Meta-analysis of Pap test accuracy

Fahey M T, Irwig L, Macaskill P

Authors' objectives
To assess the accuracy of the Papanicolaou (Pap) test by comparing the test results with histology in the diagnosis of cervical cancer.

Searching
MEDLINE was searched from 1984 to 1992, and the reference lists of retrieved articles and relevant journals were handsearched. Researchers in the field were contacted for additional material. Unpublished studies were not included. Studies not published in English were excluded.

Study selection
Study designs of evaluations included in the review
Diagnostic accuracy studies were eligible for inclusion. The median sample size was 127 (interquartile range: 87 to 300).

Specific interventions included in the review
Studies assessing Pap tests were eligible for inclusion (no further details were provided).

Reference standard test against which the new test was compared
The included studies were required to use histology as the reference standard. Two histological thresholds for disease were used in the review: cervical intraepithelial neoplasia grade 1 (CIN1) and the more severe grade 2 (CIN2). The studies had to report data for at least one of these two thresholds to be included in the meta-analysis.

Participants included in the review
The studies had to consider cervical cancer or its precursors to be included. Women who participated in cervical cancer screening programmes, including follow-up of abnormal Pap tests, were included.

Outcomes assessed in the review
The included studies were required to report sufficient data for the calculation of sensitivity and specificity. The outcome measures used in the review were sensitivity, specificity, and the diagnostic odds ratio (DOR).

How were decisions on the relevance of primary studies made?
One reviewer selected the studies for inclusion in the review.

Assessment of study quality
Validity was assessed using the following criteria: clinical use; independence of assessments; technique described; selection of study participants; reporting of point estimates of sensitivity and specificity; reporting of confidence intervals (CIs) for these estimates; definition of test threshold; selection for disease verification; sampling fractions reported; and whether accuracy was estimated independently of the test threshold. One reviewer assessed the validity of the primary studies unblinded.

Data extraction
The authors did not state how the data were extracted for the review, or how many reviewers performed the data extraction. Two-by-two tables were constructed for two thresholds of disease (CIN1 and CIN2) and used to calculate the sensitivity, specificity and the diagnostic odds ratio (DOR) for each study. Where possible, the cytologic threshold was varied and the table giving the smallest variance for the log DOR was used in the meta-analysis.
Methods of synthesis

How were the studies combined?
A summary receiver operating characteristic (sROC) curve regression model was fitted to the data. Sample size was taken into account by weighting the model by the reciprocal of the variance of the DOR. An unweighted regression analysis was also performed. The results from the model were used to combine studies in a sROC curve to illustrate the relationship between sensitivity and specificity. Pooled DORs with 95% CIs were calculated.

How were differences between studies investigated?
Sensitivity and specificity were estimated separately for subgroups with different study characteristics, namely, the clinical use of the Pap test, the independence of assessments, the choice of the histologic threshold, and the year of publication. Multiple linear regression was used to examine the relationship between overall accuracy, as measured by the DOR, and these study characteristics.

Results of the review

Sixty-two studies met the inclusion criteria. Fifty-nine studies were included in the meta-analysis; 3 studies were excluded because the data were not available for the histological thresholds used. Twenty-eight (n=10,274) of the 59 studies evaluated screening and 31 (n=7,147) evaluated follow-up Pap tests.

Of the 62 eligible studies, 82% had potential for verification bias and 63% of the studies did not report on the independence of the assessments.

Estimates of sensitivity and specificity ranged from 11 to 99% and from 14 to 97%, respectively. Specificity in the 90 to 95% range corresponded to sensitivity in the 20 to 35% range on the Pap test sROC curve. The pooled weighted DOR was 5.6 (95% CI: 4.7, 6.7), indicating a low level of accuracy. There was no significant association between sensitivity, specificity, overall accuracy (as assessed by the DOR) and the study characteristics.

Authors' conclusions

The Pap test appears to have been unable to discriminate between histologic diseased and non-diseased women with concurrently high sensitivity and specificity. The findings were limited by the poor quality of the primary studies. The poor reporting of the primary studies limited the ability of this study to demonstrate an association between accuracy and study quality.

CRD commentary

The review addressed a clearly stated research question that was adequately defined by its inclusion criteria. Some relevant studies might have been missed because a limited number of sources were searched and studies published in languages other than English were excluded. One reviewer, who selected the studies for inclusion in the review, raised some concern about the potential for selection bias. In terms of minimising errors, it was not ideal that only one reviewer conducted the quality assessment. However, the quality assessment was thorough in its content and the authors included quality variables in their investigation of differences between the studies. Appropriate statistical methods were used to synthesise the data. The authors' conclusions from the evidence reviewed are likely to be reliable.

Implications of the review for practice and research

Practice: The authors stated that the Pap screening test, as currently used in cytology laboratories, may be useful primarily to rule in, but not to rule out disease (based on data from Australia that suggests laboratories are classifying cytology in the 90 to 95% range of specificity). In addition, effort still needs to be directed at improving Pap test sampling and reading.

Research: The authors stated that future primary studies of Pap and alternative screening tests should pay more attention to methodological standards for the evaluation of diagnostic tests. This applies both to the conduct and reporting of the studies.
Bibliographic details

PubMedID
7702044

Indexing Status
Subject indexing assigned by NLM

MeSH
Bias (Epidemiology); Cross-Sectional Studies; Female; Histological Techniques /standards; Humans; Papanicolaou Test; Research Design; Sensitivity and Specificity; Vaginal Smears /standards

AccessionNumber
11995000964

Date bibliographic record published
29/02/2004

Date abstract record published
29/02/2004

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.