Effectiveness and cost-effectiveness of double reading of mammograms in breast cancer screening: findings of a systematic review

Dinnes J, Moss S, Melia J, Blanks R, Song F, Kleijnen J

Authors' objectives
The authors' stated aim was to compare the screening accuracy, patient outcomes, and costs of double reading versus single reading of mammograms for diagnosing breast cancer. In actuality, the authors focused largely on the cancer detection rates and recall rates.

Searching
The authors searched MEDLINE, CINAHL, DHSS Data, BIOSIS Previews, EMBASE, Cancerlit, databases on BIDS, Pascal, Dissertation Abstracts, the Cochrane Controlled Trials Register, SIGLE, HealthSTAR, Conference Papers Index, EconLit and the reference lists of retrieved papers; the search terms used were listed. Experts in the field were contacted for additional studies.

Study selection
Study designs of evaluations included in the review
Randomised controlled trials, population-based cohort studies and case-control studies of single versus double reading were eligible for inclusion.

Specific interventions included in the review
Studies of double versus single reading of mammograms for incident and prevalent breast cancer were eligible. 'Double reading' occurs when two radiologists interpret each mammogram. Studies of automated methods of mammogram reading were excluded.

Reference standard test against which the new test was compared
The authors did not explicitly specify a 'gold' standard as an inclusion criterion for the review, although they stated that the appropriate gold standard is follow-up to the next screening round with true interval or occult cancers excluded. Longitudinal follow-up was the primary reference standard in the included studies.

Participants included in the review
The authors did not specify any inclusion criteria relating to the participants in the primary studies. A total of 588,456 women participating in population-based national, regional, or pilot breast cancer screening programmes were included in the review. All were aged between 50 and 70 years.

Outcomes assessed in the review
To be eligible for inclusion, the studies had to present sufficient data to allow the recall rate (proportion of women called for further assessment) and cancer detection rate (number of breast cancers detected divided by the total number of women screened) to be calculated for single reading and double reading of the same sample of mammograms. Other planned outcomes in the review included sensitivity (proportion of women with breast cancer recalled for assessment), specificity (proportion of those without breast cancer who are not called for further reassessment) and breast cancer mortality. However, little data were available for these outcomes.

How were decisions on the relevance of primary studies made?
Two reviewers appraised the papers independently. Any discrepancies were resolved by discussion, or by consultation with up to two additional reviewers.

Assessment of study quality
The authors did not state that they assessed validity. Two reviewers appraised the papers independently.
Data extraction
Two reviewers extracted the data independently. Data were extracted on: the geographic location, sample size, participant characteristics, recall policy, screening round, gold standard used, blinding of assessors, number of mammography views, characteristics of mammography, characteristics of mammography readers, recall rate, cancer detection rate, and specificity and sensitivity where available. The authors calculated the recall rate and cancer detection rate from extracted data if these rates were not presented in the original papers.

Positive and negative predictive values, likelihood ratios and diagnostic odds ratios were not presented in this article, but are available in the full report (see Other Publications of Related Interest).

Methods of synthesis
How were the studies combined?
A narrative synthesis of the studies was presented, including data listed from the individual studies.

How were differences between studies investigated?
The authors did not describe a method for assessing differences between the studies.

Results of the review
Seven cohort studies and three retrospective analyses with 588,456 women were included.

Double reading increased the cancer detection rate by between 2.9 and 11.2 per 10,000 women screened. Double reading with unilateral recall increased the number of women recalled by between 38 and 149 per 10,000 women screened. Consensus or arbitration policies decreased recall by between 61 and 269 women screened.

The authors calculated sensitivity and specificity estimates based on the findings from three cohort studies. Double reading increased sensitivity by between 6.4 and 11.6% compared with single reading. Double reading increased specificity with consensus and mixed recall systems by between 0.8 and 2.8%, but specificity decreased by 1.4% with unilateral recall.

Cost information
Four economic evaluations were identified; all were based on cohort studies included in the review. The estimated incremental cost per additional cancer detected by double reading with unilateral recall was between £1,162 and £2,221 in two UK studies (excluding patient costs). A study in France drew a similar estimate of 21,838 French francs. A study based on a mixed recall system in the USA produced an estimate of US$25,523 per extra cancer detected.

Authors’ conclusions
Double reading increases the cancer detection rate by between 3 and 11 per 10,000 women screened. The findings vary depending on the recall system used. Double reading may be most beneficial for the detection of small cancers and where two readers have different levels of experience or specialist strengths.

CRD commentary
This review addressed a well-defined research question. The literature search appeared appropriate, extensive and it included attempts to locate unpublished studies. A number of databases were searched, including those comprising grey literature. It was unclear whether there were any language restrictions.

The methods used in the review were described in detail, including strategies for extracting the data and assessing relevance. There was less detail about the methods used to assess validity. The review was constrained by the quality of the included studies. The authors did not identify any randomised trials comparing single and double reading techniques. Sensitivity and specificity data were available in only three of the ten studies included in the review. The authors acknowledged the heterogeneity of the included studies, but did not discuss potential sources of bias.
The authors argued that double reading of mammograms is routinely undertaken in many countries, although there is little evidence of major benefits over single reading. The data presented in the review appear to support the authors’ conclusions, but there was little evidence on which to base conclusions about specificity and sensitivity.

**Implications of the review for practice and research**

Practice: The authors suggested that double reading of mammographic scans may increase cancer detection rates, especially for small tumours. On the other hand, double reading may also increase the costs and the number of women recalled who do not have breast cancer, especially where unilateral recall systems are used. Double reading may be most effective where radiologists have different skill levels or areas of expertise.

Research: The authors suggested that further research is needed to assess the relative benefit of double reading according to recall policy and number of mammographic views. Further research on patient outcomes is also needed.

**Funding**

UK Department of Health R&D Division.

**Bibliographic details**


**PubMedID**

14965624

**DOI**

10.1054/brst.2001.0350

**Other publications of related interest**


**Indexing Status**

Subject indexing assigned by NLM

**AccessionNumber**

12002000079

**Date bibliographic record published**

30/06/2004

**Date abstract record published**

30/06/2004

**Record Status**

This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.