Triage of women with atypical or low-grade cytological abnormalities of the cervix by HPV testing

Systematic review and meta-analysis*

*Registered by the Cochrane Gynecological Cancer Collaborative Review Group.

(Preliminary report)

Marc ARBYN\textsuperscript{1,2}, Frank BUNTINX\textsuperscript{3}, Marc VAN RANST\textsuperscript{4}

1. European Network for Cervical Cancer Screening
   Co-ordinator Evaluation of New Technologies

2. Federal Cell for Evaluation of the Cancer Screening Programmes.

3. Centre for Evidence Based Medicine, Belgian Branch of the Cochrane Collaboration,
   Department of General Practice, University of Leuven, Belgium.

4. Department of Clinical and Epidemiological Virology, University of Leuven, Belgium.

Institute of Public Health
J. Wytsmanstreet 14
1050 Brussels
Belgium

November 14th 2002
1. Abstract

1.1. Background

Management of equivocal and low-grade lesions, observed in the context of cervical cancer screening, remains until today controversial. Triage with HPV DNA testing and repeat Pap smears are two possible strategies to select women who need further follow-up.

This meta-analysis is registered by the Cochrane Gynaecological Cancer Collaborative Review Group (Oxford, February 2002) and is funded by Europe Against Cancer.

1.2. Methods

Systematic literature research targeted studies published over the last ten years, starting from 1992. Data were extracted from 30 selected references containing results of concomitant cytological and virological testing followed by colposcopically directed biopsy in cases with an index smear showing atypia or low-grade abnormality. Random effect models were used for pooling of accuracy parameters in case of significant inter-study heterogeneity.

We distinguished studies by cytological result of the index case: a) ASCUS (atypical squamous cells of uncertain significance); b) LSIL (low-grade squamous intra-epithelial lesion) and c) ASCUS/LSIL (index cases where the distinction between ASCUS and LSIL could not be separated).

All retrieved studies are assessed individually by checking a standard list of study design issues: description of the study population with inclusion and exclusion criteria, the determination of the age distribution of included subjects; the particular design of the study; description of the index cytological test; description of the enrolment triage test (HPV DNA testing system; possibly concomitant repeat cytology); description of the golden standard; timing of index, enrolment and golden standard testing; the blinding of interpreters for other test results; presentation of the absolute figures and relative accuracy parameters and finally a short discussion.

The accuracy parameters (sensitivity, specificity, predictive values, proportion of test positives and prevalence of disease are subsequently pooled using fixed or random effect models according to the level of inter-study heterogeneity. The variation in accuracy measures in the individual studies and the pooled measure are displayed graphically using forest plots. The contrast in accuracy between repeat cytology and HPV DNA testing is studied by the ratio of sensitivity and specificity of both tests applied in parallel on the same women and relative to the golden standard. The change in accuracy by combining two triage tests in comparison with one single test is assessed by the difference in accuracy. The influence of study characteristics on study outcomes was explored by metaregression. Existence of verification bias was assessed graphically using funnel plots and statistically using the rank correlation test and the asymmetry regression test. Finally the trade-off between sensitivity and specificity is studied graphically by the way of summary ROC (Receiver Operating Characteristic) curve analysis.
1.3. Results

1.3.1. Separate pooling of accuracy parameters

**Triage of ASCUS, outcome CIN2+**

Sensitivity and specificity of repeat cytology at threshold ASCUS+ was 81.8 % (95% confidence interval (CI): 73.5-84.3%) and 57.6% (CI: 49.5-65.7%). Considering all HPV tests yielded a sensitivity of 84.4% (CI: 77.6-91.1%) and a specificity of 72.9% (CI: 62.5-83.3%). The prevalence of CIN2+ ranged from 3 to 36%; its pooled mean was 10.5% (CI: 7.9-13.1%). Restriction to 8 recent studies where the Hybrid Capture-II assay was used for detection of high-risk HPV types yielded a sensitivity of 94.8% (CI: 92.7-96.9%) and a specificity of 67.3% (CI: 58.2-76.4%). The positive and negative predictive values of repeat cytology at cut-off ASCUS+ (CI of PPV: 7.6-15.9%; CI of NPV: 96.0-97.5%) were both significantly lower than values of the Hybrid Capture assay (CI of PPV: 19.2-33.6%; CI of NPV: 98.5-99.4%). The specificity of repeat cytology could be enhanced considerably by definition of higher cytological thresholds: 89.1% (CI: 82.1-96.2%) at LSIL+ and 98.0% (CI: 95.6-100%) at cut-off HSIL+. At this thresholds sensitivity became substantially lower than in modern HPV DNA detection methods.

**Triage of ASCUS, outcome CIN3+**

Using CIN3+ as outcome, pooled sensitivity and specificity of repeat cytology at threshold ASCUS+ were respectively 85.3 % (CI: 78.2-90.8%) and 42.7% (CI: 40.6-44.8%). Studies where the Hybrid Capture-II assay was used for detection of high-risk HPV types yielded a sensitivity of 96.6% (CI: 93.5-99.6%) and a specificity of 56.6% (CI: 41.0-72.2%). The prevalence of CIN3+ ranged from 1.9 to 18.9%; its pooled mean was 6.3% (CI: 1.9-10.7%). The positive (PPV) and negative predictive values (NPV) were both higher for HPV testing, but this observation was only significant for NPV.

**Triage of LSIL, outcome CIN2+**

The sensitivity of repeat cytology at cut-off ASCUS+ (91.6%; 84.4-98.8%) was comparable with that of all HPV detection systems together (87.2%; 79.3-95.1%). The sensitivity of the HC2 assay was 4.1% higher than cytology but this difference was statistically insignificant. The specificity of LSIL triage was low for all tests: 41.6% (CI: 27.4-55.8%) for repeat cytology; 51.4% (CI: 31.2-71.6%) for all HPV test systems combined and only 32.9% (CI: 17.8-48.0%) for the HC2 assay. The low specificity yielded very high test-positivity rates (62% for repeat cytology at ASCUS+; 77% for HC2) limiting the triaging potential in women with LSIL. The pooled prevalence of CIN2+ was 20.6% (CI: 13.4-27.9%).

**Triage of LSIL, outcome CIN3+**

The accuracy of triage in LSIL cases considering CIN3+ is less documented in the literature: 1 study for cytological triage; 5 studies for HPV triage among which 3 studies that described triage with HC2. The sensitivity of repeat cytology at threshold ASCUS+ (90.1%; CI: 82.1-95.4%) was similar to that of all HPV tests combined (90.6%; CI: 80.8-100%). The sensitivity of HC2 was 5.6% higher but this difference was not significant. Again specificities were low: 20.7% (CI: 17.9-23.8%) for cytology at ASCUS+; 32.9% for all HPV tests or HC2 (CI for all HPV tests: 12.1-
53.8%; CI for HC2: 17.8-48.0%). The pooled prevalence of CIN3+ was 16.4% (9.7-23.2%).

**Triage of ASCUS/LSIL**

Studies describing triage in women with a previous smear showing ASCUS or LSIL without further detail are less informative since distinction between both index groups is essential. In general, the sensitivity estimates derived from these studies approximated those observed in the ASCUS triage group while the estimates for specificity were intermediate between those derived from ASCUS and LSIL triage group.

**1.3.2. Ratio of the sensitivity and specificity of HPV DNA testing versus cytology**

In this paragraph we summarise the ratios of the accuracy of HPV detection / accuracy of repeat cytology using ASCUS+ as positivity criterion derived from a selection of the studies that provided data for the two triage methods applied on the same patients.

**Triage of ASCUS**

The sensitivity ratio for the outcome CIN2+ was 1.13 (1.08-1.19) indicating a sensitivity for HPV detection that was on average 13% higher than for repeat cytology. The specificity was significantly higher as well: 1.16 (CI: 1.04-1.29). The sensitivity of the Hybrid Capture 2 test was significantly higher as well: 1.16 (CI: 1.08-1.20). The specificity of HC2 exceeded that of repeat cytology but the difference was not significant (1.05; CI: 0.96-1.15).

**Triage of LSIL**

The sensitivity and specificity ratios did not differ significantly from unity in case of LSIL triage: 0.96 (CI: 0.86-1.07) for sensitivity and 0.99 (CI: 0.88-1.12) for specificity.

**1.3.3. Difference in accuracy between repeat cytology and the combination of cytology and HPV DNA testing**

The pooled sensitivity for CIN2+ of a repeated Pap smear at threshold ASCUS+ in women with index smear showing ASCUS could be improved with a factor 1.25 (1.10-1.42) by adding HPV DNA test. The specificity of the combination decreased to 68% (CI: 59-79%) of the specificity of cytology alone.

Adding HC2 to repeat cytology at the same threshold and for the same outcome increased the sensitivity of cytology alone with a factor 1.28 (CI: 1.16-1.47).

The combination of repeat cytology (at ASCUS+) and HPV testing in women with a LSIL index smear did not improve the sensitivity (ratio: 1.08; CI: 0.95-1.22). On the other hand, the specificity of the combination decreased significantly (ratio: 0.69; CI: 0.59-0.81).
1.3.4. Sources of between study heterogeneity

Meta-regression was used to explore the sources of inter-study heterogeneity. The variation of the log odds of the sensitivity of repeat cytology was determined by cytological threshold and blinding in histological golden standard interpretation. The degree of histological outcome was marginally not significant. Specificity was a strong predictor of sensitivity. When the logit of the false positivity rate was entered as a covariate in the meta-regression model, only cytological cut-off was required additionally to explain heterogeneity almost completely.

The sensitivity of HPV DNA detection varied significantly by type of testing, also when specificity was controlled for: the Hybrid Capture 2 assay (HC2) had higher log odds than HC1; HC2 and PCR had similar log odds. The accuracy of HPV DNA testing was further significantly influenced by blinding and quality review of histology.

Incomplete reporting of study characteristics hampered the analysis of the impact of potential explanatory factors.

1.3.5. Influence of age on the accuracy of triage

The influence of age on the accuracy of triage could only be assessed approximately in a limited number of studies. A tendency of decreasing sensitivity and increasing specificity with age was observed for HPV DNA detection. In general, specificity of triage is low for women with LSIL. However, specificity of HPV triage could reach an acceptable level in older women with LSIL.

1.3.6. Publication bias

Potential publication bias could be identified for the estimation of sensitivity of repeat cytology, but not for HPV DNA testing as triage method of women with ASCUS. A lower false positive rate of HPV DNA detection was associated with small study size. Therefore specificity of HPV DNA might have been overestimated due to publication bias.

1.3.7. Summary ROC curve analysis

The accuracy expressed by the odds ratio (odds of a positive result among women with CIN2/3+ to the odds of a positive test among women without CIN2/3+) did not differ significantly by histological degree of dysplasia (CIN2+ or CIN3+). Consequently, data for the 2 outcomes can be pooled together, as far as sensitivity and specificity are considered in pairs.

The accuracy differed significantly by triage group and by triage method (HPV testing or repeat cytology). In the ASCUS triage group, the accuracy of repeat cytology differed significantly by cytological cut-off. The prevalence of CIN2/3+ did not change significantly showing acceptable validity of the golden standard verification. Liquid based and conventional cytology showed similar accuracy.
1.4. Conclusions:

Evidence is available indicating improved cross-sectional accuracy of the Hybrid Capture-II assay in comparison with the classical repeat-Pap smear for the detection high-grade intraepithelial neoplasia of the uterine cervix among women with equivocal cytological results. This conclusion is consistent for both outcomes CIN2+ and CIN3+.

The specificity of all methods of triage in women with low-grade squamous intraepithelial lesions is low. Triage with the HC2 assay in particular yields very high positivity rates limiting its utility in most clinical situations.