Diagnostic accuracy of human papillomavirus testing in primary cervical screening: a systematic review and meta-analysis of non-randomized studies

CRD summary
This review compared different tests for human papillomavirus (HPV) with cytology-based tests for the detection of primary cervical cancer. The authors concluded that HPV tests are more sensitive but less specific than cytology. These conclusions are likely to be unreliable given the limitations in the analysis and substantial differences between the studies.

Authors' objectives
To compare the accuracy of human papillomavirus (HPV) DNA testing and cytological testing for the detection of histologically confirmed cervical intraepithelial neoplasia (CIN) 2 or worse disease in primary cervical cancer screening.

Searching
MEDLINE, EMBASE, CINAHL, LILACS, HTA, the Cochrane CENTRAL Register, and the non-trials databases and specialised register of the Cochrane Gynaecological Cancer Group were searched from 1994 to 2005; the search terms were reported. The references of identified articles and conference proceedings were screened for additional studies. No language restrictions were applied. The search was performed in duplicate.

Study selection
Study designs of evaluations included in the review
Diagnostic accuracy studies in which the two tests of interest were performed concomitantly on the same woman, or randomised trials in which the two tests were compared, were eligible for inclusion.

Specific interventions included in the review
Studies that evaluated HPV testing and cytological screening (Pap smear) were eligible for inclusion.

The HPV tests used in the included studies were: polymerase chain reaction (PCR) using the MY09/MY11, GP5+/GP6+ and/or the pU-1M/pU-2R primers; and hybrid capture techniques, either Hybrid Capture 1 (HC1) and/or Hybrid Capture 2 (HC2). For PCR, the thresholds reported by the study authors were used; a result of >10 pg/mL was considered positive for the HC1 test and a threshold of >1 pg/mL was used for the HC2 test.

The cytological screening tests assessed were conventional cytological smear and liquid-based cytology. Two criteria were used to define a positive Pap smear: a result of ASCUS (atypical squamous cells of undetermined significance) or worse, or a result of LSIL (low-grade squamous intraepithelial lesions) or worse. The 1991 Bethesda Reporting System (see Other Publications of Related Interest) was used for cytologic classification. If other systems were used then the results were converted to the Bethesda system.

Reference standard test against which the new test was compared
Studies in which the reference standard consisted of colposcopy and/or biopsy on all women who tested positive on either index test were eligible for inclusion. For studies in which colposcopy was normal, a histologic result was not required to exclude the disease. If colposcopy was abnormal, then the histologic result was used as the reference standard. The CIN nomenclature was used to describe histologic outcomes. The thresholds of CIN2 or worse and CIN3 or worse were considered.

Participants included in the review
Studies of women aged 18 to 70 years participating in cervical cancer screening programmes and who were not being followed up for previous cytological abnormalities were eligible for inclusion. The women in the included studies were
recruited from the general population either attending routine or opportunistic screening or attending family planning or other clinics, were volunteers from the general population, or were randomly recruited from the general population. In some studies women had not been previously screened.

Outcomes assessed in the review
No inclusion criteria relating to the outcomes were specified. The primary outcomes reported in the review were the sensitivity and specificity.

How were decisions on the relevance of primary studies made?
The authors did not state how the papers were selected for the review, or how many reviewers performed the selection.

Assessment of study quality
The authors did not state that they assessed validity, although some aspects of quality were discussed in the results.

Data extraction
The authors did not state how many reviewers performed the data extraction. Data were extracted as 2x2 tables of test performance for the thresholds defined a priori by the review authors.

The sensitivity, specificity, positive and negative predictive values, test positivity rates and the prevalence of disease (defined as presence of CIN2+) were calculated. Ratios of the sensitivity and specificity of HPV testing to cytology were calculated. The accuracy of both tests combined was also calculated; in this case a positive result was defined as either or both tests positive, and a negative as both tests negative. Authors were contacted to obtain missing data if necessary.

Methods of synthesis
How were the studies combined?
The sensitivity, specificity and the ratios of sensitivity and specificity for the two tests (where both were evaluated in the same women) were pooled. Studies were stratified, for pooling, by index test type and threshold. Random-effects models were used in the presence of significant heterogeneity (p<0.1 for Cochran's Q test). In the absence of heterogeneity fixed-effect models were used, weighted by the inverse variance.

How were differences between studies investigated?
Cochran's Q test was used to statistically assess heterogeneity. Forest plots were used to visually assess heterogeneity. Subgroup analyses on the accuracy of cytology at both thresholds and on HC2 testing in women aged over 30 years were performed.

Results of the review
Twenty-five studies were included. Twenty-three studies were cross-sectional and assessed concomitant testing of cytology and HPV. One study used a longitudinal design with baseline cytology and HPV testing; this study was excluded from the meta-analysis. One study was a randomised controlled trial where only baseline data were presented. The sample sizes ranged from 977 to 36,938 women.

In 12 studies histology interpretation was carried out without knowledge of the results of the screening tests, while in 3 studies the pathologists were unaware of the HPV results. In 1 study the colposcopists were unaware of the screening test results and in a further 2 studies they were blinded only to the HPV results. In 4 studies verification bias was avoided as all women received the reference standard. In 8 studies the reference standard was applied to all test positives and to a random sample of negatives to both tests. In 12 studies the reference standard was only applied to Pap smear or HPV test positives. In 1 study only women who tested positive on Pap smear received the reference standard (this study therefore did not fulfil the inclusion criteria for the review).

Accuracy of tests for detecting CIN2 or worse.
HC1 (3 studies in summary table but only two in individual studies table): sensitivity estimates were 64 and 75% and specificity estimates were 90 and 93%. The pooled sensitivity and pooled specificity were 72% (95% confidence interval, CI: 67, 77) and 93% (95% CI: 91, 94), respectively.

HC2 (15 studies): the sensitivity ranged from 46 to 100% and the specificity from 61 to 95%. The pooled sensitivity and specificity were 90% (95% CI: 86, 94) and 87% (95% CI: 83, 90), respectively.

PCR (6 studies): the sensitivity ranged from 64 to 95% and the specificity from 79 to 99%. The pooled sensitivity and specificity were 81% (95% CI: 70, 92) and 95% (95% CI: 93, 97), respectively.

Cytology based on the ASCUS threshold (18 studies): the sensitivity ranged from 27 to 96% and the specificity from 78 to 99%. The pooled sensitivity and pooled specificity were 73% (95% CI: 64, 82) and 92% (95% CI: 90, 94), respectively.

Cytology based on the LSIL threshold (12 studies): the ranges in sensitivity and specificity were unclear as individual study results were only provided for 2 studies. The pooled sensitivity and pooled specificity were 62% (95% CI: 48, 75) and 96% (95% CI: 95, 97), respectively.

HPV testing and cytology combined at the ASCUS threshold (18 studies): the sensitivity of the combination of HC2 and cytology was greater than HC2 alone (pooled ratio 1.05, 95% CI: 1.05, 1.06), but specificity was lower (pooled ratio 0.95, 95% CI: 0.94, 0.96).

Comparison of HC2 to cytology at the ASCUS threshold: the ratio of sensitivities was 1.25 (95% CI: 1.20, 1.29) and the ratio of specificities was 0.97 (95% CI: 0.96, 0.98).

Comparison of HC2 to cytology at the LSIL threshold: the ratio of sensitivities was 1.38 (95% CI: 1.30, 1.47) and the ratio of specificities was 0.89 (95% CI: 0.87, 0.90).

There was strong evidence of heterogeneity for most outcomes tested (p<0.05). Results were also reported for the reference standard threshold of CIN3 but fewer studies provided data on this. The results for women aged over 30 years appeared similar, but it was difficult to assess this as only pooled data were reported and different studies contributed to the overall analysis and subgroup analyses.

**Authors' conclusions**

HC2 and PCR were more sensitive but less specific than cytology for prevalent CIN2 or worse. The highest sensitivity was achieved by the combination of HC2 and cytology, but this had the lowest specificity.

**CRD commentary**

The objective was clearly stated and supported by defined inclusion criteria. However, the inclusion criteria were somewhat contradictory with the later text of the review in relation to study design and outcomes. In addition, 1 study that did not meet the inclusion criteria was included in the review as the reviewers felt that this was an important study. An extensive literature search was conducted and it is likely that most of the relevant studies were identified. Details of the review process were not reported, so it is not possible to determine whether appropriate steps were taken to minimise bias. A formal quality assessment was not undertaken, although some methodological details were discussed in relation to the results but were not considered in the synthesis of the results.

It is unclear whether the methods used to pool the data were appropriate, as full details were not provided. However, some of the reported CIs for specificity extended beyond 100%, suggesting that appropriate methods might not have been used. There was considerable heterogeneity between the studies and this was not adequately investigated. The pooled measures should therefore be interpreted with extreme caution. The comparisons that the authors made between tests and for subgroup analyses based on these summary measures are unlikely to be reliable.

**Implications of the review for practice and research**
Practice: The authors stated that there is still insufficient evidence to justify the use of HPV testing in primary cervical screening.

Research: The authors stated that further studies are needed to investigate whether the HPV test is less appropriate as a screening test in some countries. They also stated that longitudinal studies are required to assess whether the sensitivity of cytology is improved by repeat screening. Studies of the effect of HPV testing on the incidence of and mortality from cervical cancer are also required; several ongoing studies are addressing this issue. Economic studies that investigate the added cost of HPV testing alone or in combination with cytology for screening are also needed.

**Funding**
European Commission; DWTC/SSTC (Federal Services for Scientific, Cultural and Technical Affairs); Gynaecological Cancer Cochrane Review Collaboration.

**Bibliographic details**

**PubMedID**
17084886

**DOI**
10.1016/j.ygyno.2006.08.053

**Other publications of related interest**

**Indexing Status**
Subject indexing assigned by NLM

**MeSH**
Cervical Intraepithelial Neoplasia /pathology /virology; Female; Humans; Mass Screening /methods; Papillomavirus Infections /complications /diagnosis /virology; Uterine Cervical Neoplasms /diagnosis /pathology /virology; Vaginal Smears

**AccessionNumber**
12007000592

**Date bibliographic record published**
29/02/2008

**Date abstract record published**
29/02/2008

**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.