A systematic review of computer-assisted diagnosis in diagnostic cancer imaging

Eadie LH, Taylor P, Gibson AP

CRD summary
This review concluded that computer aided diagnostics may be beneficial, compared with interpretation by radiologists alone, in specific imaging applications for breast cancer, but that there was no evidence of benefit and some evidence of detriment in other applications (such as lung cancer imaging diagnostics). Limitations in the review methods and analysis mean that these conclusions should be interpreted cautiously.

Authors' objectives
To assess the impact of computer assisted diagnosis in cancer imaging on radiologists’ detection and diagnosis of cancer lesions.

Searching
MEDLINE, EMBASE, SCOPUS, Web of Science, Science Direct, and CINAHL/EBSCO were searched with no language restriction, from January 1990 up to August 2010; search terms were reported. Only published papers were included.

Study selection
Primary studies of cancer diagnosis (of 20 patients or more) that compared the sensitivity and specificity of computer assisted diagnosis systems with equivalent data from radiologist diagnosis of the same images were eligible for inclusion. Retrieved abstracts were searched for the terms 'sensitivity' and/or 'specificity'; abstracts that did not contain these terms were removed. The full text of the selected papers was assessed for inclusion. A small random sample (under 1%) of the rejected papers was checked; no additional included studies were identified. Studies that assessed only more difficult cancer cases, such as ‘interval’ cancers, or those missed by screening were excluded from the analysis.

Included studies were classified as detection or diagnostic. Various imaging modalities were used: X-ray (including mammography); ultrasound; computed tomography (CT, including CT colonography); magnetic resonance imaging (MRI); dermatologic imaging; single photon emission CT; and scintigraphy. Areas of the body imaged included breast, lung, liver, prostate, bone, bowel, skin, and neck. Approximately 60% of the included studies used computer assisted diagnosis systems that evaluated whole body images; the rest focused on ‘regions of interest’. In most studies, images were initially assessed by radiologists with either the computer assisted diagnosis system alone or the radiologists with the assistance of the computer assisted diagnosis system, and then used to determine a comparative diagnosis. Most studies used the results of biopsy as the reference standard. Studies that used initial diagnosis by the same radiologists as the reference standard were excluded from the analyses.

The authors did not state how many reviewers performed the study selection.

Assessment of study quality
The methodological quality of included studies was assessed using a modified version of the QUADAS tool. Two items on the reporting of uninterpretable test results and reporting of withdrawals from the study were considered not relevant to this review and were omitted. Overall QUADAS scores (maximum 12) were calculated.

The authors did not state how many reviewers performed the quality assessment.

Data extraction
Data were extracted on the sensitivity and specificity of image interpretation by radiologists with and without computer assisted diagnosis (for detection or diagnosis) overall and for various clinical applications.

The authors did not state how many reviewers performed the data extraction.

Methods of synthesis
Separate mean and standard deviation (SD) estimates of sensitivity and specificity, weighted by sample size, were calculated for radiologist interpretation with computer assisted diagnosis for overall detection, overall diagnosis, and for diagnosis stratified by clinical application (mammogram, breast ultrasound, mammogram plus breast ultrasound, lung CT, bowel CT colonography, dermatologic imaging). Corresponding sensitivity and specificity estimates were calculated for interpretation of images by the radiologist alone.

Summary receiver operating characteristic curves, for the same data sets, were constructed using the Moses and Littenberg model. Pooled estimates of diagnostic odds ratios were calculated using a DerSimonian-Laird random-effects model.

The Levine statistic was used to assess homogeneity of variance. Publication bias was assessed using funnel plots.

Results of the review
Forty-eight studies (248,772 patients) were included in the review. The median QUADAS score for computer assisted detection studies was 11 (range 8.5 to 12); the median score for diagnostic studies was 10 (range 6.5 to 12). Of the 16 detection studies, seven contained patients that 14 of the 32 diagnostic studies included a representative spectrum of patients.

**Overall detection** (16 studies; 224,668 patients): For image interpretation by radiologists with computer assisted diagnosis, the weighted mean sensitivity estimate was 84.02% (SD 1.30); specificity was 87.08% (SD 2.75). For radiologists alone (without computer assisted diagnosis), sensitivity was 80.4% (SD 1.46) and specificity was 90.10% (SD 1.97).

**Overall diagnosis** (32 studies; 24,104 patients): For image interpretation by radiologists with computer assisted diagnosis, the weighted mean sensitivity estimate was 90.66% (SD 4.07); specificity was 88.04% (SD 15.03). For radiologists alone, sensitivity was 82.79% (SD 6.12) and specificity was 83.00% (SD 14.46).

**Mammography** (eight studies; 20,071 patients): For image interpretation by radiologists with computer assisted diagnosis, the weighted mean sensitivity estimate was 90.79% (SD 1.01); specificity was 92.70% (SD 7.48). For radiologists alone, sensitivity was 82.34% (SD 2.48) and specificity was 85.91% (SD 11.70).

**Breast ultrasound** (seven studies; 1,403 patients): For image interpretation by radiologists with computer assisted diagnosis, the weighted mean sensitivity estimate was 94.21% (SD 5.42); specificity was 70.28% (SD 20.63). For radiologists alone, sensitivity was 89.91% (SD 9.80) and specificity was 62.68% (SD 17.72).

**Breast ultrasound plus mammography** (three studies; 901 patients): For image interpretation by radiologists with computer assisted diagnosis, the weighted mean sensitivity estimate was 97.54% (SD 1.96); specificity was 39.85% (SD 12.66). For radiologists alone, sensitivity was 96.92% (SD 3.80) and specificity was 51.65% (SD 10.06).

**Lung CT** (three studies; 97 patients): For image interpretation by radiologists with computer assisted diagnosis, the weighted mean sensitivity estimate was 82.05% (SD 10.69); specificity was 73.93% (SD 14.96). For radiologists alone, sensitivity was 82.38% (SD 9.88) and specificity was 80.44% (SD 18.57).

**Dermatologic imaging** (four studies; 756 patients): For image interpretation by radiologists with computer assisted diagnosis, the weighted mean sensitivity estimate was 89.49% (SD 4.32); specificity was 75.29% (SD 8.56). For radiologists alone, sensitivity was 80.22% (SD 16.06) and specificity was 91.85% (SD 8.86).

Bowel CT colonography results were not reported.

The diagnostic odds ratios were greater for mammography and breast ultrasound with computer assisted diagnosis (p<0.01) than with radiologists alone (p=0.03). Summary receiver operating characteristic curves had overlapping confidence intervals, which indicated that overall differences between imaging interpretation by radiologists alone and by radiologists with computer assisted diagnosis were not statistically significant.

Data were not reported for studies of other imaging modalities and clinical applications, which were not included in the meta-analyses. Mean and standard deviation in diagnostic odds ratio values were also reported for all data sets.
The funnel plot indicated possible publication bias.

**Authors' conclusions**
Diagnostic computer assisted diagnosis may be beneficial in specific imaging applications for breast cancer diagnosis compared with interpretation by radiologists alone. There was no evidence of a beneficial effect in other applications of computer assisted diagnosis; there was some evidence that its use may be detrimental.

**CRD commentary**
The review addressed a broad research question on the additional value provided by computer aided diagnostic systems (compared with unaided radiologists) in cancer diagnosis. Inclusion criteria were defined for population, index test(s) and outcome measures, but the reference standard was not pre-specified. A number of sources were searched for relevant studies, but unpublished studies were excluded; the authors reported some evidence for possible publication bias. It was not clear whether the review process included any measures to minimise error and or bias.

The reporting of the methodological quality of included studies was restricted to a range of overall scores, so it was not possible to adequately assess the reliability of data included in the review. However, the authors did note that fewer than half of the included studies included an appropriate spectrum of participants, so the applicability of their results to routine clinical practice was questionable (as noted by the authors). In addition, the methods used to produce summary estimates of sensitivity and specificity were of questionable validity, particularly given the apparent between study heterogeneity. Reporting of the statistical methods used to compare test performance was limited.

As acknowledged by the authors, methodological limitations in this review mean that the conclusions should be interpreted with caution.

**Implications of the review for practice and research**
**Practice:** The authors stated that their results suggested that computer assisted diagnosis could be useful to radiologists looking for breast cancer on mammograms or ultrasound, but should be avoided by those using CT for lung cancer diagnosis or using dermatologic imaging. The authors also noted that most of the included studies were conducted in experimental rather than clinical practice conditions, (a factor which may have affected radiologist behaviour) and that less than half of the included studies were of patients largely representative of the clinical population, so the reported data did not necessarily provide an accurate picture of the impact that computer assisted diagnosis might have in clinical practice.

**Research:** The authors stated that trials of computer assisted diagnosis systems in clinical settings (screening and diagnostic clinics) were needed to confirm any benefit to clinicians.

**Funding**
Not stated.

**Bibliographic details**

**PubMedID**
21345631

**DOI**
10.1016/j.ejrad.2011.01.098

**Original Paper URL**

**Indexing Status**
Subject indexing assigned by NLM
MeSH
Adult; Aged; Aged, 80 and over; Algorithms; Diagnostic Imaging /methods /statistics & numerical data; Female;
Humans; Image Interpretation, Computer-Assisted /methods; Middle Aged; Reproducibility of Results; Sensitivity and
Specificity; Young Adult

Accession Number
12012003025

Date bibliographic record published
17/04/2012

Date abstract record published
10/10/2012

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.