screened each year by the local programme in Epping.

Differences between strategies in recall rates and cancer detection rates are presented with 95% confidence intervals for observed differences between proportions for paired cases.¹²

COSTS

The objective of the costs analysis was to identify and estimate all additional resources associated with strategies B and C compared with strategy A. A broad perspective was adopted which considered not only the costs incurred by the health service but also those incurred by the women themselves. All costs were standardised to April 1994 prices by means of the health services price index.

Health service costs

The additional health service costs relate to second reading and reporting of the mammograms by the radiologists (strategies B and C) and the consensus process (strategy B). In addition, differences in recall rates among the three strategies have implications in terms of the number of women assessed and thus total assessment costs. The additional health service costs associated with the second reading and reporting of mammograms and with the consensus process were estimated by considering the labour, overheads, and capital items concerned.¹³ Data on the time taken by radiologists to read and report a mammogram were obtained by observing two radiologists who independently reported a total of 1980 images in 678 minutes.

The cost of the consensus process under strategy B

was incurred only when recall was recommended by at least one reader. The proportion of women for whom recall was recommended by at least one reader was taken from the clinical dataset. On the basis of estimates by clinicians in the study an assumption was made that for two thirds of these cases the cost of the final recall decision was equivalent to the cost of a further reading by the senior radiologist. The cost for the remaining third was assumed to be equivalent to the cost of a further reading by both readers.

The procedures required for further assessment after one view screening differ from those after two view screening.¹³ The expected cost of assessment was estimated by taking account of the proportion of women who had one and two view screening and the differential costs of assessment after one and two view screening identified by Bryan *et al.*¹³ Full details of the assessment procedures and costings were given by Bryan *et al.*

Private costs of assessment

Private travel and time costs relate to costs incurred by the women and their companions in attending the assessment clinic. Data were collected on the time incurred and travel arrangements for the clinic visit. These data were obtained by a standardised questionnaire similar to that adopted in other screening evaluation projects.¹³⁻¹⁵ The questionnaires were distributed to a separate sample of 150 consecutive attenders at St Margaret's Hospital during autumn 1992. Private costs were calculated by methods similar to those reported for other screening evaluations.¹⁶ We did not collect data on the sex of companions and assumed that all were men and that all gave up work time.

The additional total cost of strategy B compared with strategy A was estimated as $(1) n \times \{s_1 - (r_A - r_B)(a_1 + a_2)\}$ and the additional total cost of strategy C compared with strategy A was estimated as $(2) n \times \{s_2 - (r_A - r_C)(a_1 + a_2)\}$, where n = the number of women screened; s_1 = the additional health service cost of second reading, reporting, and consensus where necessary; s_2 = the additional health service cost of second reading and reporting; a_1 = the health service cost of assessment; a_2 = the private cost of assessment; r_A = the recall rate under strategy A; r_B = the recall rate under strategy B; and r_C = the recall rate under strategy C.

SENSITIVITY ANALYSIS

The purpose of the sensitivity analysis was to test the robustness of the study results to changes in the most uncertain parameters. The threshold at which the costs of strategies A and B were similar was identified by varying the additional cost of the second reading, reporting, and consensus (s_1) under strategy B.

The effect of reducing the recall rate under strategy A was also investigated. We recognised that each reader in the study knew that a second radiologist would read the film and that all queries would be reviewed. This may have introduced bias towards overestimation of the true recall rate with strategy A. In the sensitivity analysis, therefore, the recall rate with strategy A was assumed to be 4.64%, which was the lowest recall rate reported for the NHS breast screening programme by Chamberlain *et al.*¹⁷

Results

EFFECTIVENESS

Table 1 shows the recall rate with each strategy investigated. The recall rate with strategy B was significantly lower than that with strategy A (table 2). In addition, the recall rate with strategy C was significantly greater than that with strategy B (table 2). The actual number of cancers detected by double

reading and consensus was 269 (strategy B, 80 cancers detected per 10 000 women screened). Only 239 cancers would have been detected had the actual recall decision been based on the recommendation of the first reader (strategy A, 71 cancers detected per 10 000 women screened). A total of 272 cancers would have been detected had all women recommended for recall by either radiologist actually been recalled (strategy C, 81 cancers detected per 10 000 women screened). Table 2 shows that significantly more cancers were detected with strategies B and C than with strategy A. However, the difference in cancers detected between strategies C and B was not significant.

COSTS

Table 1 summarises the parameter values used in equations (1) and (2) to estimate the additional costs of strategies B and C compared with A.

Health service costs—Table 1 shows the additional health service cost associated with a second reading, reporting, and consensus. This was calculated on the basis that two films were required on average for each woman. Table 1 also shows the expected cost of assessment. For a population of 10 000 women screened strategy B was associated with a lower total health service cost than strategy A. Strategy C, on the other hand, had a higher health service cost than strategy A (table 3).

Table 1—Value of parameters used in equations (1) and (2) (see text) to calculate cost of strategies B and C compared with A

Parameter	Value
No of women screened (n)	10 000
Health service costs (£):	
Second reading, reporting, and consensus (s_1)	0.69
Second reading and reporting (s_2)	0.62
Assessment (a_1)	36.23
Private costs (£):	
Assessment (a_2)	43.53
Recall rates (%):	
Strategy A (r_A)	6.9
Strategy B (r_B)	4.2
Strategy C (r_C)	9.9

†Strategy A single reading, strategy B consensus double reading, strategy C non-consensus double reading.

Table 2—Percentage differences in recall rates and number of cancers detected with alternative strategies

Strategy	Percentage difference in recall rate (95% confidence interval)	Difference in cancers detected per 10 000 women screened (95% confidence interval)
B compared with A	-2.7 (-2.4 to -3.0)	9 (5 to 13)
C compared to A	3.0 (2.5 to 3.5)	10 (6 to 14)
C compared to B	5.7 (5.2 to 6.2)	1 (-0.2 to 2.2)

Table 3—The costs and cost effectiveness ratios, per 10 000 women screened, for the base case

Strategy	Difference in cost (£)		Additional cost (£) per additional cancer detected	
	Excluding private costs†	Including private costs†	Excluding private costs†	Including private costs†
B compared with A	-2 882	-4 853	‡	‡
C compared with A	17 069	19 259	1 707	1 926

†Private costs were those incurred by women and their companions.

‡Strategy B was dominant over strategy A—that is, it produced savings and detected more cancers than strategy A.

Private costs of assessment—One hundred and thirty two (88%) of the 150 questionnaires were completed and returned. The average cost incurred by women and their companions in attending the clinic for further assessment was £43.53 (SD £30.29). Average travel costs were £10.74 and the average time costs £32.75. Including private costs therefore increased the cost of assessment to £68.98. Table 3 shows that for a population of 10 000 women screened strategy B was still associated with a lower cost than strategy A when private costs were included. As expected, the inclusion of private costs increased the total cost of strategy C compared with strategy A.

COST EFFECTIVENESS ANALYSIS

Strategy B was dominant over strategy A in that it was more effective and less costly. Strategy C was more effective than strategy A but was also more costly. Hence incremental cost effectiveness ratios were calculated by dividing the additional cost of strategy C by the additional number of cancers detected. These are shown in table 3. As strategy C was not found to be significantly more effective than strategy B and was associated with higher cost, the subsequent analyses were exclusively on the comparison of strategies A and B.

SENSITIVITY ANALYSIS

Focus for the sensitivity analysis was on the effect of parameter changes on the difference in health service costs between strategies B and A and on the incremental cost effectiveness ratios. The results presented exclude private costs. The direction of change was the same whether private costs were included or not. The cost of second reading, reporting, and consensus under strategy B had to be increased from the base case estimate of 69p to 98p (42% increase) before the costs of strategies A and C were the same. Similarly, the cost difference between strategies A and C was sensitive to changes in the additional cost of second reading and reporting under strategy B. A change from the base case estimate of 62p to 91p (47% increase) was required before the costs of strategies A and C were the same.

The greater the difference in recall rates between strategies A and B the more likely were the savings in assessment visits to outweigh the cost of second reading, reporting, and consensus. Hence when the recall rate of strategy A was reduced to 4.64% the savings from the decrease in number of assessment visits were reduced and strategy B became the more expensive option. The additional cost of strategy B was then £5306 and the incremental cost effectiveness ratio was £590 per additional cancer detected.

Discussion

In the screening unit studied a second reading of screening mammograms was more effective in terms of early cancer detection than a single reading. The adoption of strategy B in place of a single reading led to 13% more cancers being detected. These findings accorded with those reported elsewhere.^{9,10} The results indicate no significant difference in numbers of cancers detected between strategies B and C. This may, however, be due to the fact that the sample size was insufficient to detect a significant difference. Information on the size, grade, and nodal status of the screen detected and interval cancers has been published.¹⁸ Warren and Duffy also give information on the effect of reading ability and experience on the recall rates and cancer detection.¹⁸

In economic terms strategy B was dominant over strategy A in that it detected more cancers and had a lower cost. The lower cost of strategy B can be explained by its lower recall rate. The additional costs

Key messages

- Screening mammograms are increasingly being viewed by a second reader, but this is not a requirement in the United Kingdom
- Double reading of screening mammograms detects more cancers than does single reading
- Double reading with consensus reduces recall rates and has a lower total cost than single reading
- Breast screening units should consider adopting consensus double reporting for the first screening examination in order to improve efficiency
- Double reading with consensus is also likely to confer benefits at subsequent screening examinations, though the magnitude and cost effectiveness of these benefits are not known

of double reading and consensus under strategy B were more than offset by savings in assessment costs. Strategy B was associated with a lower cost than strategy C, and the observed difference in effectiveness was not significant.

When considering implementing a double reading policy with consensus it is important also to consider the reduced anxiety in women who would otherwise unnecessarily be recalled for further assessment. That over three quarters of women sampled were accompanied to the assessment clinic may reflect such concerns.

The results were sensitive to an increase in the cost of double reading and reporting. Anderson *et al*, however, estimated that radiologists could read mammograms roughly twice as fast as in this series.⁵ This would suggest even greater savings with strategy B compared with strategy A and thus adds weight to our findings.

It was also recognised in the sensitivity analysis that each reader knew that a second radiologist would be reading the mammograms. This may have introduced bias whereby the single recall rate was overestimated. Thus in the sensitivity analysis the single reader recall rate was reduced to the lowest recall rate in the NHS breast screening programme, as reported by Chamberlain *et al*.¹⁷ Strategy B was no longer dominant over strategy A. The cost effectiveness ratio for strategy B was, however, comparatively small (£590 per additional cancer detected) with these alternative assumptions. A truer comparison, however, could be made within a randomised controlled trial.

The study design did not allow comparison of biopsy rates with the alternative strategies. Thus we do not know whether the number of benign biopsy samples differed with each policy. The findings of Anderson

et al suggested that double reading without consensus may cause a small increase in the number of biopsies, which would have additional cost implications.⁵

In conclusion, consensus double reading of mammograms is effective and cost saving when compared with single reading. This study showed that cancer detection rates were increased by 13% with consensus double reading. The study, however, focused on the prevalence screening round. It supports a policy of consensus double reading rather than single reading at the first screening examination. Consensus double reading is also likely to confer benefits at subsequent screening examinations but the magnitude and hence cost effectiveness of these benefits are not known.

We thank Jo Holland for secretarial support and Stephen Duffy, Mark Sculpher, and Martin Buxton for advice. The views in this paper are ours alone.

Funding: The Health Economics Research Group is supported by a grant awarded by the research and development division of the Department of Health.

Conflict of interest: None.

- 1 Woodman CBJ, Threlfall AG, Boggis CRM, Prior P. Is the three year breast screening interval too long? Occurrence of interval cancers in NHS breast screening programme's North Western region. *BMJ* 1995;310:224-6.
- 2 Day NE. Screening for breast cancer. *Br Med Bull* 1991;47:400-15.
- 3 Tabar L, Fagerburg CJG, Gad A, Baldetorp L, Holmberg LH, Grontoft O, *et al*. Reduction in mortality from breast cancer after mass screening with mammography. *Lancet* 1985;i:829-32.
- 4 Verbeek ALM, Hendricks JHCL, Holland R, Mravunac M, Sturmans F, Day N. Reduction of breast cancer mortality through mass screening with modern mammography. First results of the Nijmegen project, 1975-1981. *Lancet* 1984;i:1222-4.
- 5 Anderson I, Aspegren K, Janzon L, Landberg T, Lindholm K, Linell F, *et al*. Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial. *BMJ* 1988;297:943-4.
- 6 Miller AB. The Canadian national trial. In: Miller AB, ed. *Screening for breast cancer*. Toronto: Hans-Huber, 1988:51-8.
- 7 Wald NJ, Chamberlain J, Hackshaw A. Report of the European Society for Mastology Breast Cancer Screening Evaluation Committee. *The Breast* 1993;2:209-16.
- 8 Working Group. *Breast cancer screening. Report to the health ministers of England, Wales, Scotland and Northern Ireland*. London: HMSO, 1986. (Chairman Sir Patrick Forrest.)
- 9 Anderson ECD, Muir BB, Walsh JS, Kirkpatrick AE. The efficacy of double reading mammograms in breast screening. *Clin Radiol* 1994;49:248-51.
- 10 Thurffell EL, Lemevall KA, Taube AAS. Benefit of independent double reading in a population-based mammography screening program. *Radiology* 1994;191:241-4.
- 11 Clarke PR, Fraser NM. *Economic analysis of screening for breast cancer. Final report for the Scottish Home and Health Department*. Edinburgh: Scottish Home and Health Department, 1990.
- 12 Gardner MJ, Altman DG, eds. *Statistics with confidence. Confidence interval and statistical guidance*. London: BMJ Publishing, 1989.
- 13 Bryan S, Brown J, Warren R. Mammography screening: an incremental cost-effectiveness analysis of two-view versus one-view procedures in London. *J Epidemiol Community Health* 1995;49:70-8.
- 14 Ashby J, Buxton M, Gravelle H. *What costs do women meet for early detection and diagnosis of breast problems?* Uxbridge: Brunel University, 1989. (Health Economics Research Group report No 6.)
- 15 Sculpher MJ, Buxton MJ, Ferguson BA, Spiegelhalter DJ, Kirby AJ. Screening for diabetic retinopathy: a relative cost-effectiveness analysis of alternative modalities and strategies. *Health Econ* 1992;1:39-51.
- 16 Bryan S, Buxton M, McKenna M, Ashton H, Scott A. Private costs associated with abdominal aortic aneurysm screening: the importance of travel and time costs. *Journal of Medical Screening* 1995;2:62-6.
- 17 Chamberlain J, Moss SM, Kirkpatrick AE, Michell M, Johns L. National Health Service breast screening programme results for 1991-2. *BMJ* 1993;307:353-6.
- 18 Warren RML, Duffy SW. Comparison of single reading with double reading of mammograms, and change in effectiveness with experience. *Br J Radiol* 1995;68:958-62.

(Accepted 9 January 1996)

ONE HUNDRED YEARS AGO

THE PREVENTION OF SUNSTROKE.

In the hot weather which we are experiencing a popular knowledge of every precaution which can be adopted to prevent an attack of heat apoplexy is of the highest value. A timely hint from our Australian colonies has now been opportunely reported. It appears that during January, when 300 persons died of sunstroke, a Colonial Government asked the Medical Board to issue appropriate instructions for the avoidance of this grave disease. The Board is stated to have declared that, of all predisposing causes, undue indulgence in intoxicating liquor is the most common and the most dangerous. Further, that

during the attack it is dangerous to employ intoxicants as a remedy. We cordially endorse this opinion. In many cases sunstroke has practically been alcohol stroke, and in other cases an injudicious resort to alcohol therapeutically has endangered the sufferer's life. Even by the abstinent, under extreme heat conditions, it is essential that such commonsense precautions as the wearing of appropriate clothing, of light, non-radiating head-gear, and moderation of exertion should be adopted. Undoubtedly *ceteris paribus*, the strictly abstinent have the least risk of heat apoplexy.

(*BMJ* 1896;i:1520.)