Accuracy of frozen-section analysis in the diagnosis of ovarian tumors: a systematic quantitative review


CRD summary
This review assessed the accuracy of frozen-section analysis in the diagnosis of ovarian tumours. The authors concluded that the accuracy of frozen-section analysis is high for malignant and benign tumours, but lower in borderline tumours. The conclusions are supported by the data presented, but the primary studies suffered from some methodological limitations.

Authors' objectives
To determine the accuracy of frozen-section analysis in the diagnosis of benign, borderline and malignant ovarian tumours.

Searching
MEDLINE, Cancerlit, LILACS and EMBASE were searched from January 1984 to December 2003; the search terms were reported. No language restrictions were applied. The Cochrane Library and the reference lists of primary studies were also searched for additional relevant studies.

Study selection
Study designs of evaluations included in the review
Diagnostic accuracy studies in which the index test was compared with a reference standard were eligible for inclusion. Based on the grading of methodological quality, diagnostic case-control studies were excluded.

Specific interventions included in the review
Studies of frozen tissue sections were eligible for inclusion. The results of the analysis of frozen tissue sections were classified as benign, borderline or malignant.

Reference standard test against which the new test was compared
Studies that used later histologic analysis of standard paraffin sections and classified each case as benign, borderline or malignant were eligible for inclusion. Studies that only described borderline or malignant ovarian tumours and other kinds of tumour (not ovarian) were excluded.

Participants included in the review
Studies of women treated surgically for ovarian tumours were eligible for inclusion. The mean age of the participants was 49 years (range: 1 to 95).

Outcomes assessed in the review
Studies that reported sufficient data to construct 2x2 tables of test performance were eligible for inclusion. The primary outcome measures were accuracy in diagnosing ovarian lesions as: benign versus borderline and/or malignant; malignant versus benign; borderline versus benign; and borderline versus malignant. The measures of accuracy presented were the sensitivity and specificity, positive and negative likelihood ratios (LRs), and post-test probabilities of disease.

How were decisions on the relevance of primary studies made?
Four reviewers independently assessed studies for inclusion, with reference to a selection criteria checklist. Any disagreements were resolved by consensus or through discussion with a fifth reviewer.

Assessment of study quality
The studies were assessed for the following methodological features: methods of data collection and patient selection, descriptions of frozen section and the histologic reference standard, and the presence of verification bias. The studies were also assessed for methodological quality based on the Oxford Centre for Evidence-Based Medicine's 'Levels of Evidence and Grades of Recommendation' (see Other Publications of Related Interest). Only studies of levels 1 to 3 were included in the review; all other studies were excluded. The authors did not state how the papers were assessed for validity, or how many reviewers performed the validity assessment.

Data extraction
Four reviewers independently extracted data on the prevalence of benign, borderline and malignant ovarian lesions, 2x2 tables of test performance, sensitivities, specificities, LRs and post-test probabilities. Articles in English were assessed by two reviewers, while those published in other languages were assessed independently by two other reviewers following translation. Any disagreements were resolved through consensus. Cases that were deferred because of uncertain frozen-section diagnoses were excluded from the analysis. The sensitivity, specificity, LRs and post-test probability were calculated, along with 95% confidence intervals (CIs), for each included study. If 2x2 tables contained cells with zero values, 0.5 was added to each cell.

Methods of synthesis
How were the studies combined?
Spearman correlation coefficients were calculated for the correlation between sensitivity and specificity. If these were found to be independent, then summary pooled estimates of sensitivity and specificity were calculated using fixed-effect models. Summary LRs were calculated from the pooled estimates of sensitivity and specificity. The authors stated that summary post-test probabilities were calculated by multiplying the pre-test probability by the pooled LR; this was not appropriate. However, based on the data presented, it appears that the authors have in fact used the appropriate method of first transforming the pre-test probability to the pre-test odds of disease, multiplied this by the LR to give the post-test odds of disease, and then transformed this to give the post-test probability of disease.

How were differences between studies investigated?
The heterogeneity of sensitivity and specificity values was assessed using the Cochran test for chi-squared distributions (Q statistic). A sensitivity analysis was carried out to assess whether study quality affected estimates of accuracy. Studies that scored less than 50% on the quality criteria and were sublevel 3 of the grading system were excluded from this analysis.

Results of the review
Fourteen studies (n=3,659) were included in the review.

All studies were non-blind, retrospective and from narrow populations. None of the studies suffered from verification bias, and all provided sufficient descriptions of the index test and reference standard.

There were no significant correlations between sensitivity and specificity for any of the classifications investigated.

Benign versus borderline or malignant ovarian tumours (n=14).

The pooled sensitivity was 99% (95% CI: 98, 99) and the pooled specificity was 88% (95 CI: 86, 90). The pre-test probability of having a benign tumour increased from 71 to 95% (95% CI: 94, 96) in patients with a frozen section classified as benign, and decreased to 2.3% (95% CI: 1.6, 3.4) in those with a frozen section classified as borderline or malignant.

Malignant versus benign ovarian tumours (n=14).

The pooled sensitivity was 94% (95% CI: 92, 95) and the pooled specificity was 99% (95% CI: 98, 100). The pre-test probability of having a malignant tumour increased from 23 to 98% (95% CI: 97, 99) in those with a frozen section classified as malignant, and decreased to 1.6% (95% CI: 1.1, 1.9) in those with a frozen section classified as benign.
Borderline versus benign ovarian tumours (n=12).

The pooled sensitivity was 66% (95% CI: 59, 72) and the pooled specificity was 99% (95% CI: 98, 100%). The pre-test probability of having a borderline tumour increased from 5.5 to 79% (95% CI: 71, 85) in those with a frozen section classified as borderline, and decreased to 1.9% (95% CI: 1.6, 2.3) in those with a frozen section classified as benign.

Borderline versus malignant ovarian tumours (n=12).

The pooled sensitivity was 91% (95% CI: 85, 99) and the pooled specificity was 95% (95% CI: 93, 96). The pre-test probability of having a borderline tumour increased from 5.5 to 51% (95% CI: 42, 60) in those with a frozen section classified as borderline, and decreased to 0.5% (95% CI: 0.2, 0.9) in those with a frozen section classified as malignant.

The sensitivity analysis found that the exclusion of the poorer quality studies only led to very small changes in pooled estimates. The use of a random-effects model instead of a fixed-effect model did not change the results.

Inverted funnel plots showed asymmetry.

Authors' conclusions
The diagnostic accuracy rates for frozen-section analysis are high for malignant and benign tumours, but remain relatively low for borderline tumours.

CRD commentary
This was a well conducted and reported review. The review addressed a focused review question supported by clearly defined inclusion criteria. A reasonable literature search was conducted. However, the searches were limited by use of the MeSH term ‘diagnosis (sensitivity and specificity)’, and this is likely to have resulted in relevant studies being missed. In addition, no attempts were made to locate unpublished studies, so the results may be subject to publication bias; the authors also reported asymmetry in the funnel plots. Details of the review process were reported, although it was unclear whether two or four reviewers carried out the data extraction. This is unlikely to have affected the results of the review as two reviewers independently extracting data is generally considered sufficient to avoid bias in the review process. A detailed quality assessment was carried out, although it would have been helpful to have included details of the grades of recommendation used in the paper, rather than having to refer to the cited reference for further information.

The methods used to synthesise the results were appropriate and the provision of information on the post-test probability of disease, based on transforming the pre-test probability of disease, helps to set the results in the clinical context. The authors' conclusions are supported by the data presented, but should be interpreted with some degree of caution given the methodological limitations of the primary studies.

Implications of the review for practice and research
Practice: The authors stated that frozen-section diagnosis is safe for benign and malignant ovarian tumours but has low sensitivity and post-test probability for borderline tumours.

Research: The authors stated that blinded prospective studies should be performed to investigate the diagnostic accuracy of frozen-section analysis in ovarian tumours with trained gynaecological pathologists.

Bibliographic details

PubMedID
Other publications of related interest

Indexing Status
Subject indexing assigned by NLM

MeSH
Cryopreservation; Diagnosis, Differential; Female; Humans; Observer Variation; Ovarian Neoplasms /diagnosis /pathology; Quality Assurance, Health Care; Sensitivity and Specificity; Specimen Handling

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.