Performance of Commercially Available HPV Tests

Meagan Stephenson
Carolyn Doughty

New Zealand
Health Technology Assessment

Department of Public Health and General Practice
Christchurch School of Medicine and Health Sciences
Christchurch, New Zealand
Performance of commercially available HPV tests

Meagan Stephenson
Carolyn Doughty
This report should be referenced as follows:

Stephenson, M, Doughty, C. Performance of commercially available HPV tests. NZHTA Technical Brief 2007; 6(7)

Titles in this Series can be found on the NZHTA website: http://nzhta.chmeds.ac.nz/publications

2007 New Zealand Health Technology Assessment (NZHTA)

ISBN 978-1-877455-09-4 (Print)
ISBN 978-1-877455-10-0 (Web)
ISSN 1175-7884
CONTRIBUTIONS BY AUTHORS

This report was authored by Meagan Stephenson (Research Fellow) and Dr. Carolyn Doughty (Research Fellow), who conducted the critical appraisals, prepared the report and coordinated the project.

ACKNOWLEDGEMENTS

This Technical Brief was commissioned by the National Screening Unit, New Zealand Ministry of Health. It was requested by Diane Casey (Senior Policy Analyst) as part of a review of the National Cervical Screening Programmes guidelines for the management of women with abnormal or equivocal cervical smears.

The literature search strategy was developed and undertaken by Susan Bidwell (Information Specialist). Tammy Adams (Administrator) provided document formatting. Internal peer review was provided by Dr. Robert Weir.

We are grateful to Dr. Lance Jennings, Dr. Kitty Croxson and Dr. Collette Bromhead who provided valuable specialist advice regarding the commercial availability of HPV tests.

DISCLAIMER

NZHTA takes great care to ensure the accuracy of the information supplied within the project timeframe, but neither NZHTA nor the University of Otago can accept responsibility for any errors or omissions that may occur. NZHTA and the University of Otago along with their employees accept no liability for any loss of whatever kind, or damage, arising from the reliance in whole or part, by any person, corporate or natural, on the contents of this paper. This document is not intended to be used as personal health advice; people seeking individual medical advice are referred to their physician. The views expressed in this report are those of NZHTA and do not necessarily represent those of the University of Otago, or the New Zealand Ministry of Health.

COPYRIGHT

This work is copyright. Apart from any use as permitted under the Copyright Act 1994 no part may be reproduced by any process without written permission from New Zealand Health Technology Assessment. Requests and inquiries concerning reproduction and rights should be directed to the Director, New Zealand Health Technology Assessment, Christchurch School of Medicine and Health Sciences, PO Box 4345, Christchurch, New Zealand.

CONTACT DETAILS

New Zealand Health Technology Assessment (NZHTA)
Department of Public Health and General Practice
Christchurch School of Medicine and Health Sciences
PO Box 4345
Christchurch
New Zealand
Tel: +64 3 364 3696 Fax: +64 3 364 3697

Email: nzhta@chmeds.ac.nz

Website: http://nzhta.chmeds.ac.nz
LEVEL OF EVIDENCE CONSIDERED IN TECHNICAL BRIEFS

Technical Briefs are rapidly produced assessments of the best available evidence for a topic of highly limited scope. They are less rigorous than systematic reviews. Best evidence is indicated by research designs which are least susceptible to bias according to the National Health and Medical Research Council’s (NHMRC) criteria (see Appendix 1). Where methodologically acceptable and applicable, appraised evidence is limited to systematic reviews, meta-analyses, evidence based clinical practice guidelines, health technology assessments and randomised controlled trials (RCTs). Where not available, poorer quality evidence may be considered.

CONFLICT OF INTEREST

None.
EXECUTIVE SUMMARY

Objective

The primary objective of this technical brief was to examine the test performance of commercially available human papillomavirus testing methods. The focus of the review was on testing methods which could be implemented within the current cervical screening programme, and so commercial tests which are accessible and relevant for high throughput were included. The most likely utilisation of HPV testing within a cervical screening programme will be as a follow-up or adjunct test for women with abnormal or equivocal smear results. To this end, included studies were those which reported the test performance of more than one HPV test in women with an initial abnormal or equivocal smear result.

Methods

The literature was searched using the following databases: Medline, PubMed (last 90 days), Embase, Cinahl Cochrane Central Register of Controlled Trials. The following review databases were also searched: Cochrane Database of Systematic Reviews, Clinical Evidence, DARE database, NHS Economic Evaluation Database, Health Technology Assessment Database, ACP Journal Club. Further information from relevant websites detailing international guidelines for cervical screening and cytological classification systems was sought where necessary. Cited references of retrieved articles were scanned for additional potentially eligible papers.

Searches were of material published from June 2000 onwards in English.

Key results and conclusions

The most well-researched HPV tests appear to be signal amplification methods, such as Hybrid Capture 2 (Digene), and polymerase chain reaction tests utilising various primer sets. Other DNA and messenger-RNA tests which were included in this report, namely Amplicor, HPV DNA Chip and Preact Proofer, show potential but have yet to be tested in large scale well-controlled trials.

Findings were limited by the width of confidence intervals in many of the studies, caused by a small sample size and in particular, a small number of positive histological cases. In addition, many studies omitted details regarding the source, selection criteria and age range of included samples, thus lessening the ability to draw conclusions about the generalisability of their findings. While absolute values of test estimates were difficult to state, it was possible to draw some conclusions about the performance of tests relative to one another.

- The most commonly investigated comparisons were of Hybrid Capture 2 and PCR-based tests, which most often used MY09/11 or GP5+/6+ primer sets. There was little evidence of any difference in test performance with wide and overlapping confidence intervals indicating both a lack of significant difference and a lack of precision in the test estimates. The levels of sensitivity and negative predictive value achieved across studies were approximately 80-90% and consistent for both HC2 and PCR-based tests while specificity and positive predictive value was lower and varied substantially between studies. One large, well-controlled study reported a significant difference between the two tests with HC2 achieving a slightly higher sensitivity and PCR a higher specificity in the detection of CIN2 or worse.

- In studies which compared HC2, a signal amplification test, and Amplicor, a PCR-based test, the sensitivity of the two tests was the same and specificity varied slightly but inconsistently between studies. The sensitivity and negative predictive values were between 80% and 100% with specificity and positive predictive value between 30% and 50%.

- Two studies compared the performance of HPV DNA Chip, a PCR-based oligonucleotide microarray system, and HC2 and the two tests performed very similarly across all test performance characteristics (sensitivity ~85-95% and specificity ~40-50%) with wide confidence intervals. PreTect Proofer, a messenger RNA test, was compared with HC2 in one study and PCR assays in two studies. HC2 was more sensitive but less specific than PreTect Proofer. Negative predictive values were not significantly different but positive predictive value was higher for PreTect Proofer. When compared with PCR-based assays both tests achieved the same or very similar levels of sensitivity (~85-95%) while PreTect Proofer was more specific (85% versus 50%).
predictive values were very similar (~98%) while positive predictive value was low (15-40%) and varied between the two tests, although this difference was not significant.

Overall, there was little evidence of any consistent differences in test performance between the compared HPV testing methods. While Amplicor, HPV DNA Chip and PreTect Proofer all show potential as reliable methods of HPV testing, none of these have been researched to the same degree as HC2 and other PCR assays. There is a need for large-scale well-controlled studies investigating the performance of these tests to obtain more precise estimates of their test performance relative to other HPV testing methods. HC2 and PCR-based tests performed similarly when compared with each other within studies, but test estimates varied between studies and further large scale investigations would enable more precise test estimates to be calculated. The decision to implement either one of these tests will depend on more than test performance alone, and it may be that variances in cost, ease of administration and adaptability to high-volume processing will determine the most suitable test for the New Zealand Cervical Screening Programme.
# TABLE OF CONTENTS

- CONTRIBUTIONS BY AUTHORS .................................................. 1
- ACKNOWLEDGEMENTS ............................................................ 1
- DISCLAIMER ............................................................................. 1
- COPYRIGHT ............................................................................... 1
- CONTACT DETAILS .................................................................... 1
- LEVEL OF EVIDENCE CONSIDERED IN TECHNICAL BRIEFS .......... II
- CONFLICT OF INTEREST ........................................................... II
- EXECUTIVE SUMMARY ............................................................. III
  - Objective ................................................................................ 111
  - Methods ................................................................................ 111
  - Key results and conclusions .................................................. 111
- TABLE OF CONTENTS ............................................................... V
- LIST OF TABLES .......................................................................... VI
- ABBREVIATIONS ....................................................................... VII
- ABBREVIATIONS ....................................................................... VII
- GLOSSARY ................................................................................ IX
- BACKGROUND ........................................................................... 4
- SELECTION CRITERIA ............................................................... 4
  - Study inclusion criteria ......................................................... 4
  - Study exclusion criteria ........................................................ 4
- MAIN SEARCH TERMS ............................................................. 4
- SEARCH SOURCES .................................................................... 5
  - Bibliographic databases ........................................................ 5
- RESULTS ...................................................................................... 5
- APPRAISAL METHODOLOGY .................................................. 5
  - Study designs ......................................................................... 6
  - Settings and samples .............................................................. 6
  - Summary of studies .............................................................. 6
  - Conclusions ........................................................................... 14
- REFERENCES .............................................................................. 37
- APPENDIX 1: LEVELS OF EVIDENCE ............................................. 39
- APPENDIX 2: SEARCH STRATEGIES .......................................... 40
- APPENDIX 3: EXCLUDED RETRIEVED PAPERS ......................... 44
- APPENDIX 4: APPRAISED RETRIEVED PAPERS ......................... 48

---

**BACKGROUND**

**SELECTION CRITERIA**

- Study inclusion criteria
- Study exclusion criteria

**MAIN SEARCH TERMS**

**SEARCH SOURCES**

- Bibliographic databases

**RESULTS**

**APPRAISAL METHODOLOGY**

- Study designs
- Settings and samples
- Summary of studies
- Conclusions

**REFERENCES**

**APPENDIX 1: LEVELS OF EVIDENCE**

**APPENDIX 2: SEARCH STRATEGIES**

**APPENDIX 3: EXCLUDED RETRIEVED PAPERS**

**APPENDIX 4: APPRAISED RETRIEVED PAPERS**
LIST OF TABLES

TABLE 1 Classification of cervical cytology ................................................................................................. 1
TABLE 2 Studies comparing the test performance of Hybrid Capture 2 and Amplicor ................................. 7
TABLE 3 Studies comparing the test performance of Hybrid Capture 2 and PCR-based tests .................. 10
TABLE 4 Studies comparing the test performance of Hybrid Capture 2 and HPV DNA Chip .................... 12
TABLE 5 Studies comparing the test performance of PCR-based tests with PreTect Proofer ...................... 13
TABLE 6 Evidence table of appraised articles relating to clinical test performance of HPV testing .......... 16
# ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGUS</td>
<td>Atypical glandular cells of undetermined significance</td>
</tr>
<tr>
<td>AC</td>
<td>Abnormal cytology</td>
</tr>
<tr>
<td>ADCA</td>
<td>Adenocarcinoma</td>
</tr>
<tr>
<td>AIS</td>
<td>Adenocarcinoma in situ</td>
</tr>
<tr>
<td>ALTS</td>
<td>Atypical Squamous Cells of Undetermined Significance and Low-Grade Squamous Intraepithelial Lesions Triage Study</td>
</tr>
<tr>
<td>AMP</td>
<td>Roche Amplicor test</td>
</tr>
<tr>
<td>ASC-H</td>
<td>Atypical squamous cells – a high grade squamous intraepithelial lesion cannot be excluded</td>
</tr>
<tr>
<td>ASCUS</td>
<td>Atypical squamous cells of undetermined significance</td>
</tr>
<tr>
<td>BCC</td>
<td>Benign cervical cells</td>
</tr>
<tr>
<td>CIN</td>
<td>Cervical intraepithelial neoplasia (degrees of severity from CIN1 to CIN3)</td>
</tr>
<tr>
<td>CRN</td>
<td>Cancer Registry of Norway</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>HC2</td>
<td>Digene Hybrid Capture 2 test</td>
</tr>
<tr>
<td>HD-C</td>
<td>Biomedlab HPV DNA Chip test</td>
</tr>
<tr>
<td>HPV</td>
<td>Human papillomavirus</td>
</tr>
<tr>
<td>HR</td>
<td>High risk</td>
</tr>
<tr>
<td>HSIL</td>
<td>High-grade squamous intraepithelial lesion</td>
</tr>
<tr>
<td>ISH</td>
<td><em>in situ</em> hybridisation</td>
</tr>
<tr>
<td>LBC</td>
<td>Liquid-based cytology</td>
</tr>
<tr>
<td>LEEP</td>
<td>Loop electro-excisional procedure</td>
</tr>
<tr>
<td>LR</td>
<td>Low risk</td>
</tr>
<tr>
<td>LSIL</td>
<td>Low-grade squamous intraepithelial lesion</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid</td>
</tr>
<tr>
<td>NCSP</td>
<td>National Cervical Screening Programme (of New Zealand)</td>
</tr>
<tr>
<td>NPV</td>
<td>Negative predictive value</td>
</tr>
<tr>
<td>OMS</td>
<td>Oligonucleotide microarray system</td>
</tr>
<tr>
<td>OR</td>
<td>Odds ratio</td>
</tr>
<tr>
<td>Pap</td>
<td>Papanicolaou</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate-buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PPV</td>
<td>Positive predictive value</td>
</tr>
<tr>
<td>RFLP</td>
<td>Restriction fragment length polymorphism</td>
</tr>
<tr>
<td>RLB</td>
<td>Reverse line blotting</td>
</tr>
<tr>
<td>RLU/PC</td>
<td>Relative light unit per positive control</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver operating characteristic curve</td>
</tr>
<tr>
<td>SCC</td>
<td>Squamous cell carcinoma</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Se</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>Sp</td>
<td>Specificity</td>
</tr>
<tr>
<td>STD</td>
<td>Sexually transmitted disease</td>
</tr>
</tbody>
</table>
GLOSSARY

Abnormal smear  A smear showing cell abnormalities. The cells on the cervix differ in some way to normal cells.

Adenocarcinoma  A form of cervical cancer where the tumour arises from glandular tissue.

Adenocarcinoma in situ  High grade changes to the glandular cells of the cervix.

Atypical squamous cells of undetermined significance (ASCUS)  Cellular abnormalities that are more marked than those attributable to reactive changes, but that quantitatively or qualitatively fall short of a definitive diagnosis of SIL. Because the cellular changes in the ASCUS category may reflect an exuberant benign change or a potentially serious lesion, which cannot be unequivocally classified, they are interpreted as being of undetermined significance.

Bethesda system  A system for cervical/vaginal cytologic diagnoses. Reports include a descriptive diagnosis and an evaluation of specimen adequacy. This system is used in New Zealand.

Benign  Not cancerous.

Biopsy  Removal of a sample of tissue for examination under a microscope to assist in diagnosis. A cervical biopsy occurs when a tissue sample is taken from the cervix.

Case-control study  An epidemiological study involving the observation of cases (persons with the disease) and a suitable control (comparison) group of persons without the disease. The relationship of an attribute to the disease is examined by comparing retrospectively the past history of the people in the two groups with regard to how frequently the attribute is present.

Cervical carcinoma in situ  Cancer cells that are confined to the surface epithelium of the cervix and have not invaded surrounding tissues.

Cervical intraepithelial neoplasia  Precancerous changes in the surface layers of the cervix. The degree of abnormality is graded as CIN1 (low grade), CIN2 or CIN3 (high grade).

Cervical smear test  A screening test where cells are taken from the cervix, preserved on a slide and sent to a laboratory for examination. It is also known as the Pap smear. The aim of the test is to detect the precursors of cervical cancer or the earliest stage of cancer possible.

Cohort study  An epidemiological study in which subsets of a defined population can be identified who are, have been, or in the future may be exposed or not exposed in different degrees, to a factor or factors hypothesised to influence the probability of occurrence of a given disease or other outcome. Studies usually involve the observation of a large population, or a prolonged period, or both.

Colposcopist  A doctor who specialises in performing colposcopy and possesses knowledge of cytology, histology and diseases of the female genital tract. All colposcopists are obstetricians and gynaecologists specialising in and performing colposcopy.

Colposcopy  Examination of the vagina and cervix using a colposcope (a lighted magnifying instrument) to check for abnormal cells.

Cross-sectional study  A study that examines the relationship between diseases (or other health-related characteristics), and other variables of interest as they exist in a defined population at one particular time.

Cytologist  A registered medical laboratory scientist or technician who is trained in examining smears in the laboratory.

Cytology  The study of cells taken as samples during procedures such as a cervical smear test.
Dyskaryosis  Abnormal changes in the cervix graded as mild, moderate or severe. Used in the United Kingdom to classify stages of precancerous cells.

Dysplasia  Abnormal changes in the cervix graded as mild (CIN1), moderate (CIN2) or severe (CIN3). Used by the WHO to classify stages of precancerous cells.

False negative  A negative test result in a person who actually does have the condition.

False positive  A positive test result in a person who actually does not have the condition.

High grade squamous intraepithelial lesion  The more serious cervical cell changes that include CIN2 and CIN3

Histology  The microscopic study of the structure and composition of tissue.

Human Papillomavirus  Group of viruses that can cause infection in the skin surface of different areas of the body, including the genital area.

Index smear  Initial reference smear to which a subsequent sequence of follow-up management relates.

Intraepithelial lesion  Lesion confined to the surface layer of the cervix. (see also high grade and low grade squamous intraepithelial lesion)

Invasive cancer  Cancerous cells that have spread to deeper tissue.

Koilocytosis  Cell changes caused by infection with human papillomavirus.

Low grade intraepithelial lesion  Mild changes to the cells of the cervix which includes abnormalities due to HPV changes and CIN1 (mild dysplasia).

Negative smear  A normal smear report.

Neoplasia or malignant  Usually refers to abnormal growth and is the same as tumour, which may be benign or malignant.

Oligonucleotide microarray  Short sequences of nucleotide probes designed to match known target sequences of DNA

Pap test or smear  A cervical smear test named after Dr. Papanicolaou, who developed the test. See cervical smear test.

Pathologist  A doctor who specialises in identifying diseases by studying cells and tissues under a microscope and may also be called a cytopathologist/ histopathologist.

Prevalence  The number of events in a given population at a designated time point (point prevalence) or during a specified period (period prevalence)

Reference standard  An independently applied test that is compared to a screening or diagnostic test in order to evaluate it's accuracy. The reference standard provides verification of positive and negative diagnoses.

Screening  The examination of asymptomatic people in order to classify them as likely or unlikely to have the condition that is the object of screening.

Selection bias  Error to due systematic differences in characteristics between those who are selected for inclusion in a study and those who are not (or between those who are compared in a study and those who are not).

Sensitivity  The probability of a positive test result in the presence of abnormality.
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specificity</td>
<td>The ability of a screening test to correctly identify a person who is free from abnormality.</td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>The most common form of cervical cancer arising from the squamous cells in the epithelium.</td>
</tr>
<tr>
<td>True negative</td>
<td>A test correctly identifying a person without the condition</td>
</tr>
<tr>
<td>True positive</td>
<td>A test correctly identifying a person with the condition</td>
</tr>
<tr>
<td>Unsatisfactory smear</td>
<td>A smear that cannot be reported by the laboratory.</td>
</tr>
</tbody>
</table>
BACKGROUND

Ten percent of all cancers diagnosed in women world-wide are cancers of the cervix. Internationally there is wide variance in incidence and mortality with cervical cancer and there is a significant difference in the rates of cancer between developed and developing countries (World Health Organisation, 2002). The decrease in cervical cancer incidence and mortality in developed nations has been largely attributed to the implementation of effective screening programmes. The National Cervical Screening Programme (NCSP) in New Zealand has been established since 1991 with a primary aim of reducing the incidence and mortality rate of cervical cancer through an organised screening programme. Currently women between the ages of 20 and 69 years are recommended to have a cervical smear every three years, provided their results are normal. If their results are abnormal or equivocal, further screening and colposcopy or biopsy procedures may be necessary.

Monitoring women using cervical smear testing has drastically reduced the incidence of cervical cancer in countries with an organised screening programme, with the incidence of cervical cancer in New Zealand declining by 40% in the first ten years of the NCSP. However, concerns have been raised about the test performance of the cervical smear test and efforts have been made to investigate new, more objective technologies for the detection and diagnosis of cervical cancer and its precursors (Koliopoulos et al. 2007).

The NCSP utilises a New Zealand modified version of the Bethesda classification system to categorise the results from cervical smear tests and cytology. However, several other cytological classification systems are used worldwide and were utilised in the studies included in this report. For histological classifications the term cervical intraepithelial neoplasia is used, with CIN1 to CIN3 representing progressively worse outcomes. To aid in the interpretation of the patient inclusion criteria and comparisons with HPV detection utilised in the studies included in this report, the relationship between the different nomenclatures is presented in Table 1 below.

Table 1 Classification of cervical cytology (adapted from Broadstock 1999 and Cancer Research UK website1)

<table>
<thead>
<tr>
<th>Bethesda system</th>
<th>Richart</th>
<th>World Health Organisation</th>
<th>UK classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atypical squamous cells of undetermined significance (ASC-US)</td>
<td>Atypia</td>
<td>Borderline dyskaryosis</td>
<td></td>
</tr>
<tr>
<td>ASC, cannot exclude high-grade squamous intraepithelial lesion (ASC-H)</td>
<td>Cervical intraepithelial neoplasia 1 (CIN1)</td>
<td>Mild dysplasia</td>
<td>Mild dyskaryosis</td>
</tr>
<tr>
<td>Low-grade squamous intraepithelial lesion (LSIL) – encompasses CIN1 and low-grade changes due to HPV infection</td>
<td>Cervical intraepithelial neoplasia 2 (CIN2)</td>
<td>Moderate dysplasia</td>
<td>Moderate dyskaryosis</td>
</tr>
<tr>
<td>High-grade squamous intraepithelial lesion (HSIL) – encompasses both CIN2 and CIN3</td>
<td>Cervical intraepithelial neoplasia 3 (CIN3)</td>
<td>Severe dysplasia</td>
<td>Severe dyskaryosis or worse</td>
</tr>
<tr>
<td>Atypical glandular lesions</td>
<td></td>
<td>Carcinoma in situ (CIS)</td>
<td></td>
</tr>
<tr>
<td>Atypical endocervical lesions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenocarcinoma in situ</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Available from www.cancerresearchuk.org/cancerstats/types/cervix/incidence/
**Human Papillomavirus Testing**

A causal relationship has been established between cervical infection with human papillomavirus (HPV), a sexually-transmitted infection, and cervical carcinomas and precancerous abnormalities. Recent studies have suggested that approximately 99% of cervical cancers contain HPV DNA and the virus is considered a necessary but not sufficient condition for the development of cervical cancer (Metfessel et al. 2005). HPV can be divided into two main types: cutaneous and mucosal. It is mucosal virus genotypes which infect the genital tract and these are further identified as high-risk or low-risk genotypes, depending on the type of lesion they are associated with and the strength of their oncogenic risk. There is variation between countries and geographical regions in the prevalence of HPV genotypes, and some inconsistency in the categorisation of risk type, but the commonly accepted high-risk types include 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73 and 82 while low-risk types include 6, 11, 40, 42, 43, 44, 53, 54, 55, 61, 62, 66, 72, and 81. Most HPV infections resolve spontaneously and will not lead to cervical cancer; however, persistent infection with high-risk HPV types is strongly associated with later cervical cell abnormalities (Garland et al. 2006; Cuzick et al. 1999).

**Methods of HPV DNA detection**

HPV is a small double-stranded DNA virus and testing relies on molecular techniques to detect HPV DNA in cervical cell samples. Some HPV detection methods, such as Southern blot, dot blot and in situ hybridisation (ISH), involve direct probing of the DNA extracted from cells and require a large amount of cellular material, more than that obtained from a cervical smear. Using these techniques DNA is extracted from the cells, denatured to render it single-stranded, and probed using a labelled DNA or RNA complementary molecule (probe) which then binds to the target-DNA sequence (hybridises). After unhybridised molecules are removed, the target-DNA sequence can be identified by the label on the probe. ISH omits the extraction step but is otherwise a similar process. These techniques are considered unsuitable for large-scale clinical HPV detection because of the amount of cellular material required (Southern blot, dot blot), inadequate sensitivity or specificity (dot blot, ISH), and because they are labour-intensive, complicated techniques with little potential for automation (Southern blot, ISH). They are however, widely used in research studies requiring identification and typing of HPV (Garland et al. 2006; Cuzick et al. 1999).

More recent techniques utilise amplification processes which enable the identification of small amounts of DNA, and have shown higher sensitivity and specificity, potential for automation and suitability for processing the large number of samples resulting from screening programmes. These techniques fall into two broad categories, those that amplify the target-DNA and those that amplify the signal produced by the HPV DNA after detection.

**Target-Amplification Techniques:**

Target-amplification techniques for HPV detection include polymerase chain reaction (PCR) as well as oligonucleotide and DNA chip microarrays, which are able to identify individual HPV genotypes. PCR is able to produce millions of copies of an isolated portion of DNA via an amplification cycle. The first stage of the cycle involves denaturing the DNA to render it single-stranded. The goal is to replicate particular molecules on the target strand and in the second stage, called annealing, short synthetic single-stranded DNA molecules (primers) hybridise with their complementary molecules on the target strand. In the next stage, a DNA polymerase is added to the reaction which synthesises the area of target-DNA identified by the primers, creating new double-stranded DNA molecules identical to the original target-DNA strand. The cycle of denaturing, annealing and extension is then repeated to exponentially increase the number of target-DNA molecules. The types of HPV detected depend on the primer sets used on the target-DNA. Two consensus primer sets are commonly used to detect all mucosal HPV types, degenerate (MY09/11 and PGMY09/11) and mismatch acceptance sets (GP5/6 and GP5+/6+) but the SPF primer set is also commonly used and one of the studies appraised in this review utilised LCR and E7 primers in a comparison with the Hybrid Capture 2 assay.

Several commercial PCR-based assays are available but these have been evaluated to different degrees. Roche Diagnostics have produced Amplicor HPV test, a PCR-based commercial HPV assay, which detects 13 high-risk HPV types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68) but does not
allow the identification of the individual genotypes present in the sample. Roche also produce the Linear Array HPV genotyping test, which identifies 37 high- and low-risk HPV genotypes using oligonucleotides. Both Amplicor and Linear Array have CE Mark approval for use as in vitro diagnostic tests, the level of approval required for use in EU countries and accepted by IANZ. Both assays have also been submitted to the U.S. Food and Drug Administration (FDA) for review and approval is expected early next year (E. Bal, Roche Diagnostics NZ Ltd, personal communication).

Biomedlab Company (Seoul, Korea) has developed a PCR-based oligonucleotide microchip assay (HPVDNAChip) which detects and identifies 22 HPV genotypes (15 high-risk - 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68 and 69 and 7 low-risk - 6, 11, 34, 40, 42, 43 and 44). Oligonucleotide microarrays are short sequences of nucleotide probes designed to match known target sequences of DNA. During the PCR process fluorescein-tagged nucleotides are added to target-DNA and the resulting hybrids are applied onto the microchip and scanned by laser fluorescence. The advantages of the DNAChip and Linear Array systems are that they both detect and identify individual HPV genotypes and therefore multiple infections, and speed of processing.

NorChip (Norway) has developed PreTect Proofer, a microarray which utilises nucleic acid sequence based amplification to detect HPV messenger-RNA rather than DNA. The HPV messenger RNA detected by this test is a marker of the production of oncogenic proteins from the E6 and E7 region of the HPV viral genome, which occur as a result of the integration of oncogenic HPV DNA with host cervical cells and indicate malignant cell transformation. NorChip claim higher sensitivity for concurrent cervical disease because PreTect Proofer detects the production of oncogenic proteins indicating cell transformation, as opposed to other tests which detect the presence of the virus DNA but not how it is behaving. However, the current PreTect Proofer assay detects the production of E6 and E7 mRNA by only five high-risk HPV types (16, 18, 31, 33 and 45). PreTect Proofer is a CE Marked in vitro diagnostic test but it must be noted that HPV messenger-RNA detection tests have not been evaluated in large clinical trials as yet (Garland et al. 2006).

Signal Amplification Techniques:

Hybrid Capture 2 (Digene Corporation) is currently the only FDA-approved HPV DNA detection test that utilises signal-amplification, whereby the product of the HPV DNA detection process is amplified, rather than the target-DNA itself. The use of complementary probes to detect the target-DNA sequences is the same as for non-amplification and target-amplification techniques. Firstly, the cellular DNA is extracted and then denatured in alkaline solution. The target-DNA is then combined with a complementary RNA probe cocktail producing DNA-RNA hybrids. Antibodies on the wall of the reaction well then ‘capture’ the DNA-RNA hybrids and the solution is washed to remove all other molecules. Antibody molecules labelled with alkaline phosphatase bind with the DNA-RNA hybrids, amplifying their signal, and the alkaline phosphatase is then reacted with dioxetane substrate to produce light which is measured in a luminometer. This part of the process is described as chemiluminescent signal detection. The level of light measured is expressed in relative light units (RLUs) and, if it is above a cut-off level, indicates the presence of high-risk HPV DNA sequences. The level of RLUs is also broadly representative of the amount of virus DNA present in the sample.

The initial version of Hybrid Capture (HC1) detected eight high-risk and five low-risk HPV types, had relatively low test performance and is no longer in use. In the current version (HC2), two probe cocktails are available, one detecting the presence of any of 13 high-risk HPV DNA types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68) and the other any of six low-risk types (6, 11, 42, 43, 44 and 59) with neither able to identify individual HPV genotypes. The high-risk probe cocktail has FDA approval as an in vitro diagnostic test however the low-risk cocktail does not. The most recently FDA-approved applications of the test are firstly, to screen patients with ASC-US to determine the need for referral to colposcopy, and secondly, in women over the age of 30, to be used adjunctively with cervical smear results to screen for the presence of high-risk HPV types and guide patient management.

Report Scope

The most likely candidates for introduction into a national screening programme utilise PCR techniques (e.g. Roche Amplicor) or hybrid capture signal amplification methods (e.g. Digene Hybrid Capture 2). While many in-house PCR-based assays have been developed, the NCSP is interested primarily in commercially available tests which have been subjected to satisfactory levels of validation;
therefore tests which did not meet these criteria were not included in this report. In addition, the National Cervical Screening Unit (NCSU) requested that the technical brief should comment on the ability of tests to utilise liquid-based cytology techniques, thus information regarding sampling technique and transport medium has been included in the evidence tables.

This Technical Brief was requested by Diane Casey, Senior Policy Analyst, NCSP, New Zealand Ministry of Health. NZHTA has been requested to provide information to the Ministry of Health regarding the test performance of different HPV screening tests.

**SELECTION CRITERIA**

**Study inclusion criteria**

**Publication type**

Studies published from June 2000 onwards in the English language, including primary (original) research (published as full original reports) and secondary research (systematic reviews and meta-analyses) appearing in the published literature. Searches were completed on 23 April 2007.

**Context**

Studies reporting on the test performance of two or more commercially available HPV screening tests among women who have received an abnormal or equivocal cervical smear result were included. Studies were included if they reported test performance relative to a reference standard of colposcopy or histology.

**Outcomes**

Test performance relative to the reference standard, specifically, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were presented in the results.

**Study design**

Systematic review of comparative studies, randomised controlled trials, pseudorandomised controlled trials, comparative studies, case series, either post-test or pretest/post-test (“before-and-after”) studies.

**Sample size**

Studies with samples of at least 50 participants for whom there were reference standard results.

**Study exclusion criteria**

Research papers were excluded if they:

- were not published in English
- were nonsystematic reviews
- were “correspondence”, book chapters, conference proceedings, abstracts
- reported studies with samples of fewer than 50 participants for whom there were reference standard results
- did not clearly describe their methods and results, or had significant discrepancies
- did not compare two or more commercially available HPV tests

**MAIN SEARCH TERMS**

Details of the search strategy are presented in Appendix 2.

MeSH headings (Medline subject headings): papillomavirus infections, cervical intraepithelial neoplasia, reagent kits-diagnostic, exp polymerase chain reaction, “sensitivity and specificity”, predictive value of tests, false negative reactions, false positive reactions, likelihood functions, roc curve, reference standards, diagnosis[as floated subheading], comparative study[as a publication type]

Embase subject headings (where different from Medline): papilloma virus, human papillomavirus type 11, human papillomavirus type 16, human papillomavirus type 18, human papillomavirus type 33,
squamous epithelium, diagnostic kit, diagnostic accuracy, diagnostic error, false negative result, false positive result, predictive validity, roc curve

Additional free text (used in all databases): hpv, papilloma$, ASCUS, ASC-US, ASC-H, LSIL, HSIL, squamous intraepithelial, squamous intra-epithelial, cin-1, cin-2, cin-3, cin1, cin2, cin3, dyskaryosis, low grade, high grade, amplicor, cytyc, roche, imaging system, hybrid capture, biotools, real time assay, screening assay, surepath, thinprep, GenID, sensiv$, specific$, comparative, comparison, (test adj3 performance), accuracy, ppv, npv, positive predictive value, negative predictive value, likelihood ratio$, (detection adj3 rate$)

SEARCH SOURCES

The NZHTA CORE Search was employed. Characteristics of the core search include: essential sources only, major databases and secondary sources, and mostly published and indexed literature.

Bibliographic databases

Principal sources of information

Bibliographic databases:
Medline
PubMed (last 90 days)
Embase
Cinahl
Cochrane Central Register of Controlled Trials

Review databases
Cochrane Database of Systematic Reviews
Clinical Evidence
DARE database
NHS Economic Evaluation Database
Health Technology Assessment Database
ACP Journal Club

Cited references of retrieved articles were scanned for additional potentially eligible papers.

Extended searching of internet websites, meeting abstracts, hand searching of journals, and contacting of authors for unpublished data was not undertaken for this Technical Brief.

RESULTS

From the above search strategy we identified 1004 potentially relevant abstracts of which 68 were retrieved. Of these retrieved articles, 54 were excluded. These papers are presented in Appendix 3.

Articles were rejected for appraisal if they:
- did not compare two or more commercially available HPV tests (n=21)
- reported studies with samples of fewer than 50 participants for whom there were reference standard results (n=5)
- did not compare screening test performance with an adequate reference standard (n=17)
- were systematic reviews of studies which did not compare two or more HPV tests (n=5)
- were non-systematic review articles (n=4)

APPRaisal METHODOLOGY

Summaries of appraisal results will be shown in tabular form (known as Evidence Tables) which detail study design, study setting, sample, methods, results, reported conclusions and NZHTA reviewer conclusions/comments based on the limitations and validity of the study.

The evidence presented in the selected studies were assessed and classified according to the NHMRC’s revised hierarchy of evidence (Appendix 1).
OVERVIEW

The search identified fourteen eligible papers comparing the test performance of two or more commercially available HPV tests with a reference standard of colposcopy or histological diagnoses. Below is an overview of study designs and aspects of quality represented by these studies.

Full details of the papers appraised, including methods, key results, limitations and conclusions, are provided in Table 6 (pages 25 – 45).

Study designs

Nine of the studies were cross-sectional comparative studies where women were recruited into the study and then underwent HPV testing and colposcopy or histological examinations (Carozzi et al. 2007; Kraus et al. 2006; Hwang et al. 2003; Lee et al. 2005; Lie et al. 2005; Kulasingam et al. 2002; Kim et al. 2003; Morin et al. 2001; Bergeron et al. 2000). Three studies were cohort studies, with two of these studies testing for the presence of HPV high-risk virus types and then measuring disease-status at various time-points (Soderlund-Strand et al. 2005; Molden et al. 2005b) and one study using archived cytology samples and pathology records to retrospectively examine the ability of HPV positivity to predict disease-outcome (Cuschieri et al. 2007). One study randomly assigned recruited women to one of three management strategies and followed them prospectively (Schiffman et al. 2005) and one study used a case-control design nested in a cohort study, where controls were matched with cases for age (Yamazaki et al. 2001).

Settings and samples

Five studies recruited women who were undergoing routine screening as part of a national or regional cervical screening programme. Those screening programmes were set in Italy (Carozzi et al. 2007), Norway (Molden et al. 2005b), Sweden (Soderlund-Strand et al. 2005), France (Bergeron et al. 2000) and Japan (Yamazaki et al. 2001). In five of the studies women were recruited after referral to a clinic or hospital following an abnormal smear (Hwang et al. 2003; Lee et al. 2005; Kim et al. 2003; Morin et al. 2001) or histologically diagnosed carcinoma (Kraus et al. 2006), and in four studies women were recruited from multiple clinics or general practitioners (Cuschieri et al. 2007; Lie et al. 2005; Schiffman et al. 2005; Kulasingam et al. 2002).

Two studies included only women over the age of 30 years in their samples (Yamazaki et al. 2001; Molden et al. 2005b) and nine studies included women of all ages. Three studies did not report the mean age or age range of the sample (Soderlund-Strand et al. 2005; Lee et al. 2005; Kraus et al. 2006)). In several studies where the age range of the source population was reported, there was no further information provided regarding the age range or mean age of the final included sample and only one study reported the test performance of HPV tests separately for younger and older women (Kulasingam et al. 2002). Given the high prevalence of HPV virus types in younger women, this is a potential limitation of these studies.

The HPV tests compared in each of the studies varied. The test performance of Hybrid Capture 2 was compared with Roche Amplicor in two studies, with PCR-based tests in six studies, with HPV DNA Chip in two studies and with PreTect Proofer in one study. PCR-based tests were compared with HPV DNA Chip in one study and with PreTect Proofer in two studies. The overview and findings for each of the studies are presented below in separate sections for each of these comparisons.

Summary of studies

Hybrid Capture 2 compared with Amplicor

Two studies compared the performance of Digene Hybrid Capture 2 (HC2) and Roche Amplicor (see Table 2 over). Carozzi et al. (2007) assessed the analytical agreement and clinical effectiveness of two commercially available HPV tests in a population of 1032 women who were recruited to the study through the Florence Cervical Screening Programme. Most of the women (n=962) were recruited to the study following an abnormal smear result (ASCUS/AGCUS/LSIL/HSIL) with a further 70 women with histologically confirmed CIN2 or worse lesions included to increase the proportion of women with disease. Liquid-based cytology and HPV testing using HC2 were carried out on 1032 samples, followed by colposcopy and punch biopsies as necessary for women with cytology results of ASCUS or worse (n=281) as well as those who were over 35 years in age and positive for high-risk HPV.
Testing using Amplicor was completed two years later on stored samples. Clinical sensitivity and specificity of the two HPV tests was calculated using the findings from 270 women for whom there was a reference standard histological result where a diagnosis of a CIN2 lesion or higher was considered a positive and less than CIN2 was considered a negative.

Both HPV tests showed excellent sensitivity but low specificity for the detection of CIN2 lesions or worse. HC2 had a marginally higher sensitivity (97.7% vs. 96.5%) and slightly lower specificity (51.6% vs. 54.5%) than Amplicor. The positive predictive value was 48.6% and 50.0% for HC2 and Amplicor respectively with the corresponding negative predictive values being 97.9% and 97.1%.

Cuschieri et al. (2007) evaluated the clinical performance of HC2 and Amplicor HPV tests in a sample of 190 archived cervical cytology specimens collected at 15 general practices in Edinburgh. Liquid-based cytology showed borderline changes in 322 samples which were classified according to the British Society for Clinical Cytopathology guidelines. HPV testing using both HC2 and Amplicor was carried out on residual samples. Unfortunately, follow-up histology outcomes were only available for 190 women who were grouped on the basis of their three-year disease outcome. Group 1 (n=123) were cleared of their original abnormal cytology result, group 2 (n=33) had persistent low-grade disease, and group 3 (n=34) had persistent high-grade disease. Sensitivity and specificity for the two tests was calculated relative to the three-year outcome for these 190 women with a positive reference standard outcome being a diagnosis of high-grade disease (CIN2 or worse).

Both HC2 and Amplicor showed the same level of sensitivity (82.4%) and low specificity, 44.2% and 36.5% respectively. The difference in specificity was not significant. Positive and negative predictive values were unable to be calculated as were confidence intervals for other test characteristics. The authors suggested that both HC2 and Amplicor were suitable for high throughput sample processing.

The mean age of the 190 women with histological follow-up information was 31 years (SD = 10.5). Because the cytological and histological results were based on pathology records, there was no review of findings or assurance that pathologists were blind to other HPV or cytology results. The authors also examined the effect of age (<30 compared to ≥ 30 year olds) on clearance of borderline abnormalities but found no association. This may have been because the small sample size led to a lack of statistical power for this analysis, and this was acknowledged by the authors.

Table 2  Studies comparing the test performance of Hybrid Capture 2 and Amplicor.

<table>
<thead>
<tr>
<th>Study</th>
<th>Population/Screening test</th>
<th>Reference standard</th>
<th>HPV tests</th>
<th>Sensitivity % (95% CI)</th>
<th>Specificity % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carozzi et al.(2007)</td>
<td>ASCUS/SIL cytology ≥35 years + HC2 +ve N=270</td>
<td>Concurrent histology</td>
<td>HC2 AMPICOR</td>
<td>97.7 (91.9 – 99.7)</td>
<td>96.5 (90.1 – 99.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cuschieri et al. (2007)</td>
<td>Borderline cytology N=190 Mean = 31 years</td>
<td>3 year outcome of CIN2+</td>
<td>HC2 AMPICOR</td>
<td>82.4</td>
<td>44.2</td>
</tr>
</tbody>
</table>

Hybrid Capture 2 compared with PCR-based tests

Six studies compared the performance of HC2 with a PCR-based test (see Table 3 over).

Schiffman et al. (2005) examined the clinical performance of HC2 and a PCR test using PGMY09/11 primer sets within a large two year randomised clinical trial which evaluated the effectiveness of different management strategies for women with equivocal (ASCUS) and mildly abnormal cytological (LSIL) findings. Over 5000 women (ASCUS=3488, LSIL=1572) were randomly allocated to one of three management strategies; immediate colposcopy, HPV triage with colposcopy for women with HPV positive or HSIL, or conservative management where colposcopy was available for HSIL women only. Women were recruited from four university-based medical centres and six-monthly follow-up visits were completed, consisting of cytology and HPV testing, together with an exit examination and colposcopy. The PGMY09/11 PCR test identified the same 13 high-risk types of HPV as the HC2 test.
to allow for accurate comparison. Thirty-four women were excluded because they had no HPV result thus leaving a final sample of 5026 women. Cytological and histological findings were based on clinical centre pathologist’s diagnoses. However referral smears, ThinPrep specimens and histological slides were also sent to a pathology review panel (QC pathology, Johns Hopkins Hospital).

Clinical sensitivity and specificity of HC2 and PGMY09/11 PCR tests for the detection of QC pathology diagnoses of CIN3 or cancer and clinical centre pathology diagnoses of CIN2 or worse were calculated for each follow-up visit. Any women who had undergone treatment following each visit were excluded from the analyses for subsequent visits. Clinical sensitivity and specificity for 2 year cumulative cases of cancer were also calculated and comparisons made between HC2 and different versions of the PCR test with varying numbers of HPV types included.

At every time point HC2 had a higher sensitivity and lower specificity than the PCR test. This was true for the prediction of both CIN3/cancer and CIN2 or worse disease endpoints. For example, the sensitivity and specificity of the two tests when the initial HPV result was used to predict CIN3 or cancer was 93.6% and 41.2% for HC2 and 89.3% and 48.5% for PGMY09/11 PCR. When the initial HPV result was used to predict CIN2 or worse, sensitivity was 93.7% and 87.2% and specificity 44.6% and 51.7% for HC2 and PGMY09/11 PCR respectively. When an additional four HPV types were added to the PCR assay, the sensitivity increased marginally (88.1%) and specificity decreased (52.6%). Visits where the clinician knew the HPV status by HC2 were excluded to test for bias but no significant difference in test performance was found. When the overall clinical performance of the tests was compared for the two-year cumulative detection of CIN3 or cancer, HC2 was also more sensitive (92.4% vs 86.7%) and less specific (50.6% vs 55.7%) than the PCR test. There was a trend for lower sensitivity for either HPV assay for detection of cumulative cases of CIN3 or cancer or CIN2 or worse for testing performed on specimens from later visits, with a large decrease in sensitivity for the prediction of CIN3 cases diagnosed at the last exit visit.

The authors identified two potential limitations of their study, the first being that the population of women included in this study was not representative of all women with HPV infection. Secondly, HC2 testing was carried out in four laboratories whereas the PCR test was carried out in one laboratory by an expert pathologist, thereby creating a potential for bias in the performance of the tests. Overall however, this large randomised study was conducted well and provides sound comparisons of test performance.

Soderlund-Strand et al. (2005) compared the performance of HC2 with that of a PCR-based test using GP5+/6+ primers for the detection of CIN and CIN recurrence after treatment. In the initial visit, a repeat smear, colposcopy, biopsy and HPV testing was carried out in a sample of 239 women referred on the basis of an initial abnormal smear result (ASCUS or worse). Women with at least two of the following were referred for conization (n=177): repeat smear of CIN1 or greater, biopsy result showing CIN1 or greater, and/or an abnormal colposcopy result. The remaining women (n=62) had a follow-up smear test 4 months after their initial result. In the final visit another smear and HPV testing was completed. Of women who had undergone conization 114 had the final visit smear and HPV test as did 33 of the women who did not undergo conization.

Clinical sensitivity and specificity for the detection of histologically confirmed CIN1 or higher, CIN2 or higher and CIN3 were calculated for each HPV test. Similar sensitivities and positive predictive values were obtained for HC2 and GP5+/6+ PCR, with sensitivity ranging from 86.4% - 100% for CIN1 – CIN3 and PPV ranging from 28.7% (HC2 detection of CIN3) to 87% (PCR detection of CIN1 or worse). The specificity of GP5+/6+ PCR testing (23.5% – 36.7%) was greater than that of HC2 (18.9% – 24.1%) for all disease outcome levels. The same was true of negative predictive value (PCR = 35.4% – 100%; HC2 = 25.9% – 92.6%). Overall the sensitivity and negative predictive value of both tests increased while the specificity and positive predictive value decreased as the severity of disease outcome increased from CIN1 or worse to CIN3.

The blinding of cytotechnologists and pathologists to other laboratory results was not stated. Confidence intervals were wide for specificity and NPV, lessening the precision of these estimates of test performance.

Morin et al. (2001) evaluated the clinical performance of Hybrid Capture 1, Hybrid Capture 2 and a PCR test using a primer set which amplifies the L1 gene and detects 11 high-risk HPV types (Digene Probe Primer set). Women consecutively referred to a hospital clinic for colposcopy based on an equivocal or abnormal smear result (ASCUS or worse) underwent a repeat smear test, colposcopy and a colposcopy-directed biopsy as necessary (n=360). Cytology was confirmed by a senior cytologist and histologies were reviewed by one pathologist. There were 19 histologically confirmed cases of CIN2
or CIN3 and 341 disease-free controls. Three triage strategies were compared for their ability to predict CIN2 or CIN3 following an abnormal smear: repeat abnormal smear, positive HPV test by each of the three HPV tests, and a combination of repeat smear and/or a positive HPV test.

The test performance was reported separately for each of the HPV tests. The sensitivities of HC2 and the PCR test in the detection of CIN2/3 were the same (89.5%) and higher than that of HC1 (68.4%). However, the confidence intervals were very wide and the differences were not significant. HC1 obtained a significantly higher specificity (85.9%) than HC2 (74.1%) and PCR (59.0%), and HC2 had a significantly higher specificity than PCR. The positive predictive values of the three tests were low, 21.3 (HC1), 16.0 (HC2) and 11.2 (PCR), and had wide confidence intervals, but there was a significant difference in PPV between HC1 and PCR. The negative predictive values were very high (HC1=98.0%, HC2 = 99.2%, PCR = 99.0) with no significant difference between the three tests. The small number of cases of CIN2 or CIN3 (n=19) most likely contributed to the low study power and the lack of a significant difference in sensitivity and PPV is not surprising.

Yamazaki et al. (2001) used a case-control study design to examine the clinical sensitivity and specificity of HC2 and a nested-PCR test using LCR-E7 primers in the detection of LSIL or worse. The cases were 308 women selected from a pool of 1000 women who received an abnormal smear result in a population-based screening programme in Japan. These women underwent an interview and colposcopy-directed biopsy and women with negative or equivocal results were excluded, with the final 286 cases being comprised of women with abnormal cytology and follow-up histological findings of LSIL or worse. The control group were 114 disease-free women who were matched for age with the cases. Smear tests were screened by one cytotechnologist and all abnormal smears and histologic slides were reviewed by two pathologists.

The sensitivity (HC2= 83%, PCR= 81%) and specificity (HC2= 93%, PCR= 93%) were high for both tests with no significant differences. The high specificity may be because of the exclusion of LSIL women from the control group. When women with normal cytology or LSIL were included in the control group, i.e. negative according to the reference standard test, the specificity dropped to 71% for PCR and 67% for HC2. The sensitivity may have been affected by the use of HSIL or worse as the disease endpoint. The low participation rate (31%) in this study suggests that the sample may not have been representative of the general population of women with abnormal cytology and lessens the generalisability of the findings.

Kulasingam et al. (2002) evaluated the clinical test performance of HC2 and a PCR test using MY09/11 and HMB01 primer sets for the detection of women with CIN3 lesions or cancer. Women aged 18 – 50 years (mean = 25 years ± 5.7) were consecutively recruited from Planned Parenthood clinics in Washington State, U.S.A. Of 4358 women who were invited to participate, 4075 (94%) agreed and underwent cytology and HPV testing using HC2 and a PCR test. Not all women were tested using both HC2 and PCR so corrected test performance estimates were calculated. Women with a finding of ASCUS or worse following the initial smear were referred for colposcopy and biopsy. In addition, a random sample of women with a negative smear result were invited for colposcopy, of whom 202 (7.7%) accepted. Cytology and histology slides were reviewed by pathologists who had no knowledge of other laboratory results or clinical findings. Inadequate, insufficient cellular material or missing results were classified as positive because these women would usually be asked to return for a repeat smear.

Clinical performance of the HC2 and MY09/11 PCR test were compared for the detection of histologically confirmed CIN3 or worse. Corrected test sensitivities were also calculated where the correction was made on the basis that the likelihood of women with negative cytology accepting colposcopy would increase with perceived severity of smear results. HC2 and MY09/11 PCR obtained the same sensitivity (73.6%) and very similar specificities (HC2 = 88.9%, PCR = 89.8%) with no significant differences between the two. Corrected sensitivities for HC2 (60.3%) and MY09/11 PCR (59.8%) were also not significantly different and there was no difference in negative predictive value (98.5%). Test performance was also examined for different age groups with increased sensitivity and decreased specificity reported for women less than 30 years old relative to women 30 years or older, but no difference between HC2 and PCR results. The mean age of this population was relatively young with 81% of enrolled women aged less than 30 years. This may reduce the generalisability of the findings in an otherwise soundly designed and executed study.

Bergeron et al. (2000) evaluated the clinical test performance of HC2 and a PCR test based on MY09/11 primers for the detection of CIN and high-grade CIN in a large screening population whose cytology slides were processed at a private laboratory in Paris over a two year period. Over a thousand
women with an abnormal or equivocal smear result (ASCUS or worse) were asked to participate, of whom 404 (39%) agreed. The mean age of the included sample was 35 years (range = 15-75 years). Repeat cytology, HPV testing and colposcopy was carried out in all enrolled women and biopsy in 384 women. Abnormal smear tests and all biopsies were evaluated by one pathologist who was blind to the results of all other tests. Biopsies were classified as normal (including normal biopsy or no biopsy result because of a normal transformation zone), low grade CIN or high grade CIN with 26 inadequate biopsies being excluded, leaving a final sample size of 378 women. Of these, 320 women had complete PCR, HC2 and biopsy results.

Clinical sensitivity and specificity for the detection of CIN and high-grade CIN was calculated for each of the HPV tests. HC2 (consisting of high-risk and low-risk probes) had a higher sensitivity (81% versus 77%) and specificity (60% versus 52%) than MY09/11 PCR for the detection of CIN but a lower sensitivity (86%) compared to MY09/11 PCR (95%) for the detection of high-grade CIN with very similar specificities for both tests (HC2 = 41%, MY09/11 PCR = 40%). Using the high-risk probe only in the HC2 test decreased the sensitivity (77%) and increased specificity (66%) for the detection of CIN, whereas sensitivity remained basically unchanged (88%) and specificity increased (49%) for the detection of high-grade CIN. The participation rate was low (39%) in this study and complete test and reference standard results were not available for 21% of those who agreed to participate, suggesting the sample may not be representative of the screening population.

Table 3  Studies comparing the test performance of Hybrid Capture 2 and PCR-based tests

<table>
<thead>
<tr>
<th>Study</th>
<th>Population/screening test</th>
<th>Reference standard</th>
<th>HPV tests</th>
<th>Sensitivity % (95% CI)</th>
<th>Specificity % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schiffman et al. (2005)</td>
<td>ASCUS/LSIL cytology</td>
<td>Colposcopy or histology</td>
<td>HC2</td>
<td>93.7</td>
<td>44.6</td>
</tr>
<tr>
<td></td>
<td>N=5060</td>
<td></td>
<td>PCR</td>
<td>P &lt; 0.01</td>
<td>51.7</td>
</tr>
<tr>
<td></td>
<td>Range = 18-81 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soderlund-Strand et al. (2005)</td>
<td>ASCUS+ cytology</td>
<td>Colposcopy or histology</td>
<td>HC2</td>
<td>86.4 (80.9–91.9)</td>
<td>33.3 (8.2-38.5)</td>
</tr>
<tr>
<td></td>
<td>N=177</td>
<td></td>
<td>PCR</td>
<td>86.4 (80.9–91.9)</td>
<td>36.7 (19.4 – 53.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HC2</td>
<td>91.8 (86.4 – 97.3)</td>
<td>24.1 (14.6 – 33.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PCR</td>
<td>92.9 (87.8 – 98.0)</td>
<td>30.4 (20.2 – 40.5)</td>
</tr>
<tr>
<td>Kulasingam et al. (2002)</td>
<td>ASCUS+ cytology</td>
<td>Colposcopy and histology</td>
<td>HC2</td>
<td>60.3 (47.4-69.6)</td>
<td>88.9 (88.1 – 89.6)</td>
</tr>
<tr>
<td></td>
<td>7.7% of -ve cytology</td>
<td></td>
<td>PCR</td>
<td>59.8 (47.1 – 68.9)</td>
<td>89.8 (89.2 – 90.5)</td>
</tr>
<tr>
<td></td>
<td>N= 4075</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean = 25 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moin et al. (2001)</td>
<td>ASCUS cytology</td>
<td>Histology for CIN 2/3</td>
<td>HC2</td>
<td>89.5 [66.9 – 98.7]</td>
<td>74.1 (68.9 – 78.5)</td>
</tr>
<tr>
<td></td>
<td>N=360</td>
<td></td>
<td>PCR</td>
<td>89.5 [66.9 – 98.7]</td>
<td>59.0 (53.4 – 64.3)</td>
</tr>
<tr>
<td></td>
<td>Range = 18 – 50 years</td>
<td></td>
<td>HC1</td>
<td>68.4 [43.5 – 87.4]</td>
<td>85.9 [81.8 – 89.4]</td>
</tr>
<tr>
<td>Yamazaki et al. (2001)</td>
<td>Abnormal cytology</td>
<td>Colposcopy and biopsy</td>
<td>HC2</td>
<td>83</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>N=286</td>
<td></td>
<td>PCR–LCR E7</td>
<td>81</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>Sample mostly ≥ 30 yrs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3 continued. Studies comparing the test performance of Hybrid Capture 2 and PCR-based tests.

<table>
<thead>
<tr>
<th>Study</th>
<th>Population/ screening test</th>
<th>Reference standard</th>
<th>HPV tests</th>
<th>Sensitivity % (95% CI)</th>
<th>Specificity % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bergeron et al. (2000)</td>
<td>ASCUS/LSIL/HSIL N=404 Mean = 35 years</td>
<td>Colposcopy and histology</td>
<td>HC2</td>
<td>81.4 (75.4 – 87.4)</td>
<td>59.8 (52.1 – 67.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PCR</td>
<td>77.0 (70.5 – 83.5)</td>
<td>52.2 (44.4 – 60.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CIN2+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HC2</td>
<td>86.4 (72.0 – 100)</td>
<td>40.9 (35.4 – 46.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PCR</td>
<td>95.3 (86.8 – 100)</td>
<td>39.9 (34.4 – 45.5)</td>
</tr>
</tbody>
</table>

Hybrid Capture 2 compared with HPV DNA Chip

Two studies compared the test performance of HC2 and HPV DNA Chip (see Table 4 over).

Kim et al. (2003) compared the test performance of HC2 and an oligonucleotide microarray system (HPV DNA Chip, Biomedlab, Korea) in a population of women who had previously undergone HPV testing by HC2 at a hospital in Korea. Women who were recruited into the study had a repeat smear and a sample was collected for HPV testing by the two methods. Smears were classified using the Bethesda system with 81 women (58%) receiving a diagnosis of ASCUS or worse. All women underwent histopathologic examination and diagnosis and were grouped on the basis of these findings; group 1 non-specific chronic cervicitis, group 2 koilocytosis or mild dysplasia, group 3 moderate dysplasia, severe dysplasia or carcinoma in situ.

The test performance of HC2 and the oligonucleotide microarray system (OMS) was highly similar with respective sensitivities of 94.9% and 93.7%, and specificities of 50.8% and 49.2%. The positive predictive values were 71.4% (HC2) and 70.5% (OMS) and the negative predictive values 88.6% (HC2) and 85.7% (OMS).

No information was provided regarding the blinding of cytotechnologists or pathologists to other laboratory results. There was also no information regarding how the women were selected into the study, that is, whether it was a consecutive group of women or a random selection. A potential lack of blinding and bias in the selection of participants makes it difficult to interpret the accuracy of the findings. The small sample size also contributed to relatively wide confidence intervals and less accuracy in the estimates of specificity and negative and positive predictive values.

Lee et al. (2005) evaluated the clinical test performance of HC2 in comparison with HPV DNA Chip (HD-C) in a sample of 400 women referred for follow-up on the basis of abnormal cytology. Enrolled women underwent ThinPrep cytology, HPV testing, colposcopy and biopsy. Some HD-C tests were carried out on samples which were frozen after ThinPrep cytology and cytologic diagnoses were made according to the Bethesda system. Histologic diagnoses were made based on the most serious specimen obtained by colposcopy-directed biopsy or loop electrosurgical excision procedure (LEEP).

Clinical test performance for the detection of histologically diagnosed CIN1 or worse was calculated for each of the HPV tests. The sensitivities and specificities of each of the tests was very similar with a high sensitivity (HC2 = 89.9%, HD-C = 86.2%) and low specificity (HC2 = 43.2%, HD-C = 46.2%). Positive predictive values (HC2 = 76.3%, HD-C = 76.5%) and negative predictive values (HC2 = 67.9%, HD-C = 62.2%) were moderate and the two tests performed very similarly. The level of histological diagnosis considered a positive result was CIN1 for this study, whereas most other studies compared the performance of the HPV tests to diagnoses of CIN2 or worse. When the authors of the current report recalculated test performance relative to a diagnosis of CIN2 or worse, sensitivity for each of the tests was unchanged (HC2 = 89.5%, HD-C = 84.7%) but the estimates of specificity (HC2 = 32.5%, HD-C = 34.6%) and positive predictive value decreased and negative predictive value increased. The confidence intervals for the estimates of specificity and negative predictive value were relatively wide, lessening the precision of these estimates.

No information was provided regarding the blinding of cytotechnologists or pathologists to other laboratory results. There was also no information regarding how the women were selected into the
study, that is, whether it was a consecutive group of women or a random selection. A potential lack of blinding and bias in the selection of participants makes it difficult to interpret the accuracy of the findings.

### Table 4 Studies comparing the test performance of Hybrid Capture 2 and HPV DNA Chip

<table>
<thead>
<tr>
<th>Study</th>
<th>Population/screening test</th>
<th>Reference standard</th>
<th>HPV tests</th>
<th>Sensitivity % (95% CI)</th>
<th>Specificity % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim et al. (2003)</td>
<td>Positive and negative cytology N=140 Mean = 39 years</td>
<td>Histology</td>
<td>HC2</td>
<td>Low-grade SIL 94.9 (90.1 – 99.8)</td>
<td>50.8 (38.3 – 63.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HPV-OMS</td>
<td>93.7 (88.3 – 99.0)</td>
<td>49.2 (36.6 – 61.7)</td>
</tr>
<tr>
<td>Lee et al. (2005)</td>
<td>Abnormal cytology N=400</td>
<td>Colposcopy and histology</td>
<td>HC2</td>
<td>CIN1 or worse 89.9 (86.3 – 93.5)</td>
<td>43.2 (34.7 – 51.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HPV-Chip</td>
<td>86.2 (82.1 – 90.3)</td>
<td>46.2 (37.7 – 54.7)</td>
</tr>
</tbody>
</table>

**PCR compared with HPV DNA Chip**

Hwang et al. (2003) evaluated the sensitivity of an oligonucleotide microarray system (HPV DNA Chip, Biomedlab, Korea) and a PCR-based test using GDP5+/6D+ primers for the detection of CIN1 or worse. Consecutive women who were examined at a hospital clinic in Incheon, South Korea and showed abnormal cytology following a smear (n=234) were recruited to the study. Of these 234 women, 212 (90.6%) agreed to follow-up HPV testing and biopsy, while 22 women refused biopsy and were followed by repeat smear and HPV testing.

Sensitivities of each of the HPV tests were reported for the detection of CIN1/CIN2/CIN3 and carcinoma but no information was reported regarding other test characteristics. The sensitivity of the HPV DNA Chip was higher than that of the GDP5+/6D+ PCR test for all disease-endpoints; CIN1 (PCR=50.0%, HD-C = 70.8%), CIN2 (PCR=68.2%, HD-C = 77.3%), CIN3 (PCR = 76.5%, HD-C= 82.4%) and carcinoma (PCR=88.9%, HD-C=90.3%).

While the sensitivities of the tests varied from moderate to high test performance, it is impossible to ascertain true test performance because of a lack of information regarding test specificity and other test characteristics. Both the HPV tests had higher sensitivities for more severe levels of disease. There was not enough information to calculate confidence intervals or whether any differences between the tests were significant. There was no information as to whether pathologists were blind to other laboratory results when evaluating histological outcomes and no information was provided regarding the age range of the sample.

**Hybrid Capture 2 compared with PreTect Proofer**

Lie et al. (2005) compared the clinical test performance of HC2 and a messenger RNA HPV test (PreTect Proofer, NorChip) for the detection of high-grade cervical neoplasia in women referred for colposcopy on the basis of abnormal cytology. Women were recruited from five outpatient clinics and 10 gynaecologists in private practice in Norway (n=383). All recruited women underwent a repeat smear, HPV testing by HC2 and PreTect Proofer, colposcopy, and colposcopy-directed biopsy or conization. Smear samples were classified by one cytopathologist blind to HPV test results and histology was evaluated by one pathologist blind to other results.

The clinical test performance of each HPV test was reported relative to the detection of CIN2 histological diagnoses or worse. HC2 had a higher sensitivity than PreTect Proofer (94.5% vs. 77.3%) but a lower specificity (35.9% vs. 78.3%). The positive predictive values for HC2 and PreTect Proofer were 82.3% and 91.8% respectively, and the negative predictive values were 67.3% and 52.2% respectively. Confidence intervals were relatively wide for the estimates of specificity and negative predictive value.

The PreTect Proofer assay detected five high-risk HPV types while HC2 detects 13 high-risk types, so the differences in sensitivity and specificity are not surprising. The authors suggested that the
sensitivity of the PreTect Proofer assay could be improved if more high-risk HPV types were included, however this may decrease the specificity of the test. No mention was made of the sample selection method, that is, whether women were recruited consecutively or randomly from clinics. The participation rate was also not stated.

**PCR compared with PreTect Proofer**

Two studies compared the performance of PCR-based tests and PreTect Proofer (see Table 5 below).

Molden et al. (2005b) evaluated the test performance of PreTect Proofer and a PCR-based test using GP5+/6+ consensus primers. Women were recruited from a subsample of 4136 women who visited a selection of gynaecologists in Oslo as part of the Norwegian Cervical Screening Programme. All 4136 women underwent a smear test and HPV testing with their original gynaecologist. Those with a smear result of ASCUS or LSIL (n=77) were followed up longitudinally for two years using the registers of the Cancer Registry of Norway (CRN) to collect information regarding subsequent smear and biopsy results. Four women with no subsequent smear or biopsy records were excluded from the study. Women who were included for follow-up were all older than 30 years with a mean age of 47 years.

The test performance of each of the HPV tests for the detection of CIN2 lesions or worse over the two year follow-up period was reported. Both PreTect Proofer and GP5+/6+ achieved high sensitivity (85.7%) and negative predictive values (PreTect = 98.3, GP5+/6+ = 97.1). Other test characteristics varied however with PreTect Proofer having a significantly higher specificity (84.9%) than GP5+/6+ PCR (50.0%). The negative predictive value of PreTect was higher (37.5%) than the PCR test (15.4%) but the confidence intervals were wide and this difference was not significant.

Cytology and histology diagnoses were carried out by a selection of pathologists at different laboratories. This reflects standard practice but may have introduced bias in interpretation and diagnosis of cellular material. There was no expert review of diagnoses, and the blinding of pathologists to previous results was not stated. The PreTect Proofer assay identified the E6/E7 transcripts of five high-risk HPV types while the GP5+/6+ PCR test identified more than 20 types of HPV types. Only CIN2 histological diagnoses or worse were reported to the CRN. The sample size was small and confidence intervals for all test characteristics except negative predictive value were wide, decreasing the precision of the findings.

Kraus et al. (2006) also evaluated the test performance of PreTect Proofer and a PCR-based test using GP5+/6+ primers in a population of Norwegian women (n=204). Women with histologically confirmed invasive squamous cell carcinomas who were admitted to a Norwegian hospital for treatment between 1995 and 1998 were included in the study. Cell samples from these women were assessed for HPV positivity and the sensitivity of each of the tests calculated.

PreTect Proofer and GP5+/6+ PCR testing had similar levels of sensitivity (PreTect = 88.7%, PCR = 91.7%). No information regarding other test characteristics was reported and the inability to estimate specificity was a major limitation of the study. It was not clear from the study whether the cytotechnologists who assessed HPV positivity were blind to the disease-status of the women. No mention was made of the sample selection method, that is, whether women were recruited consecutively or randomly and the participation rate was also not stated.

**Table 5  Studies comparing the test performance of PCR-based tests with PreTect Proofer**

<table>
<thead>
<tr>
<th>Study</th>
<th>Population/Screening test</th>
<th>Reference standard</th>
<th>HPV tests</th>
<th>Sensitivity % (95% CI)</th>
<th>Specificity % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molden et al. (2005b)</td>
<td>ASCUS/LSIL N=77 Mean = 46 years All ≥ 30 years</td>
<td>Histologically confirmed 2 year outcome for CIN2+</td>
<td>PCR PreTect Proofer</td>
<td>85.7 (42.1 – 99.6)</td>
<td>50.0 (37.4 – 62.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>85.7 (42.1 – 99.6)</td>
<td>84.9 (73.9 – 92.5)</td>
</tr>
<tr>
<td>Kraus et al. (2006)</td>
<td>Histologically confirmed SCC N=204</td>
<td>Histology</td>
<td>PCR PreTect Proofer</td>
<td>91.7</td>
<td>88.7</td>
</tr>
</tbody>
</table>

**EFFECTIVENESS OF HPV SCREENING IN WOMEN WITH ABNORMAL OR EQUIVOCAL CERVICAL SMEAR RESULTS**
Conclusions

The most well-researched HPV tests appear to be signal amplification methods, such as HC2 (Digene), and polymerase chain reaction tests utilising various primer sets. Other DNA and messenger-RNA test methods which were included in this report, namely Amplicor, HPV DNA Chip and PreTect Proofer, show potential but have yet to be tested in large scale well-controlled trials.

In many studies, findings were limited by the imprecision of test characteristics due to a small sample size and, in particular, a small number of positive histological cases. The confidence intervals for sensitivity and positive predictive value were often wide, meaning that while one may be able to comment on the performance of one test relative to the other, it is difficult to draw conclusions about overall estimates of test performance. Many studies failed to include details about the sample, such as the average age and age range or the selection criteria for included and excluded women. In addition, details regarding the blinding of cytologists and histologists were often not specified. The problem was not so much that these studies were of poor quality but more that it was difficult to judge the sample and design quality because that information was omitted. Sample spectrum bias may have been present in many of the studies because of variation in the source sample and inclusion and exclusion criteria. There was also substantial variation in the participation rate achieved in studies, with several studies reporting low participation rates of 30-40%.

Because of differences in sample source, selection and size and because of the width of confidence intervals, absolute values of test estimates are difficult to state, however it is possible to draw some conclusions about the performance of tests relative to one another within the various comparisons. In studies which compared HC2, a signal amplification test, and Amplicor, a PCR-based test, the sensitivity of the two tests was the same and specificity varied slightly but inconsistently between studies. Confidence intervals, where they were able to be calculated, were overlapping for both sensitivity and specificity, indicating that there was no significant difference between the tests.

In studies which compared the test performance of HC2 to other PCR-based tests with different primer sets the two tests performed equally well, with sensitivity in most studies ranging between approximately 80% and 90%. In one large, well-controlled study, a significant difference between the two tests was reported with HC2 achieving a higher sensitivity and PCR a higher specificity. There was little evidence of any difference in performance in the other studies, with wide and overlapping confidence intervals indicating both a lack of significant difference and a lack of precision in the test estimates. The levels of sensitivity and negative predictive value achieved across studies were approximately 80-90% and consistent for both HC2 and PCR-based tests while specificity and positive predictive value was lower and varied substantially between studies.

Two studies compared the performance of HPV DNA Chip and HC2 and the two tests performed very similarly across all test performance characteristics. Sensitivity was between approximately 85% and 95% and specificity between 40% and 50% with wide confidence intervals. One study compared HPV DNA Chip with a PCR-based test (GP5+/6+ primers) and reported a higher sensitivity for HPV DNA Chip, however the significance of these differences and other test characteristics were not reported. PreTect Proofer, a messenger-RNA test, was compared with HC2 in one study and PCR assays in two studies. HC2 was more sensitive but less specific than PreTect Proofer. Negative predictive values were not significantly different but positive predictive value was higher for PreTect Proofer. When compared with PCR-based assays both tests achieved the same or very similar levels of sensitivity (~85-95%) while PreTect Proofer was more specific (85% versus 50%, although specificity was only able to be calculated in one study). Negative predictive values were very similar (~98%) while positive predictive value was low (15-40%) and varied between the two tests, although this difference was not significant. PreTect Proofer only detects five high-risk HPV types and the authors of these papers suggested the sensitivity of PreTect Proofer could be increased with the addition of more high-risk virus types to the assay. There is however, a risk that this could decrease the specificity of the test and this would need to be assessed in further studies.

Overall there was little evidence of any consistent and significant differences in test performance between the HPV tests. The decision to implement one test and not another will depend on the relative importance assigned to test sensitivity, specificity, and positive and negative predictive values. With a high sensitivity but low specificity and positive predictive value, a situation which occurred in many of the comparisons, the detection rate of precancerous cervical abnormalities would be high but a high proportion of women would be referred for follow-up. This has not only cost implications in terms of
unnecessary procedures, but could also result in unnecessary stress for many HPV-positive women who were subsequently found to be free of disease.

While Amplicor, HPV DNA Chip and PreTect Proofer all show potential as reliable methods of HPV testing, none of these have been researched to the same degree as HC2 and other PCR assays. There is a need for large-scale well-controlled studies investigating the performance of these tests, in particular, to obtain more precise estimates of their test performance relative to other HPV testing methods. The most thoroughly investigated HPV testing methods are HC2 and PCR-based tests, which most often used the MY09/11 or GP5+/6+ primer sets. These two tests performed similarly within studies but test estimates did vary between studies and further large scale investigations would enable more precise test estimates to be calculated. The decision to implement either one of these tests will depend on more than test performance alone, and it may be that variances in cost, ease of administration and adaptability to high-volume processing will result in one test being given preference over the other.
### Table 6  Evidence table of appraised articles relating to clinical test performance of HPV testing

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test Reference Standard</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carozzi et al. (2007) Florence, Italy Florence Cervical Screening Programme Comparative cross-sectional study Level of Evidence III-2</td>
<td>N= 1032 samples from women screened as a part of the Florence Cervical Cancer Screening Programme. This group included 70 samples from subjects with histologically confirmed CIN2+ Collected using LBC and PreservCyt medium. Samples diagnosed as: Negative, ASCUS/AGCUS LSIL, HSIL, invasive SCC, invasive adenocarcinoma HPV tests: HC II (13 HR types) Amplicor (13 HR types) 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68 HC2 testing carried out at time of specimen collection. Amplificor testing carried out 2 years later. Reference standards: Colposcopy – women with ASCUS or SIL and those aged ≥ 35 years and HC II positive with any cytology (including negative) HC II positive and &lt; 35 years were referred for colposcopy if HPV positivity persisted for 1 year Histology following punch biopsy or loop resection was reference standard for presence of CIN2+.</td>
<td>Sensitivity, specificity Agreement (kappa) between HC II and AMPLICOR 96 samples were excluded because they were β-globin negative at the time of HPV testing 6 samples excluded because of inadequate cytology</td>
<td>HPV prevalence relative to histology: Histology (n=416)</td>
<td>HC2 AMP Negative (244) 165 124 CIN1 (79) 62 57 CIN2+ (93) 91 90 Initial cytology positive: N=270 ASCUS/SIL cases</td>
<td>Authors Conclusions HC2 and Amplicor performed similarly in the detection of confirmed CIN2+ lesions, clinical sensitivities were not significantly different. Amplicor specificity for CIN2+ lesions was higher but the difference was not statistically significant. Cases were selected partially on the basis of HC2 positivity, which would bias towards increased sensitivity and decreased specificity. When performance of the two tests was compared to cytology positive cases only, performance was similar. Limitations source population age range was described, however no information was provided regarding the age of the selected sample. Also no information regarding the selection criteria for included samples. HPV testing by Amplicor was conducted on samples which had been stored for 2 years. Blinding of cytologists and histologists to previous HPV test results or other laboratory results was not stated. No negative cytology samples were tested histologically prevalence of CIN2+ samples was increased by the addition of 70 histologically confirmed cases.</td>
</tr>
</tbody>
</table>
### Table 6  Evidence table of appraised articles relating to clinical test performance of HPV testing (continued).

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test</th>
<th>Reference Standard</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cuschieri et al. (2007) Edinburgh, Scotland Cohort study Level of Evidence III-2</td>
<td>N=322 samples collected from women attending for routine cytology screening Follow-up records available for n=190 women Inclusion criteria: borderline result from initial cytology and follow-up pathology available for at least 3 years following. 15 GP practices in Edinburgh LBC specimens collected and slides prepared using ThinPrep system. Cytological classifications made using British Society cyto guidelines</td>
<td>HPV tests: HC II AmpliCor Both able to detect 13 HR-HPV types: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68</td>
<td>Reference test: Women grouped on the basis of pathology records  Group 1: Cleared their abnormality without intervention  Group 2: Persistent low-grade disease (cytologically and/or histologically confirmed i.e. borderline changes or mild dyskaryosis and/or CIN1)  Group 3: Developed high-grade disease (histologically defined as CIN2+)</td>
<td>Sensitivity Specitivity</td>
<td>HPV positivity by each test Level of concordance between the two tests</td>
<td>Group 1 – cleared abnormal cytology 123/190 = 64.7% Group 2 – low-grade disease (borderline or mild dyskaryosis) 33/190 = 17.4% Group 3 – high-grade disease (CIN2+) 34/190 = 17.9% Overall HPV positivity: HC2+ = 219/322 = 68% AmpliCor+ = 235/322 = 73% p = 0.056 [not significant] Concordance: Agreement in 80.7% (76.0 – 84.9%) of cases HPV testing relative to high-grade (CIN2+) outcome 3 years later: HC2 Sensitivity = 82.4% Specificity = 44.2% AMPLICOR Sensitivity = 82.4% Specificity = 36.5% No significant difference in sensitivity or specificity.</td>
</tr>
</tbody>
</table>
Table 6. Evidence table of appraised articles relating to clinical test performance of HPV testing (continued).

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test Procedure</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schiffman et al. (2005) ALTS study U.S.A. Randomised clinical trial Level of Evidence II</td>
<td>ALTS – randomised trial comparing 3 triage strategies for women with ASCUS or LSIL N= 5060 women with ASCUS/LSIL cytological findings</td>
<td>Cytological procedure: 2 specimens collected 1) PreservCyt for ThinPrep cytological examination 2) Digene specimen transport medium Randomisation procedures presented in separate papers HPV DNA testing: HC2 Probe set B: 13 HR HPV types 16/18/31/33/35/45/51/52/56/58/59/68 Threshold: 1.0 RLU/PC HPV PCR testing: PGMY09/11 primer set AmpliFaq Gold polymerase PCR test with reverse line blot hybridisation used for type-specific confirmation 27 HPV genotypes – first 2000 women 6/11/16/18/26/31/33/35/39/44/52/56/58/59/66/68/73/82/83/84 38 HPV genotypes – remaining samples Above plus 11 noncarcinogenic genotypes 61/62/64/67/69-72/81/82/89 PCR results from each set of tests combined after they revealed similar performance on enrolment specimens.</td>
<td>ASCUS: N=3488 median age = 26 years range = 18-81 LSIL N=1572 median age = 23 range = 18-68 Routine follow-up visits every 6 months for 2 years HSIL cyto logicaly referred for colposcopy Exit exam and colposcopy for 80% of enrolled women Pathology: 4 clinical centres included in study Cytologic and histologic diagnoses made at those centres All referral smears, ThinPrep specimens and histologic slides were sent to the Pathology QC Group (Johns Hopkins Hospital) for review diagnoses.</td>
<td>Paired results for 5026 (99.3%) women. McNemar X² was used to test for statistical differences in the number of test-positives for paired results Clinical sensitivity and specificity of HC2 and PCR for each visit for the detection of CIN3+ by the QC Pathology Group and for CIN2+ clinical centre pathology diagnoses HPV test on enrolment and detection of CIN3 or cancer: HC2 (%) Se = 93.6 Sp = 41.2 PPV = 16.1 NPV = 98.2 PCR (%) Se = 89.3 Sp = 48.5 PPV = 17.3 NPV = 97.4 HPV test on enrolment and detection of CIN2+: HC2 (%) Se = 93.7 Sp = 44.6 PCR (%) Se = 87.2 Sp = 51.7</td>
<td>Authors conclusions Moderate to good agreement on test positivity when we compared HC2 with PCR detection of the 13 types targeted by HC2. HC2 demonstrated increased clinical sensitivity but lower specificity than PCR for the detection of cervical precancer. If clinical sensitivity and specificity are considered equally important aspects of overall accuracy, the 2 tests were similarly accurate. The addition of other potentially carcinogenic HPV types to PCR testing did not significantly improve the sensitivity of detection of CIN3 or cancer, while decreasing specificity. Inclusion of all types detected by PCR in the definition of a positive test approached but did not achieve the sensitivity of HC2 with a concomitant large decrease in specificity. Overall trend of lower sensitivity for either HPV assay for detection of cumulative cases of CIN3 or cancer or CIN2+ for testing performed on specimens from later visits. The reason for the decrease is unclear but we hypothesise that with intensive screening, only the smallest CIN3 lesions remain untreated and were more likely to be missed by any screening test.</td>
</tr>
</tbody>
</table>
Table 6. Evidence table of appraised articles relating to clinical test performance of HPV testing (continued).

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schiffman et al. (2005) continued</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>Diagnostic Test Reference Standard</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations and Conclusions</th>
</tr>
</thead>
</table>
| Main analyses: PCR results considered positive if they detected at least one of the 13 HR types included in the HC2 array | HC2 and PCR tests: Baseline and 4 follow-up specimens collected for each individual: Baseline/6 mths/12 mths/18 mths/24 mths | Excluding cases where clinicians knew HPV positivity or negativity made no difference to findings. | Limitations:  
• HPV tests were run on different aliquots from different specimen collections  
• HC2 was performed on PreservCyt and PCR on specimen transport medium samples, which were always collected second  
• HC2 testing was conducted in 4 laboratories and PCR in 1 lab, less clear how generalisable the PCR results are  
• population not representative of the entire popn of women with HPV infections – authors acknowledge and discuss this |
Table 6. Evidence table of appraised articles relating to clinical test performance of HPV testing (continued).

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test</th>
<th>Reference Standard</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Söderlund-Strand et al. (2005)</td>
<td>N=239 women with atypical Pap smear result and referred for colposcopy</td>
<td>Follow-up after initial Pap</td>
<td>Visit 1: Cytology – one sample collected with cytobrush Brush immersed in NaCl and frozen PCR test</td>
<td>Visit 2: N=162 HPV positive by PCR and HC2 N=49 negative by PCR and HC2 N=28 discordant results</td>
<td>HPV results from n = 177 women with histopathology results Histopathology +ve = CIN1+ Histopathology -ve = normal or koilocytosis</td>
<td>Authors conclusions The somewhat better sensitivity of PCR-EIA than HC2 for detection of CIN3, 100% vs. 95.6% respectively, was not due to differences in probe composition since only one more HPV type (HPV-66) is included in the PCR probe cocktail, and none of the discrepant samples was positive for HPV-66. We find similar sensitivities for CIN2+ detection by HC2 (91.8%) and PCR (92.9%). However the specificities were low for both HC2 (24.1%) and PCR (30.4%). It should be noted that the specificity refers to specificity in a clinical secondary screening setting where all women have had an abnormal smear and that the specificities in a primary screening setting would most likely be substantially better. The HC2 method is easy to use and commercially available and has therefore been recommended for routine screening use. Both HC2 and PCR appear adequate for routine use in CIN2+ detection in a secondary screening setting as well as for follow-up post-treatment.</td>
</tr>
</tbody>
</table>
Table 6. Evidence table of appraised articles relating to clinical test performance of HPV testing (continued).

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test Reference Standard</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Söderlund-Strand et al. (2005) continued</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HPV testing relative to CIN3 histology</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HC2 (%) 95% CI</td>
<td>Se = 95.6 (89.5 – 101.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sp = 18.9 (12.3 – 25.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPV = 28.7 (21.4 – 33.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NPV = 92.6 (82.7 – 102.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PCR (%) 95% CI</td>
<td>Se = 100.0 (100.0 – 100.0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sp = 23.5 (16.3 – 30.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPV = 30.8 (23.3 – 38.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NPV = 100.0 (100.0 – 100.0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Limitations</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• the blinding of cytotechnologists and pathologists to other laboratory results was not stated</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• confidence intervals were wide for specificity and NPV, lessening the accuracy of these estimates of test performance</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Further results in text for test performance in post-treatment women.</td>
<td></td>
</tr>
</tbody>
</table>
Table 6. Evidence table of appraised articles relating to clinical test performance of HPV testing (continued).

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations and Conclusions</th>
</tr>
</thead>
</table>
| Kulasingam et al. (2002) Washington state, U.S.A, Comparative cross-sectional study | N=4358 women invited to participate N= 4075 enrolled Mean age = 26 SD = 5.7 years N= 3310 < 30 N= 760 ≥ 30 | Cytology: Ayres spatula and cytobrush Brush and spatula tapped and rinsed in PreservCyt ThinPrep thin-layer cytology slide produced HPV testing dacron-tipped swab placed in standard transport medium (Digene) | Cytology: Thin-layer slide screened by a cytootechnologist and reviewed by pathologists having no knowledge of other laboratory or clinical data. Classifications based on Bethesda system: ASCUS/AGUS/LSIL/HSIL/suggestive of cancer Histology: Biopsy reviewed by pathologist blind to other laboratory or clinical data. Classifications based on Bethesda and CIN systems. HPV testing: PCR: MY09, MY11, HMBO1 primers and AmpliTaq Gold polymerase Types: 16/18/26/31/33/35/39/45/51/52/53/55/56/58/59/68/73/82/84 Analyses performed separately for women <30 years, ≥ 30 years and the whole sample combined Definition of a positive screening result based on criteria for referral for colposcopy (≥ ASCUS) Inadequate, insufficient, missing results classified as positive because these women would usually be asked to return for repeat smear Corrected sensitivities calculated based on likelihood of negative smear and HPV women accepting colposcopy increasing with perceived severity of screening results. Correction method in text. Screening results: 678 (16.6%) abnormal smear 747 (18.3%) positive PCR Prevalence estimate for HC2 = 27.4% Detection of CIN3+: Thin-layer + PCR: Se = 73.6% (uncorrected) Se = 59.8 (47.1 – 68.9) Sp = 89.8 (89.2 – 90.5) NPV = 98.5 | Authors conclusions Although a similar proportion of women with HSIL were HPV-positive by HC2 and PCR, more women with LSIIL, ASCUS or normal cytology were HPV-positive by HC2 than PCR. Limiting HPV testing to women aged 30-50 years resulted in better specificity. 
Reviewers comments:
- very high participation rate (94%) a strength
- 7.7% of negative cytology women also tested for HPV status and underwent a colposcopy
- blinding adequate. 
Limitations:
- average age of sample young compared to some studies, with 81% < 30 years – so population may not be representative of general HPV positive population
- HC2 testing was not completed in a proportion of the sample but corrected test performance was calculated to account for this. |
Table 6. Evidence table of appraised articles relating to clinical test performance of HPV testing (continued).

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test</th>
<th>Reference Standard</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kulasingam et al. (2002) continued</td>
<td>Average delay between smear and colposcopy = 3 months</td>
<td>HC2: Types: 16/18/31/33/35/39/45/51/52/58/59/68</td>
<td>Data analysis: Estimates of sensitivity, specificity for detection of CIN3+ and of the percentage of women referred for colposcopy were obtained for 7 screening strategies.</td>
<td>Thin-layer + HC2: Se = 73.6 (uncorrected) Sp = 88.9 (88.1 – 89.4) NPV = 98.5 Prediction of CIN3+ by age:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Colposcopy visit: 1. cytology and HPV samples obtained 2. colposcopy 3. biopsy</td>
<td>Data analysis: Estimates of sensitivity, specificity for detection of CIN3+ and of the percentage of women referred for colposcopy were obtained for 7 screening strategies.</td>
<td>Thin-layer + HC2: Se = 73.6 (uncorrected) Sp = 88.9 (88.1 – 89.4) NPV = 98.5 Prediction of CIN3+ by age:</td>
<td>Thin-layer + HC2: Se = 73.6 (uncorrected) Sp = 88.9 (88.1 – 89.4) NPV = 98.5 Prediction of CIN3+ by age:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data analysis: Estimates of sensitivity, specificity for detection of CIN3+ and of the percentage of women referred for colposcopy were obtained for 7 screening strategies.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6. Evidence table of appraised articles relating to clinical test performance of HPV testing (continued).

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test Reference Standard</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morin et al (2001) Quebec, Canada Comparative cross-sectional study Level of Evidence III-2</td>
<td>N = 360 consecutive women with a diagnosis of ASCUS and referred for colposcopy 92% participation rate</td>
<td>Referred for colposcopy and repeat smear Directed biopsies on all lesions of the cervix. Cytology: Classified according to Bethesda system. Histology: CIN classifications Prepared and viewed by the same pathologist HPV testing: Sample technique – Dacron swab immersed in Digene Specimen Transport Buffer and frozen HC1: 9 HR types 16/18/31/33/35/45/51/52/56 HC2 13 HR types 16/18/31/33/39/45/51/52/56/58/59/68 PCR 11 HR types 16/18/31/33/39/45/51/52/56/58</td>
<td>Case control study: Cases = 19 women with CIN 2/3 diagnosis following colposcopy and biopsy Controls = 341 women without CIN 2/3 Compared three groups of strategies for predicting CIN 2/3 following an abnormal smear: 1) repeat abnormal smear [presence of ASCUS/LSL/HSIL] 2) positive HPV test by HC1 alone, HC2 alone and PCR alone 3) combination of repeat smear and/or positive HPV result se, sp. of each strategy</td>
<td>Prediction of CIN 2/3 in women with an abnormal smear result: HC2 % (95% CI) Se = 89.5 (66.9 – 98.7) Sp = 74.1 (68.9 – 78.5) PPV = 16.0 (9.6 – 24.4) NPV = 99.2 (97.2 – 99.9) Index of performance = 0.66 Referred for colposcopy =29.2 (24.8 – 34.5) PCR % (95% CI) Se = 89.5 (66.9 – 98.7) Sp = 59.0 (53.4 – 64.3) PPV = 11.2 (6.7 – 17.3) NPV = 99.0 (96.4 – 99.9) Index of performance = 0.53 Referred for colposcopy =44.0 (38.4 – 49.1) HC1 % (95% CI) Se = 68.4 (43.5 – 87.4) Sp = 85.9 (81.8 – 89.4) PPV = 21.3 (11.9 – 33.7) NPV = 98.0 (95.7 – 99.3) Index of performance = 0.59 Referred for colposcopy =16.9 (13.2 – 21.2) Data also available for repeat smear strategy and also repeat smear and/or positive HPV result strategy</td>
<td>Authors conclusions The performance of HPV testing as a strategy for referring women with ASCUS was better than the strategy based on repeat smear alone. Among the three different HPV tests, HC1 was the least sensitive; however this difference was not significantly different from HC2 or PCR, probably because of the small number of CIN cases. HC2 and PCR presented the same sensitivity, missing about 10% of women with high grade CIN. Based on the index of performance measure, HC1 and HC2 would be better than PCR for detecting women with CIN2/3. Suggests HC could be used in the triage of women with an ASCUS smear. HC appears more useful than PCR for triage. Analytic sensitivity of HC3 can be adjusted by determining the threshold level of HPV DNA copies per assay to be detected. Specificity of PCR was much lower – high number of false positives.</td>
</tr>
</tbody>
</table>

Limitations
- wide confidence intervals
- small number of cases of CIN2 or CIN3 (n=19) most likely contributed to the low study power and the lack of a significant difference in sensitivity and PPV is not surprising
- reviewers conclusions
- specificity of PCR significantly lower than HC2 or HC1
- sensitivity of HC2 and PCR the same and higher than HC1
- PPV low for both HC2 and PCR
- NPV high for both HC2 and PCR.
Table 6. Evidence table of appraised articles relating to clinical test performance of HPV testing (continued).

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test Reference Standard</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yamazaki et al. (2001) Hokuriku, Japan Case-control study nested in a cohort study Level of Evidence III-2</td>
<td>N= 286 women out of 308 selected from a cervical cancer screening programme because of abnormal cytology Inclusion criteria: Histologically confirmed LSL/HSIL /invasive cancer. Exclusion criteria: Equivocal cytological findings.</td>
<td>Cytology: Cytobrush from the ectocervix and endocervix. Samples collected for Pap and HPV tests. HPV samples stored in PBS and frozen. 1 ml resuspended in sample solution (Digene) for HC2. Screened by one cytotechnologist. HPV testing: HC2 – LR and HR probes used HR RLU cut-off of 1.0 RLU considered positive</td>
<td>Cases = 286 women Controls = women with no current evidence of cervical neoplastic lesions and STDs Matched for age from the same population as the cases HPV detection and histological diagnoses were performed independently.</td>
<td>Prevalence of HPV infection in normal/LSIL/HSIL/SCCs and ADCAs : HC2: Normal = 6% LSIL = 6% HSIL = 8% SCC = 83% ADCA = 75% PCR: Normal = 12% LSIL = 78% HSIL = 94% SCC = 93% ADCA = 88%</td>
<td>Authors conclusions In screening of HSILs and ICC, the estimated sensitivity of LCR-E7 PCR was 81% and HC2 was 83%, while the specificity was 93% for both. Higher specificity may be because the control sample excluded any women with any history of STDs. When cytologically normal and LSIL women were counted as disease-free, specificity dropped to 71% for PCR and 67% for HC2. The present results show the usefulness of HC2 and LCR-E7 PCR in screening of HSILs and ICC. We could not estimate true sensitivity or specificity since a few women with disease may have been missed by cytologic screening. We calculated estimated test performance instead. Limitations • very low participation rate (31%) lessens the generalisability of the test performance estimates • test estimates initially calculated including LSIL women as positive histology cases. This may have increased specificity estimates and when LSIL women were included in the disease-free control group, specificity for PCR and HC2 decreased.</td>
</tr>
</tbody>
</table>

---

EFFECTIVENESS OF HPV SCREENING IN WOMEN WITH ABNORMAL OR EQUIVOCAL CERVICAL SMEAR RESULTS
### Table 6. Evidence table of appraised articles relating to clinical test performance of HPV testing (continued).

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test</th>
<th>Reference Standard</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations and Conclusions</th>
</tr>
</thead>
</table>
| Bergeron et al. (2000)                                   | N= 404 women diagnosed with ASCUS/LSIL/HSIL from a sample of 1037 | Repeat cytology: Cytobrush | PCR MY09/MY11 primers | Smears and biopsies evaluated by one pathologist blind to results of other tests. | 26 biopsy specimens excluded because of inadequate samples | Authors conclusions PCR and HC2 showed good concordance and similar sensitivities for detecting CIN. Rest of authors’ discussion focuses on performance of HC2 relative to cytology. Limitations  
- participation rate was very low (39%)  
- of the women who agreed to participate, complete test results were available for 79%  
- the calculations of positive predictive value in the prediction of high-grade CIN were limited by a very small number of cases (n=22). |
| Paris, France, Comparative cross-sectional study         | Participation rate = 39% | HPV tests: HC2 | Repeat cytology: Cone brush; sample placed in Digene specimen transport medium | Cytology classified according to Bethesda system. | N= 378 women with HC2 and biopsy results |  |
| Level of Evidence III-2                                 | Mean age = 35 | HR probe B – 16/18/31/33/35/39/45/51/52/56/58/59/68 | Blinded hybridisation | Results classified as: | N= 520 women with HC2, PCR, Southern blot and biopsy results |  |
|                                                        | Range 15 – 75 | Threshold: 1.0 pg/ml | Reference standard: Colposcopy - all women (n=404) Biopsy specimen taken from transformation zone (N=384) | Outcomes: Concordance between HC2 and PCR or Southern blot hybridisation results | Sensitivities and specificities of repeat cytology, HPV triage and combined triage |  |
Table 6. Evidence table of appraised articles relating to clinical test performance of HPV testing (continued).

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test</th>
<th>Reference Standard</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bergeron et al. (2000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>continued</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Test performance relative to high-grade CIN:
  - HC2 (% 95% CI)
    - Se = 86.4 (72.0 – 100.7)
    - Sp = 40.9 (35.4 – 46.5)
    - PPV = 9.74 (5.6 – 13.9)
    - NPV = 97.6 (94.9 – 100.3)
  - PCR (% 95% CI)
    - Se = 95.5 (86.8 – 104.2)
    - Sp = 39.9 (34.4 – 45.5)
    - PPV = 10.5 (6.3 – 14.8)
    - NPV = 99.2 (97.5 – 100.8)

Results also available for ASCUS women separately and LSIL women separately.
### Table 6. Evidence table of appraised articles relating to clinical test performance of HPV testing (continued).

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test Reference Standard</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations</th>
<th>Authors’ conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim et al. (2003) Seoul, South Korea Comparative cross-sectional study Level of Evidence III-2</td>
<td>The study subjects were tested in the Department of Obstetrics and Gynaecology, the Catholic University of Korea, March 1999-February 2000. Overall n=140 and by cytology Normal/benign cellular change, n=59 ASCUS, n=33 LSL, n=15 HSL, n=33 Inclusion criteria Patients who had undergone previous HPV testing Exclusion criteria None specifically stated Overall mean age: 39.1 years (range 23-68) Mean age (range) Group I: 40 (23-56) Group II: 37 (23-61) Group III: 40 (25-68)</td>
<td>HPV oligonucleotide microarray system (HPV-oms) Hybrid Capture 2 (HC2) Conventional cytology/histology HPV types HPV-oms: 15 high risk types (HPV 16/18/31/33/35/39/45/51/52/56/58/59/66/68/69) and 7 types of low risk (HPV 6/11/34/40/42/43/44) HC2: 13 high risk types (HPV 16/18/31/33/35/39/45/51/52/56/58/59/68) Sampling technique Smears were taken using a cytobrush and a scrape for HC2. Specimens were suspended in 1ml of transport medium and stored until use.</td>
<td>Histological pathology reported for Group I: Non specific chronic cervicitis Group II: Koliocytosis, mild dysplasia Group III: Moderate dysplasia, severe dysplasia, carcinoma in situ Sensitivity (se) Specificity (sp) Positive predictive value (PPV) Negative predictive value (NPV) Statistical analysis was performing using chi-square tests and multiple logistic regression. Detection of HPV in index cytology and testing Group I (n=33) Group II (n=39) Group III (n=40)</td>
<td>Statistical analysis was performing using chi-square tests and multiple logistic regression. Detection of HPV in index cytology and testing Group I (n=33) Group II (n=39) Group III (n=40)</td>
<td>Authors’ conclusions Results suggest oligonucleotide microarray system is highly comparable to HC2 for detecting HPV in cervical specimens. Reviewers Comments ▪ two discordant cases were produced by the different methods 1) HC (-), HPV-oms-HPV-18 (+) and 2) HC (+), HPV-oms (-). Limitations ▪ blinding or masking of cytologists to other test results was not stated ▪ no information provided regarding whether sample selection was consecutive or a random selection.</td>
<td></td>
</tr>
</tbody>
</table>
### Table 6. Evidence table of appraised articles relating to clinical test performance of HPV testing (continued)

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test Reference Standard</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations Authors’ conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim et al. (2003) continued</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>relatively wide confidence intervals and, therefore, less accuracy in the estimates of specificity and negative and positive predictive values.</td>
</tr>
</tbody>
</table>

**HPV positivity by histological outcome:**

<table>
<thead>
<tr>
<th>Histology</th>
<th>HPV %</th>
<th>HPV %</th>
<th>HPV %</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronic cervicitis (n=61)</td>
<td>30 (49.2)</td>
<td>31 (50.8)</td>
<td>1 (1.6)</td>
<td></td>
</tr>
<tr>
<td>Koilocytosis (n=27)</td>
<td>24 (88.9)</td>
<td>23 (85.2)</td>
<td>2 (7.4)</td>
<td></td>
</tr>
<tr>
<td>Mild dysplasia (n=12)</td>
<td>12 (100)</td>
<td>12 (100)</td>
<td>1 (8.3)</td>
<td></td>
</tr>
<tr>
<td>Mod. Dysplasia (n=7)</td>
<td>7 (100)</td>
<td>7 (100)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Severe dysplasia (n=10)</td>
<td>10 (100)</td>
<td>10 (100)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Carcinoma in situ (n=23)</td>
<td>22 (95.7)</td>
<td>22 (95.7)</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

**HPV Test Prediction of low-grade SIL or worse:**

<table>
<thead>
<tr>
<th>HC2 (%)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Se</td>
<td>94.9</td>
</tr>
<tr>
<td>Sp</td>
<td>50.8</td>
</tr>
<tr>
<td>PPV</td>
<td>71.4</td>
</tr>
<tr>
<td>NPV</td>
<td>88.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HPV -oms (%)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Se</td>
<td>93.7</td>
</tr>
<tr>
<td>Sp</td>
<td>49.2</td>
</tr>
<tr>
<td>PPV</td>
<td>70.5</td>
</tr>
<tr>
<td>NPV</td>
<td>85.7</td>
</tr>
</tbody>
</table>
### Table 6. Evidence table of appraised articles relating to clinical test performance of HPV testing (continued).

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test</th>
<th>Reference Standard</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hwang et al. (2003) Incheon and Seoul, South Korea</td>
<td>Samples from the study subjects were collected in the Department of Obstetrics and Gynaecology, Inha University Hospital, Incheon, South Korea. Overall n=234 samples By histology, n=212 cases Chronic cervicitis (n=43), CIN I (n=24), CIN II (n=22), CIN III (n=51), Cancer (n=72) Without diagnosis (n=22)</td>
<td>HPV oligonucleotide microarray system (HPV-oms) PCR-RFLP (GDP5+/GDP6D+)</td>
<td>Conventional cytology and histology HPV types HPV-oms: 15 high risk types (HPV 16/18/31/33/35/39/45/51/52/56/58/59/66/68/69) and 7 types of low risk (HPV 6/11/34/40/42/43/44) PCR-RFLP: 7 types of HPV 16/18/31/33/35/52b and 58</td>
<td>Sensitivity (se) Test-retest reliability Kappa statistics and sensitivity were reported. Distribution of HPV by histological diagnosis Cervicitis CIN I CIN II CIN III Cancer N 43 24 22 51 72 HPV+ (%) HPV+(%) HPV + (%) HPV + (%) HPV-oms 16 (37.2) 17 (70.8) 17 (77.3) 42 (82.4) 65 PCR-RFLP 9 (20.9) 12 (50.0) 15 (68.2) 39 (76.5) 64 AC 17 (39.5) 19 (79.2) 18 (81.8) 41 (80.4) 63 AC/PV 22 (51.2) 21 (87.5) 21 (95.5) 49 (96.1) 70 AC/oms 17 (39.5) 19 (79.2) 18 (81.8) 41 (80.4) 63 Where AC abnormal cytology contains ASCUS, LSIL, HSIL, cancer and AC/PV-oms abnormal cytology or HPV microarray. -Sensitivity of diagnostic screening efficacy by HPV-oms CIN I 70.8 CIN II 77.3 CIN III 82.4 Cancer 90.3 PCR-RFLP CIN I 50.0 CIN II 68.2 CIN III 76.5 Cancer 88.9</td>
<td>Authors' conclusion Results suggest that HPV oligonucleotide microarray is a highly comparable method to PCR-RFLP methods and by combining with cytology, sensitivity improves. Comments a total of 22 patients refused biopsy and wanted to be followed by pap smear and HPV testing the kappa index between tests for detecting HPV ranged from good to excellent depending on the type of HPV Limitations blinding or masking of cytologists to other test results was not stated statistical methodology used was not described although test-retest statistics were reported cannot ascertain true test performance because of a lack of information about test specificity.</td>
<td></td>
</tr>
</tbody>
</table>
Table 6. Evidence table of appraised articles relating to clinical test performance of HPV testing (continued).

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test</th>
<th>Reference Standard</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hwang et al. (2003) Continued</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>By combining cytology and HPV testing, the detection rate was improved to 87.5, 95.5, 96.1 and 97.2% in CIN 1, CIN II, CIN III and cancer respectively</td>
</tr>
</tbody>
</table>
### Table 6. Evidence table of appraised articles relating to clinical test performance of HPV testing (continued).

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test</th>
<th>Reference Standard</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee et al. (2005) Gwangju, Korea Comparative cross-sectional study Level of Evidence III-2</td>
<td>The study subjects were tested in the Department of Obstetrics and Gynaecology, the Chonnam National University Hospital (CNUH). Overall n=400 By histology Cervicitis (n=132) CIN I (n=59) CIN II (n=42) CIN III (n=114) Cancer (n=53) Inclusion criteria Patients who had abnormal cervical cytology or cervicogram. Exclusion criteria None stated. Mean age of participants not reported.</td>
<td>Hybrid Capture II (HC2) HPV DNA chip (HD-C) ThinPrep cytology and histology HPV types HC2: 13 high risk types (HPV 16/18/31/33/35/39/45/51/52/56/58/59/68) HD-C: 15 high risk types (HPV 16/18/31/33/35/39/45/51/52/56/58/59/66/68/69) and 7 types of low risk (HPV 6/11/34/40/42/43/44) Sampling technique The first clinical test performed was cytology, followed by the HPV tests, colposcopy, and a biopsy. Some specimens used for the HD-C tests were those samples kept frozen after ThinPrep cytology.</td>
<td>Histological pathology reported for Cervicitis, CIN I, CIN2, CIN3, and cancer Sensitivity (se) Specificity (sp) Positive predictive value (PPV) Negative predictive value (NPV) Results Statistically significant differences [p-values] were reported. Association HPV (+ve) and histopathology by cytology (%) Cervicitis CIN I CIN II CIN III Cancer Cytology Cancer HSIL 6 2 18 68 7 LSIL 21 44 12 14 AGUS 6 2 2 1 ASCUS 27 6 2 16 Normal/BCC 72 5 10 14 2 Distribution of HPV by histological diagnosis HC2 HPV+ (%): Cervicitis = 75 (56.8) CIN1 = 54 (91.3) CIN2 = 37 (88.1) CIN3 = 101 (88.6) Cancer = 49 (92.5) HD-C HPV+ (%): Cervicitis = 71 (53.8) CIN1 = 54 (91.5) CIN2 = 34 (81.0) CIN3 = 96 (84.2) Cancer = 47 (88.7)</td>
<td>Authors’ conclusion Results suggest the possibility of using the HC2 and HD-C as screening tests, which have similar sensitivity as the ThinPrep cytology.</td>
<td>Comments HD-C test was more cost-effective than the HC2 Limitations blinding or masking of cytologists to other test results was not stated statistical methodology used was not described no information provided regarding sample selection, that is, whether it was a consecutive group of women or a random selection.</td>
<td></td>
</tr>
</tbody>
</table>
Table 6. Evidence table of appraised articles relating to clinical test performance of HPV testing (continued).

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test Reference Standard</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>continued</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee et al. (2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**HPV test performance for prediction of CIN1 or worse**

<table>
<thead>
<tr>
<th>Diagnostic Test</th>
<th>(%)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>HC2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Se</td>
<td>89.9</td>
<td>86.3 – 93.5</td>
</tr>
<tr>
<td>Sp</td>
<td>43.2</td>
<td>34.7 – 51.6</td>
</tr>
<tr>
<td>PPV</td>
<td>76.3</td>
<td>71.6 – 81.0</td>
</tr>
<tr>
<td>NPV</td>
<td>67.9</td>
<td>57.9 – 77.8</td>
</tr>
</tbody>
</table>

**HPV test performance for prediction of CIN2 or worse**

<table>
<thead>
<tr>
<th>Diagnostic Test</th>
<th>(%)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD-C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Se</td>
<td>86.2</td>
<td>82.1 – 90.3</td>
</tr>
<tr>
<td>Sp</td>
<td>46.2</td>
<td>37.7 – 54.7</td>
</tr>
<tr>
<td>PPV</td>
<td>76.5</td>
<td>71.7 – 81.3</td>
</tr>
<tr>
<td>NPV</td>
<td>62.2</td>
<td>52.7 – 71.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Diagnostic Test</th>
<th>(%)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD-C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Se</td>
<td>84.7</td>
<td>79.8 – 89.6</td>
</tr>
<tr>
<td>Sp</td>
<td>34.6</td>
<td>27.8 – 41.3</td>
</tr>
<tr>
<td>PPV</td>
<td>58.6</td>
<td>53.1 – 64.2</td>
</tr>
<tr>
<td>NPV</td>
<td>67.4</td>
<td>58.1 – 76.6</td>
</tr>
</tbody>
</table>
### Table 6. Evidence table of appraised articles relating to clinical test performance of HPV testing (continued).

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test</th>
<th>Reference Standard</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lie et al. (2005) Norway Comparative cross-sectional study Level of evidence II-2</td>
<td>5 hospital-based outpatients clinics, N=383 women</td>
<td>Age: Median = 35 years Range 19-85 years</td>
<td>Women with positive cytology referred for colposcopy or conization</td>
<td>Sampling technique: Cervex-brush Conventional Pap smear HPV testing- Remaining sample transferred to PreservCyt vial</td>
<td>HPV Testing: Hybrid Capture 2 – 13 HR types 16/18/31/33/35/39/45/51/52/56/58/59/68 Threshold: 1.0 pg of HPV DNA/ml PreTect HPV-Proofer assay (NorChip) 5 HR types: 16/18/31/33/45 E6/E7 mRNA PCR type-specific testing: conducted in 40 cases of low-grade benign histology and discordant HPV results and 59 cases of high-grade histology (CIN2+) and discordant HPV results and 7 cases of high-grade histology and negative HPV results</td>
<td>Pap smears evaluated by cytotecnicon blind to HPV and histology results Positive smears classified by one cytopathologist according to Bethesda system criteria Histology evaluated blindly by one pathologist based on colposcopy-directed biopsies and/or cone specimens.</td>
</tr>
</tbody>
</table>
### Table 6. Evidence table of appraised articles relating to clinical test performance of HPV testing (continued).

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test</th>
<th>Reference Standard</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molden et al. (2005b) Oslo, Norway Prospective Cohort study Level of evidence III-2</td>
<td>Subgroup from 4136 women in the Norwegian Cervical Cancer Screening Program that visited a selection of gynaecologists in Oslo, Norway and were tested in 2001. N=77 were followed up for 24 months.</td>
<td>PreTect HPV-Proofer (HPV E6/E7 viral RNA) Gp5+6+ consensus PCR (DNA detection) Conventional cytology/histology</td>
<td>HPV types PreTect: 5 carcinogenic types 16, 18, 31, 33 and 45 Gp5+6+: over 20 types</td>
<td>Histological CIN2+ lesions Reported in this study</td>
<td>Detection of HPV in index cytology</td>
<td>Authors’ conclusion With equal sensitivity and higher specificity than consensus PCR, the PreTect HPV-Proofer test might offer an improvement for the triage of women with ASCUS or LSIL Pap smear.</td>
</tr>
<tr>
<td></td>
<td>Ages: Sample were all &gt; 30 years Mean age: 46.4 and 47.5 years for the index ASCUS and LSIL groups</td>
<td>Sensitivity (se) Specificity (sp) Positive predictive value (PPV) Negative predictive value (NPV)</td>
<td></td>
<td></td>
<td>ASCUS (n=57) PreTect 12 (21.1%) 6 (30%) Gp5+6+ 27 (47.4%) 15 (75%)</td>
<td>Detection by histology within 2-year follow-up No Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Comments</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>▪ four women without a subsequent Pap smear were excluded.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Limitations</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>▪ the confidence intervals for all test characteristics except NPV were very wide, lessening the accuracy of the estimates</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>▪ the confidence interval for OR on PreTect Proofer very wide indicating lower precision for this estimate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>▪ cytology and histologic diagnoses were carried out by different pathologists at each laboratory. There was no expert review of diagnoses</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>▪ blinding or masking of cytologists to other test results was not stated.</td>
</tr>
</tbody>
</table>

Detection of HPV in index cytology
- ASCUS (n=57)
- PreTect +ve 43 (82.7%) 0 (0.0%)
- -ve 9 (17.3%) 2 (100%)
- Gp5+6+ +ve 29 (55.8%) 0 (0.0%)
- -ve 23 (44.2%) 2 (100%)
- Total 52 (100%) 2 (100%)

Detection by histology within 2-year follow-up
- ASCUS (n=19)
- PreTect +ve 13 (92.9%) 1 (20%)
- -ve 1 (7.1%) 4 (80%)
- Gp5+6+ +ve 4 (28.6%) 1 (20%)
- -ve 10 (71.4%) 4 (80%)
- Total 14 (100%) 5 (100%)

Authors’ conclusion With equal sensitivity and higher specificity than consensus PCR, the PreTect HPV-Proofer test might offer an improvement for the triage of women with ASCUS or LSIL Pap smear.
Table 6. Evidence table of appraised articles relating to clinical test performance of HPV testing (continued).

<table>
<thead>
<tr>
<th>Source, Country, Setting, Study Design, Evidence Grading</th>
<th>Sample</th>
<th>Diagnostic Test Reference Standard</th>
<th>Outcomes and verification</th>
<th>Results</th>
<th>Limitations and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kraus et al. (2006) Oslo, Norway Comparative cross-sectional study Level of Evidence III-2</td>
<td>The study subjects were Norwegian women n=204 Inclusion criteria Women with histological confirmed invasive squamous cell carcinomas. No mention of original cytological findings Exclusion criteria None stated, Mean age of participants was not reported.</td>
<td>PreTect HPV-Proofer (HPV E6/E7 viral RNA) Gp5+/6+ consensus PCR (DNA detection) Histology</td>
<td>For comparison HPV DNA was detected by consensus and type specific PCR, PCR immunoassay (IA) and reverse line blotting (RLB) and in situ hybridization. HPV types PreTect: 5 carcinogenic types 16, 18, 31, 33 and 45 Gp5+/6+: over 20 types Sampling technique Method for collecting specimens not reported.</td>
<td>Sensitivity (se) Detection of HPV by PreTect Proofer and PCR PreTect Se 88.7 Gp5+/6+ Se 91.7</td>
<td>Authors’ conclusion PreTect HPV Proofer has high agreement with other detection methods and is a valuable approach for detection of cervical carcinoma. Comments ▪ all women in the sample had confirmed invasive cancer so specificity, PPV, NPV can not be calculated. Limitations ▪ blinding or masking of cytologists to other test results was not stated ▪ a major limitation is the lack of data to calculate specificity and other test characteristics. True test performance cannot be ascertained from this study ▪ sample selection method and participation rate not stated ▪ no mention made of original cytology diagnosis or history of cytology or histology diagnoses.</td>
</tr>
</tbody>
</table>
REFERENCES


## APPENDIX 1: LEVELS OF EVIDENCE

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level I</td>
<td>Evidence obtained from a systematic review (or meta-analysis) of relevant randomised controlled trials.</td>
</tr>
<tr>
<td>Level II</td>
<td>Evidence obtained from at least one randomised controlled trial.</td>
</tr>
<tr>
<td>Level III</td>
<td>Evidence obtained from pseudorandomised controlled trials (alternate allocation or some other method).</td>
</tr>
<tr>
<td></td>
<td>2 Evidence obtained from comparative studies (including a systematic reviews of such studies) with concurrent controls and allocation not randomised, cohort studies, case control studies or interrupted time series with a control group.</td>
</tr>
<tr>
<td></td>
<td>3 Evidence obtained from comparative studies with historical control, two or more single-arm studies or interrupted time series without a parallel control group.</td>
</tr>
<tr>
<td>Level IV</td>
<td>Evidence obtained from case series, either post-test or pretest/post-test.</td>
</tr>
</tbody>
</table>
APPENDIX 2: SEARCH STRATEGIES

**Medline**

1. exp papillomavirus infections/ (7145)
2. hpv.mp. (7925)
3. papilloma$.mp. (14673)
4. or/1-3 (16389)
5. atypical squamous cells.tw. (614)
6. (ASCUS or "ASC-US" or "ASC-H").mp. (634)
7. (LSIL or HSIL).mp. (625)
8. (squamous intraepithelial or squamous intra-epithelial).mp. (1448)
9. (cin-1 or cin-2 or cin-3).mp. (450)
10. (cin1 or cin2 or cin3).mp. (306)
11. cervical intraepithelial neoplasia/ (3393)
12. dyskaryosis.mp. (150)
13. (low grade or high grade).tw. (21329)
14. or/5-13 (23987)
15. 4 and 14 (2633)
16. di.fs. (584155)
17. exp Reagent Kits, Diagnostic/ (5818)
18. linear array.mp. (622)
19. amplicor.af. (949)
20. cytyc.af. (104)
21. roche.af. (5933)
22. imaging system.mp. (2020)
23. hybrid capture.af. (552)
24. biotools.af. (6)
25. real time assay.mp. (86)
26. screening assay.mp. (679)
27. surepath.af. (34)
28. thimprep.af. (336)
29. GenID.af. (0)
30. exp Polymerase Chain Reaction/ (168768)
31. or/16-30 (746882)
32. 15 and 31 (1402)
33. limit 32 to yr=2000-2007 (1076)
34. limit 33 to English (987)
35. (news or letter).pt. (348798)
36. 34 not 35 (958)
37. human/ (3973085)
38. animal/ (1485268)
39. 38 not (37 and 38) (996951)
40. 36 not 39 (957)
41. "sensitivity and specificity"/ (134620)
42. "predictive value of test"/ or roc curve/ (62306)
43. comparative study.pt. (597054)
44. (sensitiv$ or specific$ or comparative or comparison).tw. (1115185)
45. (compare or compared).tw. (787779)
46. (test$ adj3 performance).tw. (6792)
47. accuracy.tw. (64625)
48. diagnostic errors/ or false negative reactions/ or false positive reactions/ (17750)
49. (ppv or npv or positive predictive value or negative predictive value).tw. (13061)
50. likelihood functions/ (7650)
51. likelihood ratio$.tw. (2893)
52. reference standards/ (11590)
53. (detection adj3 rate$).tw. (5321)
54. or/41-53 (2012204)
55. 40 and 54 (660)
Embase

1. exp papilloma virus/ or human papillomavirus type 11/ or human papillomavirus type 16/ or human papillomavirus type 18/ or human papillomavirus type 33/ (11344)
2. (papillomavirus or papilloma virus or hpv).tw. (10564)
3. 1 or 2 (12928)
4. squamous epithelium/ (1504)
5. atypical squamous cells.tw. (621)
6. (ASCUS or "ASC-US" or "ASC-H").tw. (624)
7. (LSIL or HSIL).tw. (646)
8. (squamous intraepithelial or squamous intra-epithelial).tw. (1474)
9. (cin-1 or cin-2 or cin-3 or cin1 or cin2 or cin3).tw. (744)
10. (cervical intraepithelial neoplasia or cervical intra-epithelial neoplasia).tw. (1810)
11. dyskaryosis.tw. (159)
12. (low grade or high grade).tw. (21319)
13. or/4-12 (24450)
14. di.fs. (842059)
15. diagnostic kit/ (1511)
16. linear array.mp. (595)
17. ampliclor.af. (923)
18. cytyc.af. (112)
19. roche.af. (4823)
20. hybrid capture.af. (554)
21. biotools.af. (6)
22. surepath.af. (39)
23. thinprep.af. (352)
24. GenID.af. (0)
25. real time assay.mp. (85)
26. screening assay.mp. (676)
27. imaging system.tw. (1828)
28. polymerase chain reaction/ or real time polymerase chain reaction/ or reverse transcription polymerase chain reaction/ (176190)
29. or/14-28 (994109)
30. 3 and 13 and 29 (1545)
31. limit 30 to yr=2000-2007 (1166)
32. limit 31 to english (1096)
33. letter.pt. (233061)
34. 32 not 33 (1083)
35. "sensitivity and specificity"/ (33814)
36. (positive predictive value or negative predictive value or ppv or npv).mp. (12890)
37. Comparative Study/ (74391)
38. (sensitiv$ or specific$ or comparative or comparison).tw. (1079473)
39. diagnostic accuracy/ (88313)
40. diagnostic error/ (13149)
41. false negative result/ or false positive result/ (1562)
42. predictive validity/ or predictive value.tw. (19912)
43. Roc Curve/ (1224)
44. likelihood ratio5.mp. (2724)
45. likelihood function5.mp. (180)
46. reference standard5.tw. (2359)
47. (detection adj3 rate$).tw. (5094)
48. (test$ adj3 performance).tw. (6712)
49. or/35-48 (1195468)
50. 34 and 49 (554)

Cinahl

1. exp papillomavirus infections/ (1334)
2. hpv.mp. (500)
3. papilloma8.mp. (1358)
4. or/1-3 (1762)
5. atypical squamous cells.tw. (62)
6. (ASCUS or "ASC-US" or "ASC-H").mp. (55)
7. (LSIL or HSIL).mp. (39)
8. (squamous intraepithelial or squamous intra-epithelial).mp. (111)
9. (cin-1 or cin-2 or cin-3).mp. (30)
10. (cin1 or cin2 or cin3).mp. (14)
11. cervical intraepithelial neoplasia/ (280)
12. dyskaryosis.mp. (6)
13. (low grade or high grade).tw. (715)
14. or/5-13 (975)
15. 4 and 14 (214)
16. df.fs. (92921)
17. exp Reagent Kits, Diagnostic/ (448)
18. linear array.mp. (19)
19. ampicor.af. (50)
20. cytnc.af. (19)
21. roche.af. (2212)
22. imaging system.mp. (460)
23. hybrid capture.af. (39)
24. biotools.af. (0)
25. real time assay.mp. (2)
26. screening assay.mp. (46)
27. surepath.af. (5)
28. thinprep.af. (48)
29. GenID.af. (0)
30. exp Polymerase Chain Reaction/ (2964)
31. or/16-30 (97531)
32. 15 and 31 (145)
33. limit 32 to yr=2000-2007 (138)
34. limit 33 to english (138)
35. (news or letter).pt. (41619)
36. 34 not 35 (131)
37. human/ (0)
38. animal/ (890)
39. 38 not (37 and 38) (890)
40. 36 not 39 (131)
41. "sensitivity and specificity"/ (9694)
42. "predictive value of tests"/ or roc curve/ (5394)
43. (sensitiv$ or specific$ or comparative or comparison).tw. (83298)
44. (compare or compared).tw. (64991)
45. (test$ adj3 performance).tw. (1510)
46. accuracy.tw. (6222)
47. diagnostic errors/ or false negative reactions/ or false positive reactions/ (3424)
48. (ppv or npv or positive predictive value or negative predictive value).tw. (1067)
49. likelihood ratio$tw. (345)
50. (detection adj3 rate$).tw. (281)
51. reference standard$.tw. (191)
52. likelihood function$.tw. (2)
53. Comparative Studies/ (39360)
54. or/41-53 (159247)
55. 40 and 54 (65)

**Cochrane Central Register of Controlled Trials**

1. exp papillomavirus infections/ (315)
2. hpv.mp. (232)
3. papilloma$.mp. (315)
4. or/1-3 (562)
5. atypical squamous cells.tw. (31)
6. ASC-H.tw. (1)
7. (LSIL or HSIL).mp. (34)
8. (squamous intraepithelial or squamous intra-epithelial).mp. (67)
9. (cin-1 or cin-2 or cin-3).mp. (41)
10. (cin1 or cin2 or cin3).mp. (28)
11. cervical intraepithelial neoplasia/ (158)
12. dyskaryosis.mp. (17)
13. (low grade or high grade).tw. (1196)
14. or/5-13 (1344)
15. 4 and 14 (97)
16. di.fs. (19667)
17. exp Reagent Kits, Diagnostic/ (160)
18. linear array.mp. (20)
19. amplicor.af. (84)
20. cytyc.af. (7)
21. roche.af. (571)
22. imaging system.mp. (70)
23. hybrid capture.af. (36)
24. biotools.af. (0)
25. real time assay.mp. (0)
26. screening assay.mp. (4)
27. surepath.af. (0)
28. thinprep.af. (20)
29. GenID.af. (0)
30. exp Polymerase Chain Reaction/ (953)
31. or/16-30 (21190)
32. 15 and 31 (49)
33. limit 32 to yr=2000-2007 (42)
34. limit 33 to english [Limit not valid; records were retained] (42)
35. (news or letter).pt. (4506)
36. 34 not 35 (42)
37. human/ (0)
38. animal/ (0)
39. 38 not (37 and 38) (0)
40. 36 not 39 (42)

Current Contents and PubMed (last 60 days)
1. papillomavirus OR papilloma virus OR hpv
2. cervical intraepithelial neoplasia OR cin-1 OR cin-2
3. ASCUS OR ASC-H OR ASC-US OR LSIL OR HSIL
4. squamous intraepithelial OR squamous intra-epithelial OR dyskaryosis
5. atypical squamous cells OR low grade OR high grade
6. #1 AND (#2 OR #3 OR #4 OR #5)
7. linear array OR amplicor OR roche OR cytyc OR imaging system OR hybrid capture OR biotools OR surepath OR thinprep
8. real time assay OR screening assay OR GenID OR polymerase chain reaction OR pcr
9. diagnosis
10. diagnostic OR triage
11. #6 AND (#7 OR #8 OR #9 OR #10)
12. sensitivity OR accuracy OR specificity
13. test SAME performance
14. comparative
15. comparison
16. compare OR compared
17. detection SAME rate
18. ppv OR npv OR false positive* OR false negative*
19. likelihood ratio OR positive predictive value OR negative predictive value OR reference standard OR likelihood function
20. #11 AND (#12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19)
APPENDIX 3: EXCLUDED RETRIEVED PAPERS


APPENDIX 4: APPRAISED RETRIEVED PAPERS


