

ORIGINAL ARTICLE

Optional screening strategies for cervical cancer using standalone tests and their combinations among low- and medium-income populations in Latin America and Eastern Europe

Luis Otavio Sarian, Sophie Derchain, Irena Shabalova, Silvio Tatti, Paulo Naud, Adhemar Longatto-Filho, Stina Syrjänen and Kari Syrjänen, On behalf of members of the Latin American Screening Study (LAMS) and the New Independent States of the Former Soviet Union (NIS) groups

J Med Screen 2010;17:195–203
DOI: 10.1258/jms.2010.010016

See end of article for authors' affiliations

Correspondence to:
Professor Kari Syrjänen MD
PhD FIAC, Department of
Oncology and
Radiotherapy, Turku
University Hospital,
Savitehtaankatu 1, FIN-
20521 Turku, Finland;
kari.syrjanen@tyks.fi

Accepted for publication
13 August 2010

Background The performance of cervical cancer (CC) screening can be improved by combining Pap smear with human papillomavirus (HPV) testing or visual methods, addressing local demographic, clinical and economic characteristics.

Objectives To examine the performance of standalone and combined screening tools in populations with variable prevalence of cervical intraepithelial neoplasia (CIN) and CC.

Methods Merged data-sets from the Latin American Screening Study and New Independent States cohorts provided results for 15,000 women, screened using Pap smear, HPV testing and visual inspection with acetic acid, in Brazil, Argentina, Russia, Belarus and Latvia. Bayesian correction for verification bias was used.

Results At CIN2+ cut-off, HPV detection alone was the most sensitive technique. There was an improvement (88.5% to 92.7%) in Hybrid Capture 2 (HC2) sensitivity among women ≥ 35 years old. Using HPV detection alone was the least specific screening tool, regardless of the age group (69.9% [95% CI 66.5–72.8%] and 86.4% [95% CI 84.6–88.2%], in $<$ or ≥ 35 years, respectively). Of the test combinations, Pap smear (LSIL threshold) with HC2 had the highest specificity (98.7%; 95% CI 98.3–99.0%). However, in women ≥ 35 years, the sensitivity of Pap alone was superior to that of the combination.

Conclusions The Pap test is a highly specific screening option in populations with medium-range CC prevalence. Combined testing for HPV in this scenario may yield slightly better positive predictive values in women ≥ 35 years of age with LSIL, but at a high incremental cost.

INTRODUCTION

Despite the success of organized Pap test-based screening programmes in reducing the burden of cervical cancer (CC) in many privileged countries, inherent limitations of the test preclude complete control of CC even in countries with well-implemented screening programmes.^{1,2} Some countries with economic and logistical shortcomings have been unable to exploit the full potential of cervical cytology screening, and continue to deal with a substantial disease burden.³

There are limitations to the Pap test. The full implementation of such a screening programme requires a well-structured and costly network of health services, including sample collection sites, processing and reading laboratories, centralized cancer registry and professional staff to convey results to the screened population. Facilities to treat detected

CC and precursor lesions are also costly.⁴ The Pap test itself has limited sensitivity (40–50% even in best laboratories), necessitating repeat screening at relatively short intervals (2–3 years).⁴ Verification of positive results using colposcopy and biopsy necessitates additional visits.^{4,5}

The technical shortcomings of cervical cytology led to the failure of many countries to reduce their CC burden, but have also prompted the development of alternative and/or adjunct screening methods, such as human papillomavirus (HPV) testing, visual inspection with acetic acid (VIA) and combinations of these tests. Results from most large trials suggest that screening performance can be improved with such combinations. Combinations of screening tests are usually trialled either sequentially (referral to colposcopy when screen positive in both tests) or in parallel (colposcopy performed when one test is positive). Sequential

combinations generally yield high specificity at the expense of sensitivity; the reverse is true for parallel testing. However, these improvements seem to be critically dependent on local circumstances and availability of resources.

Based on an encouraging experience from the USA,⁶ the Food and Drug Administration (FDA) recently approved the use of HPV testing with Hybrid Capture 2 (HC2) in combination with the Pap test for screening women over 30 years old. This strategy aims to maximize the high sensitivity of the HPV test and the almost 100% specificity of the Pap test. By contrast, in low-resource settings like India, researchers advocate the use of VIA as the primary CC screening tool, with advantages over the Pap test such as the 'see and treat' possibility. However, when tested in other settings (e.g. Latin America), VIA was not particularly valuable, because of its low specificity and low positive predictive values in populations with moderate to low prevalence of CC and its precursors.^{7,8}

We evaluated the performance of different screening tests compared with conventional and liquid-based cervical cytology in two trials, the New Independent States (NIS) Cohort and the Latin American Screening Study (LAMS), from Russia, Belarus and Latvia, and Brazil and Argentina.^{9,10} These trials were designed to compare the performance of several screening tools in populations with variable prevalence of CIN and CC and with different facilities to conduct organized screening. The standardized structure and test quality of the two studies, along with comparable epidemiological characteristics of the women in the two cohorts, allowed us to pool the data-sets to create a combined cohort of over 15,000 women. We then assessed the performance of different standalone screening tests and multiple sequential combinations of these tests in the detection of CC and its precursor lesions. We have also produced simulations based on the data, and summarize the available information on the optimal screening strategy for CC in each setting.

SUBJECTS AND METHODS

Participants

The analysis is based on a combined sample of 3187 women from the NIS study, and 12,114 women in the LAMS studies. Both studies were international multicentre trials testing optional screening tools in low-resource settings of three NIS countries (Russia, Belarus and Latvia)¹¹ and in two Latin American countries (Brazil and Argentina).⁹ Both studies shared similar patient accrual, data collection methods and test standards, facilitating the data merger. Cytology was universal for all women; the number of women screened with each of the other tests varied for logistic and budgetary reasons (see Table 2).^{9,11}

The 3187 women in the NIS study cohort attended six outpatient clinics between 1998 and 2002. This study sample comes from three distinct backgrounds: (i) women participating in CC screening; (ii) women attending gynaecology outpatient clinics; and (iii) women examined at sexually transmitted disease clinics. The mean age at enrolment was 32.6 (± 10.7 SD) years (median 30.6, range 15–85 years).¹¹ All eligible women underwent a Pap smear and

were tested for HR-HPV using HC2 assay. The first 1500 women were also tested with polymerase chain reaction (PCR) and confirmative hybridization. Patients with ASC-US or higher Pap had biopsy confirmation at baseline.¹¹

The LAMS study combined a population-based, cross-sectional study and a prospective cohort study of women in regions with different (low, intermediate, high) incidence of CC in Brazil and Argentina.⁹ In the first phase (the original LAMS Study cohort), the four clinics examined 12,114 women between February 2002 and June 2003. Mean age at enrolment was 37.9 years (range 14–67; median 37.7 years). In this trial, eight different diagnostic tests were compared. Cervical cytology (conventional and liquid-based cytology [LBC]) was compared (1) with four optional screening tools suggested for low-resource settings (a) VIA, (b) visual inspection with Lugol iodine (VILI), (c) cervicography, and (d) screening colposcopy; and (2) with the new molecular diagnostic tools (HPV testing by HC2), performed (a) in samples collected by physicians, and (b) in those taken by self-sampling devices.^{12,13} Women testing positive with any of these techniques were examined by colposcopy at the second visit. In addition, a 5% random sample of all test-negative (Pap, VIA, VILI, HC2) women was submitted to colposcopy to assess false-negative exams, and 20% of baseline HC2-negative women were referred for new HC2 to assess the rates of incident HPV infections.

Tests and test providers

Screening tests were performed by trained professionals. In both studies, epidemiological data were obtained by doctors with formal training and previous expertise in obstetrics and gynaecology. Screening tests were performed by the same professional who collected the clinical and epidemiological data (patient history).

Cervical cytology (Pap test)

In the NIS study, all women were examined using conventional Pap smear only,¹¹ whereas in the LAMS study, three methods were used: conventional Pap and two different LBC techniques (DNA-Citoliq[®]; Digene Brazil, Sao Paulo, Brazil and SurePath[®]; TriPath, Durham, NC, USA). In our analysis we have used only the results of the conventional Pap test, which were available from all patients of the combined NIS-LAMS cohort. Results from the LAMS study were classified using the Bethesda System of 2001 (TBS 2001). Modified Papanicolaou classification was used for the NIS study, subsequently transformed to TBS 2001 to enable joint analyses.

Other screening tests

The technical aspects pertaining to VIA and VILI, cervicography, detection of HPV DNA with HC2 assay, and detection and quantification of HR-HPV genotypes by realtime PCR have been described in full detail in previous reports from the LAMS and NIS study groups.^{9–13}

Directed punch biopsy

Directed punch biopsies (and cone biopsies) were fixed in formalin, embedded in paraffin and processed into 5- μ m-thick haematoxylin–eosin-stained sections for light microscopy, following the routine procedures. All biopsies were examined in the Pathology Departments, and diagnosed using the commonly agreed CIN nomenclature. For the study purposes, the pathologists were also asked to notify the morphological changes suggestive for the presence of HPV in cases with no CIN, i.e. HPV-NCIN (=flat condyloma).

Statistical analysis

All data analyses were performed with the R environment for statistical computing¹⁴ and using 95% confidence intervals (95% CI) and $P = 0.05$ throughout. We defined arbitrary thresholds for Pap tests (atypical squamous [ASC] or higher, low-grade squamous intraepithelial lesion [LSIL] or higher, high-grade squamous intraepithelial lesion [HSIL]), and used the customary 1 relative light unit threshold for HC2. VIA was considered positive when the image, second to the examiner subjective assessment, was classified as suggestive of condyloma, cervical intraepithelial neoplasia (CIN) 1, CIN 2, CIN 3 or cancer.⁷ We then calculated the number of expected colposcopies after each test or combination of tests (a positive test = colposcopy), and calculated the ratio of performed:expected colposcopies for each case (=verified positives). We considered the disease-free women who (1) had all tests negative; (2) had one or more tests positive but had a normal colposcopy and/or normal biopsy result. Test combinations were considered positive when both screening tests were positive (sequential testing). We then calculated the number of prevalent CIN1+ or CIN2+ lesions detected using the defined thresholds for each test or test combination. We calculated the performance indicators (sensitivity, specificity, positive [PPV] and negative [NPV] predictive values) with standard formulae, and used generalized formulae for verification bias correction, according to Begg and Greenes¹⁵ and Gray *et al.*¹⁶

The entire cohort was divided into two groups, 'verified' and 'non-verified'. The empirical probability of verification (i.e. receiving the gold standard test) was calculated for each test result category. Within each test result category, the observed frequency counts among verified patients were divided by the empirical probability to obtain unbiased estimates of the frequency counts in perfect study design, i.e. in a setting where all patients received the gold standard test.

The final corrected sensitivity was obtained with the following relation:

Corrected sensitivity

$$= \frac{v_{11} \div [(v_{11} + v_{12}) \div n_1]}{v_{11} \div [(v_{11} + v_{12}) \div n_1] + v_{21} \div [(v_{21} + v_{22}) \div n_2]}$$

where v_{11} = diseased verified women with 'positive' test; v_{12} = verified non-diseased women with 'positive' screening test, v_{21} = verified diseased women with 'negative' screening test; v_{22} = verified non-diseased women with 'negative'

screening test; n_1 = total number of women with 'positive' screening tests; n_2 = total number of women with 'normal' screening tests. Corrected specificity was calculated likewise. Confidence intervals were constructed using bootstrap re-sampling methodology. Two thousand replications were performed for each calculation to ensure consistent asymptotically.

RESULTS

Table 1 lists the main epidemiological and demographic characteristics of the women enrolled in the LAMS and NIS studies. The mean age at enrollment was slightly higher (37.8 versus 32.5 years) for the women in the LAMS study compared with women in the NIS cohort. Considering both studies together, around two-thirds of the women were aged 25–50 years. Most (72%) women started their sexual activity at the age 17+ years, and approximately the same proportion were living with a partner at the time of the interview. In the NIS cohort, the proportion of women reported to have started sexual activity before 14 years of age was lower (1.2% versus 8.8%) than that in the LAMS cohort. Nulliparous women comprised more than 40% of the sample in both studies, while approximately 21% of the women in the LAMS cohort had three or more offspring, compared with only 3.0% in women of the NIS cohort.

Table 1 Key epidemiological and clinical characteristics of the women

Region of origin	No. (%)	Study
LAMS		
Buenos Aires (Argentina)	3437 (22.46)	LAMS
Porto Alegre (Brazil)	3043 (19.89)	LAMS
São Paulo (Brazil)	3000 (19.60)	LAMS
Campinas (Brazil)	2634 (17.2)	LAMS
Novgorod (Russia)	1088 (7.11)	NIS
Moscow (Russia)	1062 (6.94)	NIS
Minsk (Belarus)	590 (3.85)	NIS
Riga (Latvia)	447 (2.92)	NIS
	LAMS	NIS
	No. (%)	No. (%)
Total	12114 (79.2)	3187 (20.8)
Age (years)		
Mean (SD)	37.8 (11.2)	32.5 (10.7)
18–20	385 (3.2)	364 (11.4)
21–35	4601 (38.0)	1638 (51.4)
36–50	4907 (40.6)	1009 (31.7)
Above 50	2204 (18.2)	176 (5.5)
Onset sexual activity		
14 or younger	1063 (8.8)	36 (1.2)
15–16	2667 (22.1)	403 (14.0)
17–19	4702 (39.0)	1328 (46.0)
20 and older	3638 (30.1)	1118 (38.8)
Marital status		
Single	3814 (31.6)	760 (26.6)
Living with partner	8270 (68.4)	2096 (73.4)
Parity		
Zero	5021 (41.5)	1163 (40.5)
One	2331 (19.3)	1030 (35.9)
Two	2109 (17.4)	590 (20.6)
Three or more	2635 (21.8)	86 (3.0)

LAMS, Latin American Screening Study; NIS, New Independent States

Table 2 Verification of positive test results and consequent CIN detection

Screening strategy	Women screened	Colposcopy expected	Colposcopy performed (%)	Performed/expected	Colposcopy expected (age ≥35)	Colposcopy performed (age ≥35)	Performed/expected (age ≥35)	CIN2+ detected	CIN2+ detected age ≥35
LSIL or higher	15,187	558	452	0.81	338	274	0.81	185	143
ASC or higher	15,187	1150	854	0.74	649	497	0.76	221	157
HC2 positive	7498	1767	1279	0.72	747	596	0.79	169	102
PCR positive	17,74	689	361	0.52	312	147	0.47	71	41
LSIL or higher and HC2 positive	7453	297	234	0.78	157	121	0.77	103	74
ASC or higher and HC2 positive	7453	497	371	0.74	205	157	0.76	120	78
HSIL	15,187	202	162	0.80	167	133	0.79	127	111
VIA positive	12,093	1395	1357	0.97	802	787	0.98	83	52
VIA and LSIL	12,085	85	84	0.98	59	58	0.98	47	40
VIA and ASC or higher	12,085	154	153	0.99	95	94	0.98	152	47
VIA and HC2 positive	4411	146	142	0.97	64	64	1	59	15

CIN, cervical intraepithelial neoplasia; PCR, polymerase chain reaction; VIA, visual inspection with acetic acid; HC2, Hybrid Capture 2

Table 2 gives detail on the data from which the performance figures for the different screening tests were calculated. The proportion of positive tests verified with colposcopy ranged from 52% for PCR to 99.3% for women who had a positive VIA associated with ASC, or higher. As per study design, colposcopy was performed at the same visit as VIA or VILI, yielding high positive verification proportions (see discussion). Among women with no VILI or VIA, the verification of positive screening modalities was on average 76% in the whole cohort and almost 78% in women aged 35 years or older. The number of CIN1+ and CIN2+ detected for each screening strategy is also displayed.

Table 3 shows the corrected and uncorrected performance of the different standalone and combined screening strategies, having CIN2+ as the screening endpoint (cut-off). HPV detection used alone (by HC2 or PCR) was the most sensitive technique. There was a slight improvement (88.5–92.7%) in HC2 sensitivity in women ≥35 years of age. After Bayesian correction, the HPV detection techniques sustained their superiority in terms of sensitivity (for HC2, 79.0% [95% CI 71.2–86.8%] in women <35 years and 85.1% [95% CI 78.4% to 91.8%] in those ≥35 years). HPV detection, when used alone, was also the least specific screening tool, regardless of the age group (69.9% [95% CI 66.5–72.8%] and 86.4% [95% CI 84.6–88.2%], after correction, in each age stratum). However, in women above 35 years of age, when combined with the Pap test, at both ASC and LSIL thresholds, good balances between sensitivity and specificity were obtained (corrected sensitivity [LSIL + HC2] = 44.8% [95% CI 38.1–51.6%], specificity = 98.6 [98.3–99.0%]). Of the test combinations, that of Pap smear (LSIL threshold) with HC2 was the best option in terms of specificity (98.7% [95% CI 98.3–99.0%], after correction). However, in this age stratum, Pap alone performed almost similarly in terms of specificity (98.2% – [95% CI 97.8–98.6%], after correction), but used alone had higher sensitivity than when combined with HC2 (54.8% [95% CI 47.9–61.9%] alone versus 44.8% [95% CI 38.1–51.5%] when combined). This loss in sensitivity among women ≥35 years is counter-balanced by a gain in PPV (47.9% alone versus 61.1% when combined). In women <35 years, the combination of Pap (LSIL threshold) and HC2 yielded a corrected sensitivity of only 22.7% (95% CI 15.3–30.2%), compared with 32.8% (95% CI 24.4–40.3%) for Pap (LSIL) alone. However, the gain in PPV observed when combining Pap and HC2 in women ≥35 years was not observed in the younger age stratum: 29% for Pap (LSIL) alone versus 29.7% for Pap and HC2 combined. Using a higher threshold for Pap (HSIL) resulted in unacceptably low corrected sensitivity in women <35 years (12.7%; 95% CI 7.4–18%). This figure was better in women ≥35 years (29.2%; 95% CI 22.8–35.5%), but still much lower than that obtained with less aggressive Pap thresholds (LSIL or ASC). By contrast, intermediate values of sensitivity and specificity were obtained with VIA. While combined with Pap tests or HC2, the improvement in specificity was counter-balanced by losses in sensitivity and PPV. Except for VIA, all screening strategies were more accurate, with better PPV and sensitivity in women ≥35 years than in their younger counterparts.

Table 3 Performance of different screening strategies for the detection of CIN 2 or worse (CIN2+)

Screening strategy	Age < 35 years					Age > 35 years						
	Sensitivity	(Low to high)	Specificity	(Low to high)	NPV	PPV	(Low to high)	Specificity	(Low to high)	NPV	PPV	(Low to high)
LSIL or higher	47.68	(39.70 to 55.60)	94.15	(93.30 to 95.00)	28.98	97.3	(66.70 to 79.20)	96.13	(95.53 to 96.74)	47.98	98.64	(95.53 to 96.74)
LSIL or higher (corr)	32.8	(24.4 to 40.3)	96.8	(96.13 to 97.41)	28.98	97.3	(47.93 to 61.87)	98.21	(97.80 to 98.63)	47.98	98.64	(97.80 to 98.63)
ASC or higher	66.22	(58.68 to 73.73)	87.65	(86.41 to 88.89)	21.09	98.11	(74.51 to 85.69)	91.13	(90.21 to 92.04)	30.60	98.94	(90.21 to 92.04)
ASC or higher (corr)	53.88	(45.93 to 61.83)	92.25	(91.25 to 93.26)	21.09	98.11	(59.80 to 73.02)	95.43	(94.76 to 96.11)	30.60	98.94	(94.76 to 96.11)
HC2 positive	88.46	(82.32 to 94.60)	53.00	(49.57 to 56.43)	11.45	98.52	(87.90 to 97.60)	74.06	(71.76 to 76.36)	17.37	99.42	(71.76 to 76.36)
HC2 positive (corr)	78.99	(71.16 to 86.81)	69.69	(66.54 to 72.85)	11.45	98.52	(78.44 to 91.75)	86.43	(84.64 to 88.23)	17.37	99.42	(84.64 to 88.23)
PCR positive	91.11	(82.79 to 99.42)	30.36	(21.16 to 39.56)	16.26	95.83	(67.74 to 89.94)	42.78	(32.78 to 52.78)	26.97	88.29	(32.78 to 52.78)
PCR positive (corr)	75.93	(63.44 to 88.42)	58.61	(48.76 to 68.46)	16.26	95.83	(38.52 to 65.68)	71.92	(62.83 to 81.00)	26.97	88.29	(62.83 to 81.00)
LSIL or higher and HC2 positive	35.77	(27.30 to 44.24)	96.50	(95.84 to 97.16)	29.72	97.32	(48.41 to 65.43)	94.07	(92.43 to 95.70)	61.15	93.01	(92.43 to 95.70)
LSIL or higher and HC2 positive (corr)	22.75	(15.34 to 30.16)	98.11	(97.62 to 98.61)	29.72	97.32	(38.15 to 51.56)	98.65	(98.30 to 99.01)	61.15	93.01	(98.30 to 99.01)
ASC or higher and HC2 positive	52.18	(43.04 to 61.30)	93.07	(92.12 to 94.02)	22.90	98.01	(55.41 to 72.45)	97.40	(96.89 to 97.90)	43.82	98.84	(96.89 to 97.90)
ASC or higher and HC2 positive (corr)	38.17	(29.29 to 47.05)	95.95	(95.22 to 96.69)	29.90	70.20	(35.51 to 53.14)	98.81	(98.47 to 99.16)	43.82	98.84	(98.47 to 99.16)
HSIL	22.51	(15.85 to 29.17)	99.20	(98.89 to 99.51)	58.62	96.25	(53.83 to 69.34)	98.86	(98.50 to 99.23)	72.09	98.19	(98.50 to 99.23)
HSIL (corr)	12.67	(7.36 to 17.98)	99.60	(99.38 to 99.82)	58.62	96.25	(32.67 to 48.33)	99.51	(99.27 to 99.75)	72.09	98.19	(99.27 to 99.75)
VIA positive	48.27	(37.77 to 58.77)	73.96	(72.02 to 75.90)	5.83	97.71	(34.40 to 54.72)	84.11	(82.70 to 85.51)	7.80	98.04	(82.70 to 85.51)
VIA positive (corr)	32.15	(22.34 to 41.97)	84.83	(83.24 to 86.41)	5.83	97.71	(18.06 to 36.23)	91.94	(90.90 to 92.99)	7.80	98.04	(90.90 to 92.99)
VIA and LSIL	11.86	(6.02 to 17.69)	99.15	(98.83 to 99.48)	35.89	96.57	(22.54 to 40.30)	99.71	(99.52 to 99.89)	78.57	97.75	(99.52 to 99.89)
VIA and LSIL (corr)	5.97	(1.69 to 10.24)	99.6	(99.37 to 99.82)	35.89	96.57	(9.02 to 23.07)	99.88	(99.76 to 100.00)	78.57	97.75	(99.76 to 100.00)
VIA and ASC or higher	22.01	(14.23 to 29.79)	98.06	(97.56 to 98.56)	30.00	97.09	(25.10 to 43.52)	99.57	(98.62 to 99.32)	52.23	97.88	(98.62 to 99.32)
VIA and ASC or higher (corr)	11.69	(5.66 to 17.72)	99.08	(98.73 to 99.42)	30.00	97.09	(10.39 to 25.24)	99.97	(99.34 to 99.80)	52.23	97.88	(99.34 to 99.80)
VIA and HC2 positive	16.07	(6.45 to 25.69)	98.85	(98.46 to 99.24)	11.25	98.39	(6.39 to 23.45)	98.84	(98.45 to 99.24)	23.80	97.96	(98.45 to 99.24)
VIA and HC2 positive (corr)	8.24	(1.03 to 15.44)	98.85	(98.46 to 99.24)	11.25	98.39	(0.45 to 12.05)	99.55	(99.31 to 99.80)	23.80	97.96	(99.31 to 99.80)

DISCUSSION

Our study represents a major attempt to evaluate alternative strategies for CC screening in low- and middle-resource settings. In general terms, comparing several standalone and combined screening strategies, HC2 was by far the most sensitive tool, whereas the Pap test was highly specific at the LSIL threshold in diagnosing CIN2+. However, our results must be interpreted in the light of the variable screening options and epidemiological peculiarities of the screened populations. For instance, combining HC2 and the Pap test at ASC threshold resulted in a well balanced option for women ≥ 35 years, whereas this option was unacceptably non-specific among younger women. Overall, for the studied population, the PPV of the visual tests (VIA, VILI) was also unsatisfactory, largely due to the relatively low prevalence of CIN. In a previous report on women screened in the LAMS study, we detected a similar trend in the performance indicators, although the corrected figures are not directly comparable because the method used for verification bias correction was different in these two studies.¹⁷

The most convenient method for assessing the population effectiveness of a screening is programme sensitivity.¹⁷ However, programme sensitivity is dependent on the sensitivity of the screening test (or the combination of tests), together with the compliance with further follow-up and diagnostic work-up (verification), and the natural history of disease.¹⁸ However, sensitivity alone does not cope with the other important requirements of a screening programme, most notably the impact of false-negatives in the screening, the effects of over-screening, and the medical and legal consequences of false-positives.¹⁹ Bearing this in mind, we compared our cohort with other recent major studies to assess the potential implications of these different data. For instance, our corrected specificity estimate for the Pap test (LSIL and ASC thresholds) (98.8%) is higher than that (82.1%) reported in studies with comparable design conducted in the USA.^{20,21} These differences highlight the importance of considering the disease dynamics in the screened population. In the two US studies, the prevalence of CIN2+ was as low as $<1\%$, whereas in our combined LAMS-NIS cohort, 303 CIN2+ cases were found among 15,301 women (1.9%), but considering that on average 75% of the positive tests were verified, we can estimate a 2.5% true prevalence of CIN2+ in the entire population. Mathematical dissimilarities in the verification bias correction procedures between the two studies, combined with differences in accuracy in the interpretation of the Pap tests, may all have contributed to the observed differences.

Correction for verification should be a major concern in cross-sectional exam validity studies with incomplete designs. We chose the Bayesian approach to estimate the performance indicators because this method may be the most appropriate when the number of verified negatives is very low.^{15,16} In our study, the percentage of verified negatives only rarely surpassed 5% (e.g. for VIA). In a previous report¹⁷ using the algorithm for verification bias correction proposed by Reichenheim and Ponce de Leon,²² the corrected sensitivity of the Pap smear (HSIL threshold) in detecting CIN2 or worse in women above 35 years of age was only 17.2% (95% CI 13.8–20.6%), in contrast to the uncorrected figure (67.4%; 95% CI 56.8–76.8%). The mathematical

output of this type of correction method seems to heavily underestimate the sensitivity of the test when there is a small proportion of verified negatives, although this effect is less marked for the specificity of the test.

In a recent meta-analysis of 25 studies with comparable study subjects (women aged 18–70 years, participating in CC screening programmes, not previously followed-up for cytological abnormalities), types of interventions (HPV tests + cytology) and verification of disease status (colposcopy and/or biopsy on at least all Pap smear or HPV test positives),²³ the sensitivity of HPV testing was higher than that of cytology even at the ASC threshold, whereas the specificity was significantly higher for cytology at the threshold of LSIL. This advantage was much less obvious at the ASC cut-off. In our study, we detected similar levels in sensitivities of these two tests, but our cytology was far more specific than HPV test at both ASC and LSIL thresholds, before correction for verification bias. After correction, our estimated specificity for HC2 is closer to the figures from the meta-analysis, highlighting the importance of using appropriate mathematical corrections when the gold standard is not applied to the women with negative screening test results (i.e. those considered *a priori* true negatives). In this meta-analysis, the pooled specificity of HC2 was 90.4% (95% CI 87.1–93.6%) for CIN3, while we obtained a corrected specificity of 86.4% (95% CI 84.6–88.3%) among women ≥ 35 years of age. All our estimates were made using CIN2+ as the screening endpoint.

By contrast, our experience with the visual inspection techniques (VIA, VILI) was far less encouraging than the Indian and African experience.^{5,8} This is probably due to the marked contrasts between the Indian and Sub-Saharan populations and the LAMS-NIS cohort in prevalence of CIN2+ lesions (high in the former, intermediate in the latter). Another possible interpretation of this reduced performance, as compared with that described in other publications, is that differences in test standardization might have been responsible for a large number of false-positives. Our corrected performance indicators for VIA were favoured by the almost perfect verification of the VIA-positive cases, resulting in only minor corrections in sensitivity (downwards) and specificity (upwards). These tests allow for immediate confirmation of the findings in settings where colposcopy is available at the screening site or nearby.⁸ In our setting (LAMS), this was relevant. Despite efforts to obtain colposcopy confirmation of all screen-positive women, we were able to verify only approximately 75% of these. This number was much lower for PCR+ women (below 50%) (PCR done only for a small fraction of women, as a supplement to HC2), which resulted in more marked deviation in the performance figures obtained after correction for verification bias. In clinical practice, the failure to provide the results of Pap test and HPV assay immediately is a potential shortcoming, especially in low-resource settings, where patient compliance is a problem, and many women never return to obtain their test results.

The logical next step should be to combine the Pap test (with low sensitivity but high specificity) with HPV DNA testing, with high sensitivity but lower specificity,²³ to create a screening strategy that should cope with the known inherent problems of the standalone tests. Many reports advocating the feasibility of this strategy have been

published.^{17,20,21} Meta-analysis of the pooled data from these studies, however, suggests that while the combined use of Pap test and HPV DNA assay has the highest sensitivity, it also has the lowest specificity.²³ In our study we did not consider screening test combinations where a positive test in only one of the two tests resulted in colposcopy referral. In our previous reports,^{7,17} we tested those options, with unsatisfactory results in specificity as an outcome, thus confirming the data of the meta-analysis. Also other recent studies have assessed the possibility of referring for colposcopy all women who test positive for HC2 or ASC-US,²⁰ reporting excellent sensitivity but unacceptably low specificity. This suggests that adding HPV DNA testing to cytology increases the sensitivity of the latter, but will seriously compromise the high specificity of the Pap test. However, repeat positive HC2 is an FDA-approved recommendation for colposcopy in the USA, because repeat ASC mandates colposcopy in several countries. We could not evaluate these repeat test options in our combined cohort, because most women were excluded from the prospective follow-up after a normal colposcopy.

The explanation for this unexpected result of the pooled data for Pap and HPV test combinations is the different performance of the tests in different age groups, emphasizing that screening strategies for young women should be different from those for older women.¹⁷ This was confirmed in our recent report, where all standalone tests and combined HC2 and Pap performed significantly better among older (>35 years of age) women, using the one-of-two test positive option.¹⁶ In our present analysis, co-testing positive with HC2 and Pap smear did not efficiently improve the overall specificity of the combination, as shown in other studies.^{20,21} Also, we did not test the combination of HC2 and Pap test at the HSIL threshold, because most women with HSIL also test HPV-positive. Among younger women, the Pap test (LSIL threshold) was far more sensitive as a standalone test in detecting CIN2+ than when combined with HC2, and the gain in specificity was minimal for the combination. As expected from our previous report,¹⁷ all standalone tests and test combinations showed better performance in the subcohort of women ≥ 35 years of age, and the marked difference between Pap alone and the Pap-HC2 combination seen in younger women almost disappeared in the older women. In the economic context of the LAMS-NIS countries, the addition of HPV testing without significant gains in performance would not be advisable. Moreover, if such a test combination results in increased referral for colposcopy, incremental costs would arise. The only (mathematical) advantage derived from the co-testing with HC2 and Pap smear was that our simulations yielded better PPVs than those of either test used alone. These improved PPVs could prove to be particularly meaningful in a screening setting, because an optimal screening test should have the highest possible PPV, i.e. to detect only positive lesions with no (or little) need for other confirmatory tests that would result in incremental costs.²⁴

Co-testing with HC2+ and ASC+ was clinically the most advantageous combination in the simulations performed. Because the proportion of verified positives for HC2 and ASC+ was quite similar in all age strata, we may consider

the corrected performance figures without having to take into account distortions caused by dissimilarities in verification ratios (i.e. groups with lower positive verification ratios would display lower values of sensitivity and higher specificity after correction). Our results also indicate that co-testing with HPV testing and Pap smear may be advisable only in settings where the prevalence of HPV is relatively low and where economic resources are not constrained. In our study populations neither of these conditions applies, and we found that co-testing with HC2 and Pap smear was advantageous only in women ≥ 35 years of age with ASC. HPV testing would also be advantageous for women ≥ 35 years of age with LSIL, because of an increased PPV of this test combination compared with that of the Pap test alone. We also detected that in our context, where the prevalence of disease is not as high as it is in some previously studied African and Indian populations, the main advantage of VIA (i.e. the possibility of immediate confirmation of positive test results) brings little relief to the unacceptably low PPV of this test, relative to the other options.

Our results may be valuable for health-care planners in medium-resource countries. In these contexts, the use of HPV testing as a screening tool may not be advisable, except in a few specific situations. In most instances, due to the high prevalence of HPV even among women ≥ 35 years of age, the specificity of the HC2-Pap combination is reduced to unacceptably low levels. The Pap test is a highly specific option to screen for CC in populations with a disease prevalence ranking within the medium global range. In spite of the relatively low sensitivity of cytology compared with that of, for example, HPV testing, the protracted natural history of CIN makes repeated rescreening an acceptable option to improve the overall sensitivity of the screening programme. On the other hand, even small decreases in screening specificity can make the programme impractical, by dramatically increasing the demand for confirmatory colposcopy. Co-testing with HPV in this scenario may yield slightly better PPVs in women ≥ 35 years of age with LSIL, but at a high incremental cost.

Further research should now aim to find the wisest way to combine the properties of HPV testing and cytology. Evidence suggests that the use of ASC-US cytology with HPV testing may avoid a large number of wasteful colposcopies.²⁵ The use of LSIL cytologies may also be effective in reducing the number of unnecessary colposcopies to 50% in women >35 years of age. Some recent studies also emphasize the potential role of HPV testing in future screening scenarios, and there are indications that one-time HPV-based screening is superior to Pap tests and VIA for reducing CC mortality.^{26,27} Although evidence favoring the use of HPV tests in several screening scenarios accumulates, a comment from the Discussion section of a 2007 meta-analysis: 'Caution should be taken to make sure that these findings (HPV testing in primary screening improving the control of cervical cancer) are applicable to all countries', is still contemporary and valid.²³ From the results of the present study, we cannot advocate using direct visual inspection (VIA) of the cervix in screening populations where the prevalence of CIN is not extremely high.

Authors' affiliations

Luis Otavio Sarian, Associate Professor of Gynecologic Oncology, Department of Obstetrics and Gynecology, Universidade Estadual de Campinas, Campinas, Brazil

Sophie Derchain, Associate Professor of Gynecologic Oncology, Department of Obstetrics and Gynecology, Universidade Estadual de Campinas, Campinas, Brazil

Irena Shabalova, Professor of Clinical Cytology, Russian Academy of Post-Graduate Medical Education, Moscow, Russia

Silvio Tatti, Professor of Gynecology and Obstetrics, First Chair of Gynecology, Hospital de Clinicas, Buenos Aires, Argentina

Paulo Naud, Associate Professor of Obstetrics and Gynecology, Hospital de Clinicas de Porto Alegre, and Department of Gynecology and Obstetrics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil

Adhemar Longatto-Filho, Biomedical Senior Researcher, Department of Pathology, Medical School of São Paulo University (LIM14) and Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, Braga, Portugal

Stina Syrjänen, Professor & Chair of Oral Pathology, Department of Oral Pathology, Institute of Dentistry, University of Turku, Finland

Kari Syrjänen, Professor and Senior Scientist, Department of Oncology & Radiotherapy, Turku University Hospital, Turku, Finland

ACKNOWLEDGEMENTS

The NIS Cohort study was supported by the INCO-Copernicus Program of the European Commission (Contract No. ERB IC15-CT98-0321), and the LAMS study by the European Commission, INCO-DEV Program (Contract# ICA4-CT-2001-10013). We express special thanks to the former Digene Corp. for providing the Hybrid Capture analyser, samplers and the test kits at our disposal. We express our thanks to all women who participated in these two cohort studies. Finally, all the members of the NIS and LAMS Study research groups are acknowledged for their invaluable contribution to these studies.

The NIS Cohort Study Group:

K Syrjänen¹, S Syrjänen², I Shabalova^{3,4}, N Petrovichev³, V Kozachenko³, T Zakharova³, J Pajanidi³, J Podistov³, G Chemeris³, L Sozaeva⁴, E Lipova⁴, I Tsidaeva⁴, O Ivanchenko⁵, A Pshpurko⁵, S Zakharenko⁶, R Nerovjina⁷, L Kljukina⁸, O Erokhina⁸, M Branovskaja⁹, M Nikitina¹⁰, V Grunjberga¹⁰, A Grunjberg¹⁰, A Juschenko¹⁰, R Santopietro¹¹, M Cintorino¹¹, PTosi¹¹

¹Department of Oncology and Radiotherapy, Turku University Hospital, Turku, Finland; ²Department of Oral Pathology, Institute of Dentistry, and MediCity Research Laboratory, University of Turku, Finland; ³N.N. Blokhin Cancer Research Centre of Russian Academy of Medical Sciences (RAMS), Moscow, Russia; ⁴Russian Academy of Post-Graduate Medical Education, Moscow, Russia; ⁵Novgorod Clinical Regional Hospital, Centralised Cytology Laboratory, Novgorod, Russia; ⁶Novgorod Municipal Dermato-venerological Dispensary, Department of Gynaecology, Novgorod, Russia; ⁷Novgorod Female Consultative Outpatient Hospital, Department of Gynaecology, Novgorod, Russia; ⁸Research Institute of Oncology and Medical Radiology, Republican Centre of Clinical Cytology, Minsk, Belarus; ⁹Minsk State Medical Institute, Department of Gynaecology and Obstetrics, Minsk, Belarus; ¹⁰Latvian Cancer Centre, Department of Gynaecology, and Laboratory of Cytology, Riga, Latvia;

¹¹Department of Human Pathology and Oncology, University of Siena, Italy.

The LAMS Study Group:

K Syrjänen¹, P Naud², S Derchain³, C Roteli-Martins⁴, A Longatto-Filho⁵, STatti⁶, M Branca⁷, M Eržen⁸, L S Hammes², J Matos², R Gontijo³, L Sarian³, J Bragança³, F C Arlindo⁴, M Y S Maeda⁵, A Lörincz⁹, G B Dores¹⁰, S Costa¹¹, S Syrjänen¹²

¹Department of Oncology & Radiotherapy, Turku University Hospital, Turku, Finland; ²Hospital de Clinicas de Porto Alegre, Brazil; ³Universidade Estadual de Campinas, Campinas, Brazil; ⁴Hospital Leonor M de Barros, Sao Paulo, Brazil; ⁵Instituto Adolfo Lutz, Sao Paulo, Brazil and (ALF) Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, Braga, Portugal; ⁶First Chair Gynecology Hospital de Clinicas, Buenos Aires, Argentina; ⁷Unit of Cytopathology, National Centre of Epidemiology, Surveillance and Promotion of Health, National Institute of Health (ISS), Rome, Italy; ⁸SIZE Diagnostic Center, Ljubljana, Slovenia; ⁹Digene Corp., Maryland, USA; ¹⁰Digene Brazil, Sao Paulo, Brazil; ¹¹Department of Obstetrics and Gynecology, S.Orsola-Malpighi Hospital, Bologna, Italy; ¹²Department of Oral Pathology, Institute of Dentistry, University of Turku, Finland.

REFERENCES

- Solomon D, Schiffman M, Tarone R, ALTS Study group. Comparison of three management strategies for patients with atypical squamous cells of undetermined significance: baseline results from a randomized trial. *J Natl Cancer Inst* 2001;**93**:293-9
- Stoler MH, Schiffman M. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. *JAMA* 2001;**285**:1500-5
- Smith JS, Lindsay L, Hoots B, Keys J, Franceschi S, Winer R. Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions: a meta-analysis update. *Int J Cancer* 2007;**121**:621-32
- Miller AB, Nazeer S, Fonn S, et al. Report on consensus conference on cervical cancer screening and management. *Int J Cancer* 2000;**86**:440-7
- Arbyn M, Ronco G, Cuzick J, Wentzensen N, Castle PE. How to evaluate emerging technologies in cervical cancer screening? *Int J Cancer* 2009;**125**:2489-96
- Wright TC Jr, Cox JT, Massad LS, et al. 2001 Consensus guidelines for the management of women with cervical cytological abnormalities. *JAMA* 2002;**287**:2120-9
- Sarian L, Derchain SM, Naud P, et al. Evaluation of visual inspection with acetic acid (VIA), Lugol's iodine (VILI) cervical cytology and HPV testing as cervical screening tools in Latin America. This report refers to partial results from the LAMS (Latin American Screening) study. *J Med Screen* 2005;**12**:142-9
- Muwonge R, Walter SD, Wesley RS, et al. IARC Multicentre Study Group on Cervical Cancer Early Detection. Assessing the gain in diagnostic performance when two visual inspection methods are combined for cervical cancer prevention. *J Med Screen* 2007;**14**:144-50
- Syrjänen K, Naud P, Derchain SM, et al. Comparing PAP smear cytology, aided visual inspection, screening colposcopy, cervicography and HPV testing as optional screening tools in Latin America. Study design and baseline data of the LAMS Study. *Anticancer Res* 2005;**25**:3469-80
- Syrjänen S, Shabalova IP, Petrovichev N, et al. Age-specific incidence and clearance rates of high-risk human papillomavirus (HPV) infections in women subjected to HPV screening in the new independent states. *Int J STD & AIDS* 2005;**16**:217-33
- Syrjänen S, Shabalova IP, Petrovichev N, et al. Human papillomavirus testing and conventional PAP smear cytology as optional screening tools of women at different risk for cervical cancer in countries of former Soviet Union. *J Lower Genital Tract Dis* 2002;**6**:97-110
- Longatto-Filho A, Maeda MYS, Eržen M, et al. Conventional PAP smear and liquid-based cytology as screening tools in low-resource settings of Latin America. Experience of the Latin American Screening Study. *Acta Cytol* 2005;**49**:500-6

- 13 Hammes LS, Naud P, Passos E, *et al.* Value of the International Federation of Colposcopy and Cervical Pathology (IFCPC) terminology in predicting cervical pathology. *J Lower Genital Tract Dis* 2007;**11**:158–65
- 14 R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3–900051-07-0, See <http://www.R-project.org> (last checked 4 January 2010)
- 15 Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics* 1983;**39**:207–15
- 16 Gray R, Begg CB, Greenes RA. Construction of receiver operating characteristic curves when disease verification is subject to selection bias. *Med Decis Making* 1984;**4**:151–64
- 17 Syrjänen K, Naud P, Derchain SF, *et al.* Value of conventional Pap smear, liquid based cytology, visual inspection (VIA) and human papillomavirus (HPV) testing as optional screening tools among Latin American women <35 and >35 and >35 years of age. Experience from the LAMS Study. *Acta Cytol* 2008;**52**:641–53
- 18 Arbyn M, Dillner J, Schenck U, *et al.* Chapter 3: methods for screening and diagnosis. In: Arbyn M, Anttila A, Jordan J, Ronco G, Schenck U, Segnan N, Wiener H, Daniel J, von Karsa L, European Commission, eds. *European Guidelines for Quality Assurance in Cervical Cancer Screening*. Luxembourg: Office for Official Publications of the European Communities, 2008:69–152
- 19 Arbyn M, Kyrgiou M, Simoens C, *et al.* Peri-natal mortality and other severe adverse pregnancy outcomes associated with treatment of cervical intraepithelial neoplasia: a meta-analysis. *BMJ* 2008;**337**:a1284:1–11
- 20 Baseman JG, Kulasingam SL, Harris TG, *et al.* Evaluation of primary cervical cancer screening with an oncogenic human papillomavirus DNA test and cervical cytologic findings among women who attended family planning clinics in the United States. *Am J Obstet Gynecol* 2008;**199**:26.e1–8
- 21 Kulasingam SL, Hughes JP, Kiviat NB, *et al.* Evaluation of human papillomavirus testing in primary screening for cervical abnormalities: comparison of sensitivity, specificity, and frequency of referral. *JAMA* 2002;**288**:1749–57
- 22 Reichenheim ME, Ponce de Leon A. Estimation of sensitivity and specificity arising from validity studies with incomplete designs. *Stata J* 2002;**2**:267–79
- 23 Koliopoulos G, Arbyn M, Martin-Hirsch P, Kyrgiou M, Prendiville W, Paraskevaidis E. Diagnostic accuracy of human papillomavirus testing in primary cervical screening: A systematic review and meta-analysis of non-randomized studies. *Gynecol Oncol* 2007;**104**:232–46
- 24 dos Santos Silva I. *Cancer Epidemiology: Principles and Methods*. Lyon: IARC, 1999:355–83
- 25 Cuzick J, Arbyn M, Sankaranarayanan R, *et al.* Overview of human papillomavirus-based and other novel options for cervical cancer screening and developing countries. *Vaccine* 2008;**26**:K29–K41
- 26 Sankaranarayanan R, Nene BM, Shastri SS, *et al.* HPV screening for cervical cancer in rural India. *N Engl J Med* 2009;**360**:1385–94
- 27 Schiffman M, Wacholder S. From India to the world: a better way to prevent cervical cancer. *N Engl J Med* 2009;**360**:1453–5