Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review


Authors' objectives
To evaluate the accuracy of conventional and new methods of Papanicolaou (Pap) testing when used to detect cervical cancer and its precursors.

Searching
The databases searched to October 1999 included MEDLINE (from 1966), EMBASE (from 1980), HealthSTAR (from 1975), Cancerlit (from 1983) and CINAHL (from 1983); full details of the search strategy were given. The searches were limited to studies of humans that were reported in English. Newly published relevant journal issues, bibliographies of included studies, and recent systematic reviews were searched manually. Unpublished studies were sought by contacting relevant professional societies and manufacturers of cytological devices.

Study selection
Study designs of evaluations included in the review
No inclusion criteria relating to the study design were specified. All of the included studies appeared to be of a diagnostic cohort design.

Specific interventions included in the review
Studies evaluating methods of Pap testing, including conventional methods (with or without manual rescreening), primary computer screening (AutoPap or PAPNET), computer rescreening (AutoPap or PAPNET), and monolayer slide preparation (ThinPrep) were eligible for inclusion. Other recently developed methods, the AutoCyte PREP system and the AutoCyte SCREEN system (TriPath Imaging), were not evaluated.

Reference standard test against which the new test was compared
The included studies were required to use a concurrent (within 3 months) reference standard (histological examination, colposcopy, or cytology).

Separate criteria were used for studies of new methods: prospective comparison of screening tests, or test and reference standard on the same set of patients or slides; if cytological examination was the reference standard, discordant results from the two study tests must be adjudicated by an independent panel of experienced cytology professionals; at least 50% of patients testing positive for high-grade squamous intra-epithelial lesion (HSIL) must be verified by histological examination or colposcopy.

Participants included in the review
No inclusion criteria relating to participant characteristics were specified. Women undergoing Pap testing for primary screening and those undergoing evaluation for previous cytological abnormality were included.

Outcomes assessed in the review
The included studies were required to report sufficient data for the construction of 2x2 contingency tables. The main outcome measures reported in the review were the sensitivity, specificity and positive and negative likelihood ratios (LRs) of the cytological test. Cytological abnormality was defined by one of three thresholds: atypical squamous cells of undetermined significance (ASCUS); low-grade squamous intra-epithelial lesions (LSIL); and HSIL. 'Cases' were defined as histological diagnosis of cervical intra-epithelial neoplasia (CIN) grades I to III, or carcinoma. Equivalent categories in other classification schemes were also used (Bethesda System, Richart System and Reagan/World Health Organization System).

How were decisions on the relevance of primary studies made?
Two investigators independently screened each study, with any differences of opinion resolved by consensus.
Assessment of study quality
Validity was evaluated using the following criteria: type of reference standard; independence of the assessment; method of verification; sample selection; spectrum of disease or non-disease; publication type; and whether industry related. Nine members of the study's working group developed a numeric quality score (maximum 11 points) by consensus (full details of the process and scoring according to validity criteria were given). The authors did not state how many reviewers performed the validity assessment.

Data extraction
Two reviewers independently completed 2x2 tables for each study. Where available, data were abstracted for four different combinations of cytological and histological thresholds: ASCUS/CIN-I, LSIL/CIN-I, LSIL/CIN-II-III and HSIL/CIN-II-III. Cases that were indeterminate were documented but excluded from the calculation of sensitivity and specificity.

Methods of synthesis
How were the studies combined?
Minimum, maximum and median values of sensitivity, specificity, positive and negative LRs, and prevalence were calculated for four different combinations of cytological and histological thresholds: ASCUS/CIN-I, LSIL/CIN-I, LSIL/CIN-II-III and HSIL/CIN-II-III. The results were summarised in a narrative synthesis.

How were differences between studies investigated?
Sensitivity and specificity values were reported separately for different combinations of test and reference standard, and for studies that identified low-risk patients undergoing screening.

Results of the review
Ninety-four studies of conventional Pap test and 3 studies of monolayer cytology were included.

Most of the studies on thin-layer cytology (ThinPrep), computerised re-screening device (PAPNET) and algorithmic classifier (AutoPap) were excluded from the review for the following reasons: many studies did not apply the new technology and conventions for Pap test prospectively to the same sample of women; almost all studies of thin-layer cytology, computer screening or rescreening failed to verify the disease status of women who had negative tests results on cytological screening tests; and little evidence was available with which to assess the effects of thin-layer cytology, computer screening or rescreening on specificity.

Thin-layer cytology relative to histological examination or conventional Pap smear (3 studies, number of patients not reported).

One study compared ThinPrep diagnosis of LSIL or higher with a histological reference standard of CIN-II-III or higher, and reported a specificity of 57.7% and a sensitivity of 94.2%. Using the same reference, conventional Pap smears reported a specificity of 37.0% and a sensitivity of 84.6%. Most of the negative results were not verified by histological examination. Two studies compared conventional and ThinPrep slides with a split-sample technique using a combination of cytological and histological examination as the reference standard. The sensitivity and specificity could not be calculated directly. ThinPrep and conventional Pap smears were compared. Both studies found higher relative true-positive rates for ThinPrep (rates of 1.13 and 1.19, indicating greater sensitivity) and lower relative false-positive rates (rates of 1.12 and 2.05, indicating lower specificity).

Conventional Pap tests.
Estimates of sensitivity and specificity varied greatly in the individual studies. Most of the studies were conducted in women who were referred for previous cytological abnormalities, had visible cervical lesions, or were at high risk for cervical cancer. Few studies evaluated low-prevalence screening samples. Most of the studies used histological examination as a reference standard, but only 51% obtained verification of all or a random sample of patients with
negative tests. Few studies independently assessed the test and reference standard. Fifteen per cent did not provide adequate information on the spectrum of disease. The sensitivity and specificity at different combinations of test and reference standard thresholds were as follows:

for ASCUS/CIN-I (37 studies), the overall specificity ranged from 17 to 99% and the sensitivity ranged from 18 to 98%;

for LSIL/CIN-I (71 studies), the overall specificity ranged from 9 to 100% and the sensitivity ranged from 17 to 99%;

for LSIL/CIN-II-III (54 studies), the overall specificity ranged from 6 to 99% and the sensitivity ranged from 23 to 100%; and

for HSIL/CIN-II-III (43 studies), the overall specificity ranged from 21 to 100% and the sensitivity ranged from 6 to 100%.

Studies with least biased estimates (12 studies identified low-risk patients undergoing screening).

The quality scores ranged from 6 to 10. The overall sensitivity ranged from 30 to 99% and the specificity ranged from 86 to 100%. For LSIL/CIN-I (9 studies), the overall specificity ranged from 86 to 100% and the sensitivity ranged from 30 to 87%; in 8 studies sensitivity was less than 60%. For LSIL/CIN-II-III, the overall specificity ranged from 91 to 98% and the sensitivity ranged from 44 to 99%.

**Authors' conclusions**

Insufficient high-quality data exist to estimate test operating characteristics of new cytological methods for cervical screening. Future studies of these technologies should apply adequate reference standards. Most studies of the conventional Pap test were severely biased. The best estimates suggested that it is only moderately accurate and does not achieve concurrently high sensitivity and specificity. Cost-effectiveness models of cervical cancer screening should use more conventional estimates of Pap test sensitivity.

**CRD commentary**

The aims and inclusion criteria were clearly stated. The search was restricted to English language articles, thus some relevant studies might have been omitted, although attempts were made to locate unpublished studies. Details were given of the methods used to select the primary studies and develop validity criteria. Results from studies without verification bias were considered separately.

The discussion considered the following limitations in the evidence: bias in many studies of conventional Pap test; studies conducted only in patients with abnormalities, leading to verification bias; studies with various gold standards (histological examination with cone biopsy, hysterectomy, or punch biopsy, and colposcopy), which themselves were subject to inaccuracy; and methodological quality and frequency of histological abnormalities varied greatly between the studies. Details were only provided of 12 studies considered to be without verification bias. No comment was made on potential sources of heterogeneity among these higher quality studies, and no attempt at meta-analytic pooling was made.

The evidence presented supports the authors' conclusions.

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

Practice: The authors stated that studies of Pap screening in low-prevalence samples found high specificity, but sensitivity estimates that were lower than generally believed.

Research: The authors stated that future decision models, cost-effectiveness studies, and health policy decisions should consider in their analyses the lower than generally believed sensitivity estimates found in this review.
Bibliographic details

PubMedID
10819705

Original Paper URL
http://www.annals.org/cgi/content/full/132/10/810

Other publications of related interest

Indexing Status
Subject indexing assigned by NLM

MeSH
Cervix Uteri /cytology /pathology; Diagnosis, Computer-Assisted; Female; Humans; Mass Screening /methods /standards; Papanicolaou Test; Sensitivity and Specificity; Uterine Cervical Neoplasms /diagnosis; Vaginal Smears /methods /standards

AccessionNumber
12000008298

Date bibliographic record published
31/07/2005

Date abstract record published
31/07/2005

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.