# Comparison of Predictors for High-Grade Cervical Intraepithelial Neoplasia in Women with Abnormal Smears

**Anne Szarewski,[1] Laurence Ambroisine,[1] Louise Cadman,[1] Janet Austin,[1] Linda Ho,[1] George Terry,[1] Stuart Liddle,[2] Roberto Dina,[3] Julie McCarthy,[5] Hilary Buckley,[6] Christine Bergeron,[7] Pat Soutter,[4] Deirdre Lyons,[8] and Jack Cuzick[1]**

[1]Cancer Research UK Centre for Epidemiology, Mathematics and Statistics, Wolfson Institute of Preventive Medicine; [2]The Doctors Laboratory; Departments of [3]Cytopathology and [4]Gynaecology, Hammersmith and Queen Charlotte's & Chelsea Hospitals; [5]Cytopathology Department, St. Mary's Hospital, London, United Kingdom; [6]Department of Pathology, St. Mary's Hospital, Manchester, United Kingdom; [7]Laboratoire Pasteur-Cerba, Cergy Pontoise, France; and [8]Gynaecology Department, St. Mary's Hospital, London



## Abstract

Background: **The detection of high-risk human papillomavirus (HPV) DNA provides higher sensitivity but lower specificity than cytology for the identification of high-grade cervical intraepithelial neoplasia (CIN). This study compared the sensitivity and specificity of several adjunctive tests for the detection of high-grade CIN in a population referred to colposcopy because of abnormal cytology.**

Methods: **953 women participated in the study. Up to seven tests were carried out on a liquid PreservCyt sample: Hybrid Capture II (Digene), Amplicor (Roche), PreTect HPV-Proofer (NorChip), APTIMA HPV assay (Gen-Probe), Linear Array (Roche), Clinical-Arrays (Genomica), and CINtec p16$^{INK4a}$ Cytology (mtm Laboratories) immunocytochemistry. Sensitivity, specificity, and positive predictive value (PPV) were based on the worst histology seen on either the biopsy or the treatment specimen after central review.**

Results: **273 (28.6%) women had high-grade disease (CIN2+) on worst histology, with 193 (20.2%) having CIN3+. For the detection of CIN2+, Hybrid Capture II had a sensitivity of 99.6%, specificity of 28.4%, and PPV of 36.1%. Amplicor had a sensitivity of 98.9%, specificity of 21.7%, and PPV of 33.5%. PreTect HPV-Proofer had a sensitivity of 73.6%, specificity of 73.1%, and PPV of 52.0%. APTIMA had a sensitivity of 95.2%, specificity of 42.2%, and PPV of 39.9%. CINtec p16$^{INK4a}$ Cytology had a sensitivity of 83.0%, specificity of 68.7%, and PPV of 52.3%. Linear Array had a sensitivity of 98.2%, specificity of 32.8%, and PPV of 37.7%. Clinical-Arrays had a sensitivity of 80.9%, specificity of 37.1%, and PPV of 33.0%.** (Cancer Epidemiol Biomarkers Prev 2008;17(11):3033–42)

## Introduction

High-risk human papillomavirus (HR-HPV) is a necessary factor for the development of cervical cancer (1), but the presence of HR-HPV DNA does not invariably lead to disease. Recently, we showed that the detection of HR-HPV provides high sensitivity but has lower specificity than cytology for the identification of high-grade cervical lesions in a screening population in the United Kingdom (2), and this finding has been replicated in several other studies (3-9). In addition, prospective studies have shown that HPV DNA-positive women are significantly more likely to develop high-grade squamous intraepithelial lesions within 10 years than women with a negative HPV DNA test (10, 11). If testing for HR-HPV DNA is to be used as a primary cervical screening test, refinements or additional tests are highly desirable to improve specificity while retaining its very high sensitivity.

The introduction of a liquid-based medium for collection of cytologic specimens has allowed other molecular techniques to be evaluated more easily as adjunctive or triage tests (12). The aim of this study was to compare directly the sensitivity and specificity of several tests from the same sample for the detection of high-grade cervical intraepithelial neoplasia (CIN) in a population referred to colposcopy because of abnormal cytology. All the tests were compared against the gold standard of histopathology.

## Materials and Methods

The study population comprised 953 women who had been referred to the colposcopy clinics at the Hammersmith and St. Mary's Hospitals in London between August 2005 and January 2007. Although not a screening population, the advantage was a broad range of outcomes and a high disease rate, which would enable accurate evaluation of sensitivity and specificity in a relatively small sample. Women were eligible if they had been referred as a result of one or more abnormal

Received 6/3/08; revised 8/22/08; accepted 9/2/08.

**Grant support:** Cancer Research UK Programme Grant. Financial contributions and/or assay kits from each of the manufacturers whose tests were involved (Digene/Qiagen, Gen-Probe, Genomica, mtm Laboratories, NorChip, and Roche Diagnostics).

**Note:** Supplementary data for this article are available at Cancer Epidemiology Biomarkers and Prevention Online (http://cebp.aacrjournals.org/).

**Requests for reprints:** Anne Szarewski, Cancer Research UK Centre for Epidemiology, Mathematics and Statistics, Wolfson Institute of Preventive Medicine, Charterhouse Square, London EC1M 6BQ, United Kingdom. Phone: 44-20-7014-0243; Fax: 44-20-7014-0269. E-mail: anne.szarewski@cancer.org.uk



doi:10.1158/1055-9965.EPI-08-0508

cervical smears, were not pregnant, had not been treated previously for CIN, nor had a hysterectomy. All women received a patient information sheet explaining the study and provided written consent. Approvals were obtained from the relevant local research ethics committees.

Before colposcopy, a cervical sample was obtained using a Cervex broom and placed in PreservCyt transport medium. Colposcopy was then done in the usual manner. The liquid-based cytology samples were transported to The Doctors' Laboratory, where an aliquot was first removed for cytology, processed using the ThinPrep system, and returned to the relevant cytopathology departments for reporting. The processing of the other tests was then done as follows:

NorChip PreTect HPV-Proofer tests were done by The Doctors' Laboratory using a second aliquot removed from the liquid-based cytology sample. The remainder of the sample was then sent to the laboratory at Cancer Research UK, where aliquots were removed for the other tests in the order: Digene Hybrid Capture II, Roche Amplicor, Gen-Probe APTIMA, Roche Linear Array, Genomica Clinical-Arrays, and mtm Laboratories CINtec p16$^{INK4a}$ Cytology. To look for the potential of sample carryover, we did autocorrelations on results obtained from sequential samples assayed by NorChip PreTect HPV-Proofer, Hybrid Capture II, Amplicor, Linear Array, and a PCR test for HPV-16 and HPV-18, respectively, and found no evidence of a pattern. All tests were carried out in the Cancer Research UK laboratory according to standard manufacturers' instructions, except the APTIMA test, which was carried out by the manufacturer. The molecular testing laboratories were blinded to the cytology and histopathology results.

Reporting of CINtec p16$^{INK4a}$ Cytology was carried out by Dr. Christine Bergeron according to a scoring method described recently (13, 14).

Histopathology was first reported locally and then centrally reviewed by either Dr. Hilary Buckley (92% of samples) or Dr. Christine Bergeron (8% of samples), who were blinded to all study test results but did have access to the concurrent cytology. All results are presented based on the reviewed histopathology and the highest grade of abnormality seen in the biopsy or treatment specimen was used.

**Laboratory Methods.** In this study, the following assays were carried out and scored in strict accordance with the manufacturer's protocol.

*(a) DNA-Based Detection Assays*

- Hybrid Capture II (Digene/Qiagen) detecting 13 HR-HPV genotypes collectively. The Hybrid Capture II assay requires 2 mL PreservCyt sample and is based on the hybridization of HPV DNA to a 13

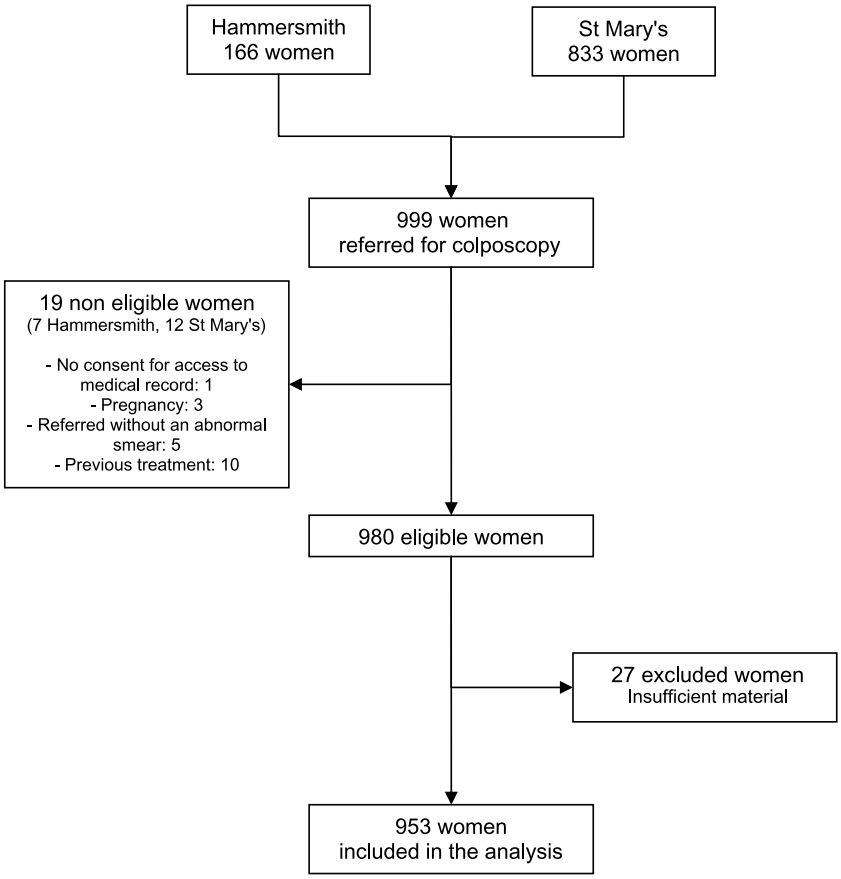


**Figure 1.** Study flow chart.

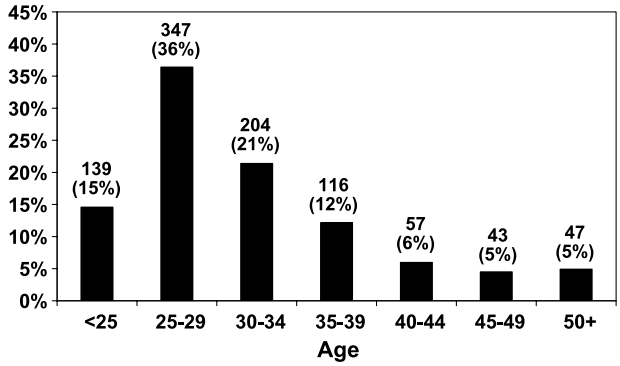


**Figure 2.** Age distribution.

HR-HPV RNA probe cocktail. The DNA:RNA hybrid is captured by an anti-DNA:RNA antibody and detected by chemiluminescence.

- Amplicor (Roche Diagnostics) detecting the same 13 HR-HPV genotypes collectively. The assay uses 1 mL PreservCyt and is based on PCR amplification of 13 HR-HPV types using pooled-type specific primers (sequence not given). β-Globin amplification acts as sample control. Amplicons are hybridized to oligonucleotide probes before colorimetric detection. DNA samples were amplified by 40 cycles of PCR.

*(b) DNA-Based Genotyping Assays*

- Linear Array (Roche Diagnostics) detecting 37 high-risk and low-risk genotypes individually. HPV DNA was amplified from 1 mL PreservCyt using biotinylated PGMY primers and β-globin acted as sample control. Amplicons are hybridized to an array of genotype-specific and cross-reactive oligonucleotide probes before colorimetric detection. Linear Array aims to identify the DNA of 35 HPV genotypes with more appropriate primers (PYGM). Unfortunately, the manufacturer does not provide clear guidelines as to how the results should be scored. In this study, we only scored clear positive bands at least as strong as the low level β-globin band. DNA samples were amplified by 40 cycles of PCR.
- Clinical-Arrays (Genomica) detecting 35 high-risk and low-risk genotypes individually. HPV DNA was amplified from 1 mL PreservCyt using biotinylated MY09/11 consensus primers hybridized to a low-density microarray of 37 HPV-type specific oligonucelotides for 45 cycles of amplification before colorimetric detection. A DNA sample control (CFTR) gene and an amplification reaction control are included.

*(c) $p16^{INK4a}$ Immunocytochemical Detection of Overexpression of the $p16^{INK4a}$ Tumor Suppressor Gene (CINtec Cytology, mtm Laboratories) and Scoring Using Nuclear Morphology.* Immunostaining was done using monoclonal antibody to $p16^{INK4a}$ with anti-mouse as second antibody and detected by DAB chromogen.

$p16^{INK4a}$ score was based on nuclear assessment of brown stained cells by four criteria (13).

(*a*) Increased size,
(*b*) Granular or hyperchromatic chromatin,
(*c*) Irregular shape, and
(*d*) Variable morphology from cell to cell.

Cells positive for any one of these criteria were scored as 2. Cells positive for A and one other criterion were scored as 3, and cells positive for A and more than one other criterion were scored as 4. The sample score was the highest score observed.

In addition, two RNA-based detection assays were also carried out. These were

- PreTect HPV-Proofer (NorChip) assayed by The Doctors' Laboratory. PreTect HPV-Proofer is a real-time multiplex NASBA assay for isothermal amplification of E6/E7 mRNA expressed by 5 HR-HPV types (16, 18, 31, 33, and 45) using proprietary primer sets.
- APTIMA (Gen-Probe) assayed by the manufacturer. The APTIMA assay is based on target capture, transcription-mediated amplification, and hybridization protection for the detection of E7 mRNA expression of 14 HR-HPV types.

Sample adequacy was assessed by a cellular DNA or RNA marker for all molecular assays, except Hybrid Capture II. Assay variability was controlled by including



**Table 1. Results of referral smears and smears taken on the day of colposcopy (concurrent smears)**

| Referral smear | Concurrent smear | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Normal/ unsatisfactory | Borderline | Mild dyskaryosis | Moderate dyskaryosis | Severe dyskaryosis | Possible glandular | Possible invasive | Total (%) |
| Normal smear/ unsatisfactory/ clinical suspicion | 12 | 1 | 2 | 2 | 0 | 0 | 0 | 17 (1.8) |
| Borderline | 43 | 31 | 16 | 8 | 4 | 2 | 0 | 104 (10.9) |
| Mild dyskaryosis | 171 | 115 | 235 | 73 | 23 | 0 | 0 | 617 (64.7) |
| Moderate dyskaryosis | 20 | 9 | 21 | 36 | 28 | 1 | 0 | 115 (12.1) |
| Severe dyskaryosis | 6 | 2 | 6 | 11 | 65 | 0 | 2 | 92 (9.6) |
| Possible glandular neoplasia | 1 | 1 | 0 | 0 | 0 | 4 | 0 | 6 (0.6) |
| Possible invasive cancer | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 2 (0.2) |
| Total (%) | 253 (26.5) | 159 (16.7) | 280 (29.4) | 130 (13.6) | 121 (12.7) | 7 (0.7) | 3 (0.3) | 953 (100.0) |

**Table 2.**

(A) Sensitivity, specificity, and PPV of different tests for the detection of high-grade disease (sensitivity only for CIN2 alone)

| Test (no. assessed) | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) |
|---|---|---|---|
| Hybrid Capture II ($n$ = 936) | | | |
| CIN3+ | 99.5 (97.1-100.0) | 25.4 (22.3-28.7) | 25.6 (22.5-28.9) |
| CIN2+ | 99.6 (98.0-100.0) | 28.4 (25.0-32.0) | 36.1 (32.6-39.6) |
| CIN2 | 100.0 (95.4-100.0) | — | — |
| Amplicor ($n$ = 949) | | | |
| CIN3+ | 99.5 (97.1-100.0) | 19.7 (16.9-22.7) | 23.9 (21.0-27.0) |
| CIN2+ | 98.9 (96.8-99.8) | 21.7 (18.6-25.0) | 33.5 (30.3-36.9) |
| CIN2 | 97.5 (91.2-99.7) | — | — |
| PreTect HPV-Proofer ($n$ = 918) | | | |
| CIN3+ | 82.2 (75.9-87.4) | 70.4 (66.9-73.7) | 41.2 (36.1-46.4) |
| CIN2+ | 73.6 (67.8-78.8) | 73.1 (69.5-76.4) | 52.0 (46.8-57.2) |
| CIN2 | 52.6 (40.8-64.2) | — | — |
| APTIMA ($n$ = 949) | | | |
| CIN3+ | 97.4 (94.1-99.2) | 38.8 (35.3-42.3) | 28.9 (25.4-32.5) |
| CIN2+ | 95.2 (92.0-97.4) | 42.2 (38.4-46.0) | 39.9 (36.2-43.8) |
| CIN2 | 90.0 (81.2-95.6) | — | — |
| CINtec p16$^{INK4a}$ Cytology ($n$ = 524) | | | |
| CIN3+ | 92.7 (86.0-96.8) | 65.8 (61.0-70.3) | 41.6 (35.3-48.0) |
| CIN2+ | 83.0 (76.1-88.6) | 68.7 (63.7-73.4) | 52.3 (45.8-58.7) |
| CIN2 | 59.1 (43.2-73.7) | — | — |
| Linear Array ($n$ = 932) | | | |
| CIN3+ | 99.0 (96.3-99.9) | 29.6 (26.4-33.1) | 26.9 (23.6-30.3) |
| CIN2+ | 98.2 (95.8-99.4) | 32.8 (29.2-36.5) | 37.7 (34.1-41.4) |
| CIN2 | 96.3 (89.4-99.2) | — | — |
| Clinical-Arrays ($n$ = 737) | | | |
| CIN3+ | 83.9 (76.9-89.5) | 36.0 (32.2-40.0) | 24.0 (20.3-28.0) |
| CIN2+ | 80.9 (74.8-86.0) | 37.1 (33.0-41.4) | 33.0 (28.9-37.3) |
| CIN2 | 73.8 (60.9-84.2) | — | — |
| Referral smear: mild or worse ($n$ = 953) | | | |
| CIN3+ | 93.3 (88.8-96.4) | 14.2 (11.8-16.9) | 21.6 (18.9-24.6) |
| CIN2+ | 93.4 (89.8-96.0) | 15.1 (12.5-18.1) | 30.6 (27.5-33.9) |
| CIN2 | 93.7 (86.0-97.9) | — | — |
| Concurrent smear: mild or worse ($n$ = 953) | | | |
| CIN3+ | 95.9 (92.0-98.2) | 53.2 (49.5-56.8) | 34.2 (30.2-38.4) |
| CIN2+ | 93.8 (90.2-96.3) | 58.1 (54.3-61.8) | 47.3 (43.0-51.6) |
| CIN2 | 88.7 (79.7-94.7) | — | — |

(B) Sensitivity, specificity, and PPV of different tests for the detection of high-grade disease in women ages <30 y (sensitivity only for CIN2 alone)

| Test (no. assessed) | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) |
|---|---|---|---|
| Hybrid Capture II ($n$ = 481) | | | |
| CIN3+ | 100.0 (96.3-100.0) | 20.7 (16.7-25.1) | 24.6 (20.5-29.1) |
| CIN2+ | 100.0 (97.5-100.0) | 23.4 (19.0-28.3) | 35.6 (30.9-40.5) |
| CIN2 | 100.0 (92.0-100.0) | — | — |
| Amplicor ($n$ = 485) | | | |
| CIN3+ | 99.0 (94.4-100.0) | 15.0 (11.6-18.9) | 22.8 (18.9-27.1) |
| CIN2+ | 98.6 (95.0-99.8) | 16.6 (12.8-21.0) | 32.9 (28.4-37.5) |
| CIN2 | 97.7 (88.0-99.9) | — | — |
| PreTect HPV-Proofer ($n$ = 468) | | | |
| CIN3+ | 85.3 (76.5-91.7) | 65.7 (60.6-70.5) | 38.8 (32.1-45.7) |
| CIN2+ | 74.3 (66.1-81.4) | 67.5 (62.1-72.5) | 48.3 (41.4-55.3) |
| CIN2 | 48.8 (32.9-64.9) | — | — |
| APTIMA ($n$ = 484) | | | |
| CIN3+ | 96.0 (30.0-98.9) | 35.1 (30.3-40.1) | 27.5 (22.9-32.6) |
| CIN2+ | 92.3 (86.7-96.1) | 37.5 (32.4-42.9) | 38.3 (33.1-43.6) |
| CIN2 | 84.1 (69.9-93.4) | — | — |
| CINtec p16$^{INK4a}$ Cytology ($n$ = 275) | | | |
| CIN3+ | 89.8 (79.2-96.2) | 63.9 (57.1-70.3) | 15.6 (6.6-37.1) |
| CIN2+ | 78.0 (67.5-86.4) | 65.3 (58.1-72.0) | 48.9 (40.0-57.7) |
| CIN2 | 47.8 (26.8-69.4) | — | — |
| Linear Array ($n$ = 477) | | | |
| CIN3+ | 98.0 (92.9-99.8) | 23.8 (19.6-28.4) | 25.2 (20.9-29.8) |
| CIN2+ | 97.9 (94.0-99.6) | 26.6 (22.0-31.7) | 36.4 (31.5-41.4) |
| CIN2 | 97.7 (88.0-99.9) | — | — |
| Clinical-Arrays ($n$ = 368) | | | |
| CIN3+ | 86.8 (76.4-93.8) | 33.3 (28.0-39.0) | 22.8 (17.8-28.4) |

(*Continued on the following page*)

**Table 2. (Cont'd)**

(B) Sensitivity, specificity, and PPV of different tests for the detection of high-grade disease in women ages <30 y (sensitivity only for CIN2 alone)

| Test (no. assessed) | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) |
|---|---|---|---|
| CIN2+ | 82.2 (73.3-89.1) | 34.1 (28.4-40.1) | 32.0 (26.4-38.1) |
| CIN2 | 72.7 (54.5-86.7) | — | — |
| Referral smear: mild or worse ($n$ = 486) | | | |
| CIN3+ | 93.9 (87.3-97.7) | 11.9 (8.8-15.5) | 21.4 (17.7-25.6) |
| CIN2+ | 93.7 (88.4-97.1) | 12.5 (9.2-16.5) | 30.9 (26.6-35.5) |
| CIN2 | 93.2 (81.3-98.6) | — | — |
| Concurrent smear: mild or worse ($n$ = 486) | | | |
| CIN3+ | 98.0 (92.9-99.8) | 50.4 (45.3-55.5) | 33.6 (28.1-39.3) |
| CIN2+ | 95.1 (90.2-98.0) | 55.4 (50.0-60.7) | 47.1 (41.2-53.0) |
| CIN2 | 88.6 (74.4-96.2) | — | — |

(C) Sensitivity, specificity, and PPV of different tests for the detection of high-grade disease in women ages ≥30 y (sensitivity only for CIN2 alone)

| Test (no. assessed) | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) |
|---|---|---|---|
| Hybrid Capture II ($n$ = 455) | | | |
| CIN3+ | 98.9 (94.2-100.0) | 30.4 (25.7-35.4) | 26.7 (22.1-31.8) |
| CIN2+ | 99.2 (95.7-100.0) | 33.5 (28.4-38.9) | 36.6 (31.5-42.0) |
| CIN2 | 100.0 (89.7-100.0) | — | — |
| Amplicor ($n$ = 464) | | | |
| CIN3+ | 100.0 (96.2-100.0) | 24.6 (20.3-29.3) | 25.2 (20.9-29.9) |
| CIN2+ | 99.2 (95.8-100.0) | 26.9 (22.2-32.0) | 34.3 (29.5-39.4) |
| CIN2 | 97.1 (85.1-99.9) | — | — |
| PreTect HPV-Proofer ($n$ = 450) | | | |
| CIN3+ | 78.9 (69.0-86.8) | 75.3 (70.5-79.6) | 44.4 (36.5-52.4) |
| CIN2+ | 72.8 (64.1-80.4) | 78.8 (73.9-83.1) | 56.9 (48.8-64.7) |
| CIN2 | 57.1 (39.4-73.7) | — | — |
| APTIMA ($n$ = 465) | | | |
| CIN3+ | 98.9 (94.2-100.0) | 42.6 (37.5-47.8) | 30.4 (25.3-35.9) |
| CIN2+ | 98.5 (94.6-99.8) | 46.9 (41.4-52.4) | 41.8 (36.2-47.6) |
| CIN2 | 97.2 (85.5-99.9) | — | — |
| CINtec p16$^{INK4a}$ Cytology ($n$ = 249) | | | |
| CIN3+ | 96.0 (86.3-99.5) | 67.8 (60.9-74.3) | 42.9 (33.5-52.6) |
| CIN2+ | 88.7 (79.0-95.0) | 72.5 (65.3-78.9) | 56.3 (46.6-65.6) |
| CIN2 | 71.4 (47.8-88.7) | — | — |
| Linear Array ($n$ = 455) | | | |
| CIN3+ | 100.0 (96.2-100.0) | 35.7 (30.8-40.9) | 28.8 (24.0-34.1) |
| CIN2+ | 98.5 (94.6-99.8) | 39.1 (33.7-44.6) | 39.3 (33.9-44.8) |
| CIN2 | 94.4 (81.3-99.3) | — | — |
| Clinical-Arrays ($n$ = 369) | | | |
| CIN3+ | 81.3 (70.7-89.4) | 38.8 (33.2-44.6) | 25.3 (19.9-31.3) |
| CIN2+ | 79.6 (70.5-86.9) | 40.2 (34.3-46.4) | 34.0 (28.1-40.4) |
| CIN2 | 75.0 (55.1-89.3) | — | — |
| Referral smear: mild or worse ($n$ = 467) | | | |
| CIN3+ | 92.6 (85.3-97.0) | 16.6 (13.0-20.8) | 21.9 (17.9-26.2) |
| CIN2+ | 93.1 (87.3-96.8) | 17.8 (13.9-22.3) | 30.4 (25.9-35.2) |
| CIN2 | 94.4 (81.3-99.3) | — | — |
| Concurrent smear: mild or worse ($n$ = 467) | | | |
| CIN3+ | 93.6 (86.6-97.6) | 56.0 (50.8-61.1) | 34.9 (29.0-41.2) |
| CIN2+ | 92.3 (86.3-96.2) | 60.8 (55.4-66.1) | 47.6 (41.3-54.0) |
| CIN2 | 88.9 (73.9-96.9) | — | — |

a pooled positive control and a pooled negative control in each run for each test. Reagents from the same batch were used throughout and the controls provided in the kits showed little assay variability over time.

Data entry and statistical analysis were carried out at the Cancer Research UK Centre for Epidemiology, Mathematics and Statistics.

**Statistical Methods.** All statistical analyses were carried out using Stata 9.2 (Stata Corp.)

We calculated that a minimum of 500 patients would be needed to have 80% power to detect a change in sensitivity from 80% to 90% assuming a 25% prevalence of CIN2+ and a discordance rate of 3.6% (18 of 500). This number of subjects would give at least 90% power to detect improvements in specificity from 50% (common in triage) to 75%.

The gold standard was histologically confirmed high-grade CIN (CIN2+), but calculations were also made for histologically confirmed CIN3+. The sensitivity, specificity, and positive predictive value (PPV) of each test

**Table 3. Sensitivity, specificity, and PPV of different tests for the detection of high-grade disease in women with referral smear less or equal to single mild (sensitivity only for CIN2 alone)**

| Test (no. assessed) | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) |
|---|---|---|---|
| Hybrid Capture II ($n$ = 567) | | | |
| CIN3+ | 100.0 (92.6-100.0) | 26.0 (22.3-30.0) | 11.1 (8.3-14.5) |
| CIN2+ | 100.0 (95.9-100.0) | 28.2 (24.2-32.4) | 20.4 (16.7-24.5) |
| CIN2 | 100.0 (91.2-100.0) | — | — |
| Amplicor ($n$ = 573) | | | |
| CIN3+ | 100.0 (92.6-100.0) | 21.0 (17.5-24.7) | 10.4 (7.7-13.5) |
| CIN2+ | 97.7 (91.9-99.7) | 22.2 (18.6-26.2) | 18.4 (14.9-22.2) |
| CIN2 | 94.9 (82.7-99.4) | — | — |
| PreTect HPV-Proofer ($n$ = 558) | | | |
| CIN3+ | 89.4 (76.9-96.5) | 72.8 (68.7-76.6) | 23.2 (17.3-30.0) |
| CIN2+ | 70.6 (59.7-80.0) | 74.4 (70.2-78.3) | 33.1 (26.3-40.5) |
| CIN2 | 47.4 (31.0-61.2) | — | — |
| APTIMA ($n$ = 573) | | | |
| CIN3+ | 97.9 (88.9-99.9) | 39.4 (35.2-43.8) | 12.9 (9.6-16.8) |
| CIN2+ | 96.6 (90.4-99.3) | 42.3 (37.8-46.8) | 23.3 (19.0-28.0) |
| CIN2 | 95.0 (83.1-99.4) | — | — |
| CINtec p16$^{INK4a}$ Cytology ($n$ = 323) | | | |
| CIN3+ | 88.0 (68.8-97.5) | 67.1 (61.5-72.4) | 18.3 (11.9-26.4) |
| CIN2+ | 66.7 (52.5-78.9) | 68.8 (62.9-74.3) | 30.0 (22.0-39.0) |
| CIN2 | 48.3 (29.4-67.5) | — | — |
| Linear Array ($n$ = 562) | | | |
| CIN3+ | 100.0 (92.6-100.0) | 32.1 (28.1-36.3) | 12.1 (9.0-15.7) |
| CIN2+ | 96.6 (90.4-99.3) | 34.2 (29.9-38.6) | 21.4 (17.5-25.8) |
| CIN2 | 92.5 (79.6-98.4) | — | — |
| Clinical-Arrays ($n$ = 456) | | | |
| CIN3+ | 89.2 (74.6-97.0) | 37.2 (32.6-42.1) | 11.1 (7.8-15.3) |
| CIN2+ | 80.9 (69.5-89.4) | 37.9 (33.0-42.9) | 18.6 (14.3-23.5) |
| CIN2 | 71.0 (52.0-85.8) | — | — |

were computed and 95% exact binomial confidence intervals (95% CI) were obtained using exact binomial methods when necessary. Sensitivity and specificity were further studied by ROC analyses for concurrent cytology, Hybrid Capture II, Amplicor, and CINtec p16$^{INK4a}$ Cytology score using different cutoff levels. Additional calculations of sensitivity, specificity, and PPV were carried out for different age groups. Agreement between the pairs of tests was assessed using Cohen's $\kappa$ statistic. Results for concurrent cytology are also given, but they are not comparable with the study tests because they were available to the pathologists when reading the histology specimens.

In the present analyses, only HR-HPV types were included for Linear Array and Clinical-Arrays (Genomica), and samples showing multiple unspecified types for Clinical-Arrays were removed from the statistical analyses.

## Results and Discussion

Figure 1 shows the study recruitment. Nine hundred ninety-nine women consented to participate in the study; however, 19 proved to be ineligible. In 27 cases, there was insufficient material for analysis. Thus, the tests from 953 women were analyzed. The median length of time between the referral smear and colposcopy was 2.4 months (interquartile range, 1.9-3.3; range, 0.6-50 months). The median age of the women was 29.9 years (interquartile range, 26.4-35.6). The age distribution is shown in Fig. 2.

Table 1 shows the results of the referral smears, where it can be seen that ~23% of women had high-grade dyskaryosis and >75% had low-grade disease (borderline or mild dyskaryosis). Table 1 also shows the results of the cervical smear taken on the day of the colposcopy (concurrent smear). Although 27% of women had high-grade dyskaryosis, the percentage with low-grade dyskaryosis was only 46%, whereas in 26% of cases, this smear was normal. This may reflect the time interval between the two smears, during which some low-grade disease may have regressed, and also technique differences and the subjectivity of cytologic assessment. It should be noted that not all the referral smears were taken using liquid-based cytology, because the area was

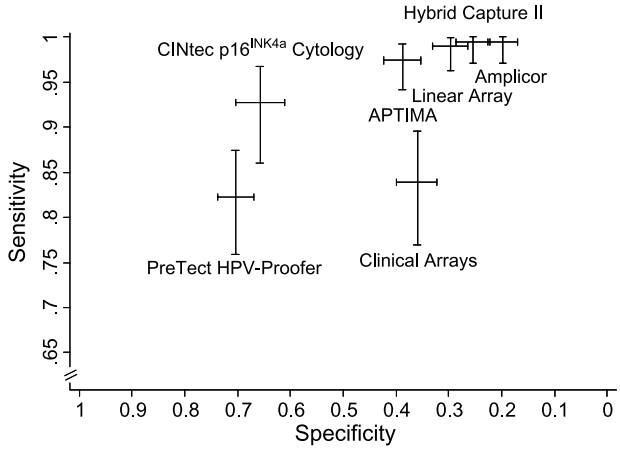


**Figure 3.** Summary graph of the sensitivity and specificity results for the detection of CIN3+ (with 95% CI).

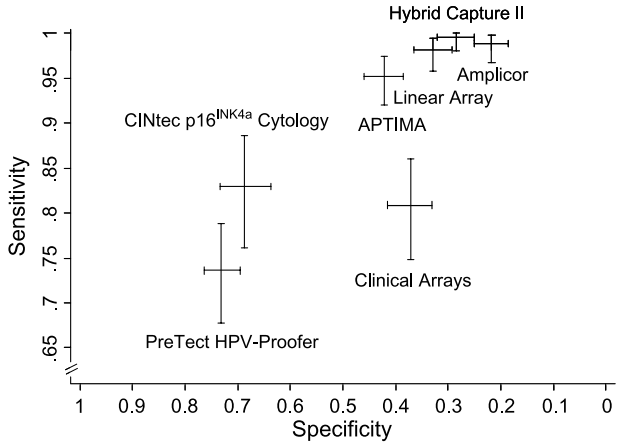


**Figure 4.** Summary graph of the sensitivity and specificity results for the detection of CIN2+ (with 95% CI).

in the process of a changeover during the study. All the colposcopy clinic (concurrent) smears were taken using liquid-based cytology. The results of the concurrent cytology were available to the pathologist when determining histologic outcome and so could not be independently evaluated. However, the results are shown in Table 2 for comparison.

Over 45% of the women had either a normal colposcopy with no biopsy or a negative biopsy. Around 25% of women were found to have low-grade CIN, 8% had CIN2, and just below 20% had CIN3 or worse.

Sensitivity, specificity, and PPV for CIN2+ and CIN3+ of the different tests are reported in Table 2A to C, respectively, for all women, women aged <30 and ≥30 years. We have chosen to report separately results for CIN2+ and CIN3+ because CIN2 is less likely to progress and has greater variability in diagnosis. CIN3 has been shown to have greater reproducibility. Tables 2A to C and 3 also include the sensitivity for CIN2 alone.

Unfortunately, CINtec p16$^{INK4a}$ Cytology testing was the last evaluation in the testing sequence, and there was not always sufficient sample to perform the test. The results pertaining to CINtec p16$^{INK4a}$ Cytology are therefore from a subset of 534 (56%) women, including 10 women for which the results were not available. We therefore present the CINtec p16$^{INK4a}$ Cytology results with the caveat that they should be interpreted with caution. In addition due to technical problems during the evaluation, the Genomica Clinical-Arrays results should also be interpreted with caution.

Sensitivity, specificity, and PPV for all the tests are shown in Table 2A. The sensitivity and specificity results for the tests are summarized in Figs. 3 and 4, respectively, for CIN2+ and CIN3+. Four adjunctive tests had a sensitivity greater than 95% for high-grade disease (CIN2+ and CIN3+, respectively): Amplicor (98.9% and 99.5%), Hybrid Capture II (99.6% and 99.5%), Linear Array (98.2% and 99.0%), and APTIMA (95.2% and 97.4%). Of these, APTIMA showed the highest specificity (42.2% and 38.8%; Table 2A). Linear Array had similar sensitivity to Hybrid Capture II and Amplicor, as has been shown previously, but in contrast to earlier studies (15) had a higher specificity

**Table 4. Sensitivity, specificity, and PPV of Hybrid Capture II, Amplicor, and CINtec p16$^{INK4a}$ Cytology at different cutoffs for the detection of high-grade disease**

| Test | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) |
|---|---|---|---|
| Endpoint CIN3+ | | | |
| Hybrid Capture II | | | |
| ≥1 RLU/PC | 99.5 (97.1-100.0) | 25.4 (22.3-28.7) | 25.6 (22.5-28.9) |
| ≥2 RLU/PC | 98.4 (95.5-99.7) | 28.8 (25.5-32.2) | 26.3 (23.1-29.7) |
| ≥4 RLU/PC | 96.9 (93.3-98.8) | 31.7 (28.4-35.2) | 26.8 (23.5-30.3) |
| ≥10 RLU/PC | 94.8 (90.6-97.5) | 36.4 (33.0-40.0) | 27.8 (24.4-31.4) |
| Amplicor | | | |
| ≥0.2 | 99.5 (97.1-100.0) | 19.7 (16.9-22.7) | 23.9 (21.0-27.0) |
| ≥1 | 99.0 (96.3-99.9) | 24.5 (21.5-27.8) | 25.0 (22.0-28.3) |
| ≥2 | 98.4 (95.5-99.7) | 28.2 (25.1-31.6) | 25.9 (22.8-29.2) |
| CINtec p16$^{INK4a}$ Cytology score | | | |
| ≥2 | 97.2 (92.2-99.4) | 35.9 (31.3-40.7) | 28.5 (24.0-33.4) |
| ≥3 | 92.7 (86.0-96.8) | 65.8 (61.0-70.3) | 41.6 (35.3-48.0) |
| ≥4 | 36.7 (27.7-46.5) | 94.2 (91.5-96.3) | 62.5 (49.5-74.3) |
| Endpoint CIN2+ | | | |
| Hybrid Capture II | | | |
| ≥1 RLU/PC | 99.6 (98.0-100.0) | 28.4 (25.0-32.0) | 36.1 (32.6-39.6) |
| ≥2 RLU/PC | 98.5 (96.3-99.6) | 32.0 (28.5-35.7) | 37.0 (33.5-40.6) |
| ≥4 RLU/PC | 96.7 (93.8-98.5) | 35.0 (31.4-38.7) | 37.6 (34.0-41.3) |
| ≥10 RLU/PC | 94.4 (91.0-96.9) | 39.9 (36.2-43.8) | 38.9 (35.2-42.8) |
| Amplicor | | | |
| ≥0.2 | 98.9 (96.8-99.8) | 21.7 (18.6-25.0) | 33.5 (30.3-36.9) |
| ≥1 | 98.2 (95.8-99.4) | 27.0 (23.7-30.5) | 35.1 (31.7-38.6) |
| ≥2 | 96.7 (93.8-98.5) | 30.7 (27.2-34.3) | 36.0 (32.5-39.6) |
| CINtec p16$^{INK4a}$ Cytology score | | | |
| ≥2 | 91.5 (85.9-95.4) | 37.5 (32.5-42.6) | 37.6 (32.7-42.8) |
| ≥3 | 83.0 (76.1-88.6) | 68.7 (63.7-73.4) | 52.3 (45.8-58.7) |
| ≥4 | 27.5 (20.6-35.2) | 94.1 (91.2-96.2) | 65.6 (52.7-77.1) |

at 32.8% for CIN2+ and 29.6% for CIN3+. Overall, the highest specificity (>70%) was achieved with the PreTect HPV-Proofer, but this test had relatively low sensitivity (Table 2A). The CINtec p16$^{INK4a}$ Cytology test (on a subset of samples) also gave lower sensitivity (83.0% and 92.7%), compared with HPV DNA tests, but had higher specificity (68.7% and 65.8%), approaching that of the PreTect HPV-Proofer (73.1% and 70.4%). The Genomica test (Clinical-Arrays) had performance characteristics that were dominated by those of the other tests. The Genomica results were less favorable than those obtained in other studies (16, 17) and may be due to the fact that 187 samples had ''multiple types'' results. The reasons for this are unclear despite a careful review of laboratory procedures on several occasions. When focusing on CIN3+, overall, sensitivity is slightly improved and specificity is slightly reduced; however, the ordering of the tests remains the same for both measures.

All tests showed a similar sensitivity in both younger (<30 years) and older women (≥30 years), although CINtec p16$^{INK4a}$ Cytology did better in the older age group and PreTect HPV-Proofer and Clinical-Arrays did slightly better in the younger age group for CIN3+. All tests, however, showed higher specificity in the older age group (Table 2B and C).

When the population was restricted to women who had a single mild or only borderline dyskaryotic smears, the results were generally similar (Table 3). For the four tests that showed the highest sensitivity (Hybrid Capture II, Amplicor, Linear Array, and APTIMA), specificity was ~50% higher when only women with borderline dyskaryotic smears were considered (Supplementary Table S1). However, the relative performance of the tests was unchanged in these lower risk categories.

The results for CINtec p16$^{INK4a}$ Cytology were less favorable than those in other recent publications (14), which found a sensitivity of 100% in a low-grade squamous intraepithelial lesions population, with a specificity of 81.7%. However, there is a wide variation in the literature, which may be at least partly related to changes in and a lack of standardization in reporting of p16$^{INK4a}$ cytology results (18). It should also be remembered that the p16$^{INK4a}$ samples in this study may have been suboptimal and further studies are required to fully evaluate this test.

Table 4 shows the effects on sensitivity, specificity, and PPV of using different cutoffs for Hybrid Capture II, Amplicor, and CINtec p16$^{INK4a}$ Cytology score in predicting histologically confirmed high-grade disease (CIN2+ and CIN3+). If the cutoffs for both Hybrid Capture II and Amplicor were raised (from 1 to ≥2 or ≥4 RLU/PC for Hybrid Capture II and from 0.2 to ≥1 or ≥2 RLU/PC for Amplicor), the sensitivity remained relatively unchanged, whereas the specificity slightly improved. Using a cutoff of ≥2 instead of ≥3 for the CINtec p16$^{INK4a}$ Cytology score improved the sensitivity (83.0-97.2%) but had an unacceptable effect on specificity, which decreased from 68.7% to 35.9%. High concordance between Hybrid Capture II and Amplicor has been shown previously (15, 19-21) However, our results suggest that both tests might benefit from adjusting their positivity cutoff values (to ≥2 RLU/PC for Hybrid Capture II and ≥1 for Amplicor), which would improve their specificity and PPV, while having

**Table 5. Summary of discordant results for all individuals (top half of table) and for CIN2+ individuals missed by at least one test (bottom half of the table)**

| | Hybrid Capture II | Amplicor | PreTect HPV-Proofer | APTIMA | CINtec p16$^{INK4a}$ Cytology | Linear Array | Clinical-Arrays |
|---|---|---|---|---|---|---|---|
| Hybrid Capture II | | (41,84)* $\kappa$ = 0.55 | (358,5) $\kappa$ = 0.28 | (109,8) $\kappa$ = 0.68 | (203,26) $\kappa$ = 0.16 | (72,33) $\kappa$ = 0.66 | (146,58) $\kappa$ = 0.28 |
| Amplicor | (1,3) | | (402,3) $\kappa$ = 0.22 | (171,22) $\kappa$ = 0.45 | (211,18) $\kappa$ = 0.16 | (92,2) $\kappa$ = 0.68 | (168,45) $\kappa$ = 0.23 |
| PreTect HPV-Proofer | (1,68) | (0,65) | | (17,272) $\kappa$ = 0.41 | (68,89) $\kappa$ = 0.37 | (1,316) $\kappa$ = 0.35 | (38,234) $\kappa$ = 0.28 |
| APTIMA | (1,13) | (3,13) | (59,3) | | (160,33) $\kappa$ = 0.28 | (41,104) $\kappa$ = 0.61 | (108,100) $\kappa$ = 0.34 |
| CINtec p16$^{INK4a}$ Cytology | (1,26) | (1,25) | (21,10) | (1,25) | | (23,178) $\kappa$ = 0.24 | (57,125) $\kappa$ = 0.11 |
| Linear Array | (1,5) | (0,2) | (64,0) | (13,5) | (24,2) | (0,35) | (121,70) $\kappa$ = 0.33 |
| Clinical-Arrays | (0,37) | (0,37) | (28,17) | (3,34) | (7,20) | | |

NOTE: Discordant results between two tests (A and B): A (positive for the test in the first column and negative for the test in the first row) and B (negative for the test in the first column and positive for the test in the first row) for the top half, the value of Cohen's $\kappa$ statistic, and the $P$ associated. For the top half, $P < 0.0001$, except for CINtec p16$^{INK4a}$ Cytology versus Clinical-Arrays ($P = 0.007$).
*For example, 41 individuals positive for Hybrid Capture II and negative for Amplicor and 84 individuals negative for Hybrid Capture II and positive for Amplicor.

minimal effect on sensitivity (Table 4). The ROC curves for Hybrid Capture II, Amplicor, and CINtec p16$^{INK4a}$ Cytology score are shown in Supplementary Figs. S1 to S3, respectively.

Table 5 indicates the concordance between the tests and the number of discordant cases. There was a fair to good agreement between Amplicor, Linear Array, APTIMA, and Hybrid Capture II ($\kappa$ values = 0.45-0.68). The other 2 $\times$ 2 comparisons showed a poor to moderate agreement between the tests ($\kappa$ values between 0.11 and 0.41).

It should be noted that the tests did not all miss the same cases, making comparisons more complex. There were 101 CIN2+ individuals missed by at least one test (details in Supplementary Table S2): 64 individuals were missed by one test only (35 by PreTect HPV-Proofer, 16 by Clinical-Arrays, and 10 by CINtec p16$^{INK4a}$ Cytology using the cutoff of 3, and 3 by APTIMA); 24 individuals were missed by two tests (1 by Hybrid Capture II and Clinical-Arrays, 5 by APTIMA and PreTect HPV-Proofer, 8 by Clinical-Arrays and PreTect HPV-Proofer, 8 by CINtec p16$^{INK4a}$ Cytology score and PreTect HPV-Proofer, and 2 by CINtec p16$^{INK4a}$ Cytology score and Clinical-Arrays); 9 individuals were missed by three tests (3 by CINtec p16$^{INK4a}$ Cytology score, Clinical-Arrays, and PreTect HPV-Proofer, 4 by APTIMA, Clinical-Arrays, and PreTect HPV-Proofer, 1 by Linear Array, Clinical-Arrays, and PreTect HPV-Proofer, and 1 by Linear Array, PreTect HPV-Proofer, and Amplicor); and 4 individuals were missed by four or more tests. Data on individual HPV types were complicated by the large number of multiple infections and will be reported separately.

This study is the first to compare a wide range of adjunctive tests on the same sample. The population studied had a wide range of cytologic outcomes that facilitated the ability to look at relative sensitivity. It is likely that sensitivity will be similar to that observed with only borderline or negative cytology, but further work is needed to fully evaluate specificity in this context. Four tests showed very high sensitivity, indicating that they are unlikely to miss any significant disease, and of these, the APTIMA test had the best specificity. However, when alternative cutoffs were used, the differences between APTIMA, Hybrid Capture II, and Amplicor were small and not significant. The Clinical-Arrays test had several cases with unevaluable multiple types. This had a negative effect on the results and led to a relatively poor performance in both sensitivity and specificity. Further work is needed to fully evaluate CINtec p16$^{INK4a}$ Cytology. High sensitivity is clearly important in these circumstances. However, it is known that many CIN2 and some CIN3 lesions will regress spontaneously and it is possible that tests with lower sensitivity still identify the lesions that are destined to progress to cancer.

Where triage is most needed (e.g., women with a single mildly dyskaryotic or borderline dyskaryosis on smears), absolute sensitivity and the relative performance of the tests were similar to that for the whole cohort, but specificity was higher, especially in the case of borderline dyskaryotic smears, indicating the value of a triage test for these women. These data provide useful relative comparisons of the performances of these tests in such a situation.

## Disclosure of Potential Conflicts of Interest

Financial contributions to the study were received as above. J. Cuzick is a consultant for Roche, on advisory boards for Gen-Probe, and a speaker for Qiagen.

## Acknowledgments

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank the staff in the Departments of Cytopathology and Colposcopy at the Hammersmith Hospital and St. Mary's Hospital for help with this study.

## References

1. Walboomers JM, Jacobs MV, Manos MM, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol 1999;189:12–9.
2. Cuzick J, Szarewski A, Cubie H, et al. Management of women who test positive for high risk types of human papillomavirus: the HART study. Lancet 2003;362:1871–6.
3. Cuzick J, Clavel C, Petry K-U, et al. Overview of the European and North American studies on HPV testing in primary cervical screening. Int J Cancer 2006;119:1095–101.
4. Almonte M, Ferreccio C, Winkler J, et al. Cervical screening by visual inspection, HPV testing, liquid-based and conventional cytology in Amazonian Peru. Int J Cancer 2007;121:796–802.
5. Bulkmans NW, Berkhof J, Rozendaal L, et al. Human papillomavirus DNA testing for the detection of cervical intraepithelial neoplasia grade 3 and cancer: 5-year follow-up of a randomised controlled implementation trial. Lancet 2007;370:1764–72.
6. Mayrand MH, Duarte-Franco E, Rodrigues I, et al.; Canadian Cervical Cancer Screening Trial Study Group. Human papillomavirus DNA versus Papanicolaou screening tests for cervical cancer. N Engl J Med 2007;357:1579–88.
7. Naucler P, Ryd W, Törnberg S, et al. Human papillomavirus and Papanicolaou tests to screen for cervical cancer. N Engl J Med 2007; 357:1589–97.
8. Kotaniemi-Talonen L, Anttila A, Malila N, et al. Screening with a primary human papillomavirus test does not increase detection of cervical cancer and intraepithelial neoplasia 3. Eur J Cancer 2008;44: 565–71.
9. Ronco G, Giorgi-Rossi P, Carozzi F, et al.; New Technologies for Cervical Cancer Screening Working Group. Results at recruitment from a randomized controlled trial comparing human papillomavirus testing alone with conventional cytology as the primary cervical cancer screening test. J Natl Cancer Inst 2008;100: 492–501.
10. Bulkmans NW, Rozendaal L, Voorhorst FJ, Snijders PJ, Meijer CJ. Long-term protective effect of high risk human papillomavirus testing in population-based cervical screening. Br J Cancer 2005;92: 1800–2.
11. Khan MJ, Castle PE, Lorincz AT, et al. The elevated 10-year risk of cervical precancer and cancer in women with human papillomavirus (HPV) type 16 or 18 and the possible utility of type-specific HPV testing in clinical practice. J Natl Cancer Inst 2005; 97:10729.
12. Yoshida T, Fukuda T, Sano T, Kanuma T, Owada N, Nakajima T. Usefulness of liquid-based cytology specimens for the immunocytochemical study of p16 expression and human papillomavirus testing: a comparative study using simultaneously sampled histology materials. Cancer 2004;102:100–8.
13. Wentzensen N, Bergeron C, Cas F, Eschenbach D, Vinokurova S, von Knebel Doeberitz M. Evaluation of a nuclear score for p16$^{INK4a}$-stained cervical squamous cells in liquid-based cytology samples. Cancer 2005;105:461–7.
14. Wentzensen N, Bergeron C, Cas F, Vinokurova S, von Knebel Doeberitz M. Triage of women with ASCUS and LSIL cytology: use of qualitative assessment of p16$^{INK4a}$ positive cells to identify patients with high-grade cervical intraepithelial neoplasia. Cancer 2007;111:58–66.
15. Stevens MP, Garland SM, Rudland E, Tan J, Quinn MA, Tabrizi SN. Comparison of the Digene Hybrid Capture 2 Assay and Roche Amplicor and Linear Array Human Papillomavirus (HPV) tests in detecting high-risk HPV genotypes in specimens from women

with previous abnormal Pap smear results. J Clin Microbiol 2007;45: 2130–7.
16. Lillo F. HPV genotyping: are different hybridization methods comparable in detecting high risk infections? Presented at Eurogin 2006, FC4-04.
17. Verdasca N, Pista A, Oliveira A, et al. Comparison of Papillocheck, Clinical Array and Linear Array HPV genotyping assays in a group of Portuguese women. Presented at ISPV Beijing 2007, PS6-20.
18. Kalof AN, Cooper K. $p16^{INK4a}$ immunoexpression: surrogate marker of high-risk HPV and high-grade cervical intraepithelial neoplasia. Adv Anat Pathol 2006;13:190–4.
19. Carozzi F, Bisanzi S, Sani C, et al. Agreement between the Amplicor Human Papillomavirus Test and the Hybrid Capture II Assay in detection of high-risk human papillomavirus and diagnosis of biopsy-confirmed high-grade cervical disease. J Clin Microbiol 2007;45:364–9.
20. Halfon P, Trepo E, Antoniotti G, et al. Prospective evaluation of the Hybrid Capture 2 and Amplicor human papillomavirus (HPV) tests for detection of 13 high-risk HPV genotypes in atypical squamous cells of uncertain significance. J Clin Microbiol 2007;45: 313–6.
21. De Francesco MA, Gargiulo F, Schreiber C, Ciravolo G, Salinaro F, Manca N. Comparison of the Amplicor Human Papillomavirus Test and the Hybrid Capture 2 Assay for detection of high-risk human papillomavirus in women with abnormal PAP smear. J Virol Methods 2008;147:10–7.