DOI:10.1111/j.1365-2303.2011.00849.x

Large performance variation does not affect outcome in the Finnish cervical cancer screening programme

S. Lönnberg*, P. Nieminen†, L. Kotaniemi-Talonen*, H. Kujari‡, J. Melkko§, G. Granroth¶, M. Vornanen**, T. Pietiläinen††, J. Arola‡‡, J. Tarkkanen§§, T. Luostarinen* and A. Anttila*

*Finnish Cancer Registry, Helsinki, Finland, †Department of Obstetrics and Gynaecology, Helsinki University Central Hospital, Helsinki, Finland, ‡Lounais-Suomen Patologian Laboratorio Oy, Turku, Finland, §Suomen Syöpäyhdistyksen Sytologian Laboratorio, Oulu, Finland, ¶Syöpäjärjestöjen Sytologian Laboratorio, Vaasa, Finland, **Kotkan Syöpäpoliklinikka, Kotka, Finland, ††Puijon Solu- Ja Kudoslaboratorio Oy, Kuopio, Finland, ‡‡Syöpäjärjestöjen Poliklinikka, Helsinki, Finland and §§Kätilöopiston Sairaalan Patologian Laboratorio, Helsinki, Finland

Accepted for publication 29 November 2010

S. Lönnberg, P. Nieminen, L. Kotaniemi-Talonen, H. Kujari, J. Melkko, G. Granroth, M. Vornanen, T. Pietiläinen, J. Arola, J. Tarkkanen, T. Luostarinen and A. Anttila

Large performance variation does not affect outcome in the Finnish cervical cancer screening programme

Objective: Cytology screening for prevention of cervical cancer can reduce incidence and mortality by more than 80% in settings with good organization and rigorous quality control. Audit studies are essential for reaching and maintaining a high quality of screening. The aim of this study was to evaluate variation in performance indicators by screening laboratory and assess the impact on the effectiveness of screening as indicated by cervical intraepithelial neoplasia grade 3 and above (CIN3+) rates after a negative screen.

Methods: Seven cytology screening laboratories operating during 1990–1999 with a total of 953 610 screening tests performed were included in the study. By linking screening and cancer register files, all cases of CIN3+ diagnosed in the screened population were identified. For 395 CIN3+ cases with a preceding negative screen and 787 controls, a re-evaluation of smears was undertaken to uncover false negative screening tests. Performance parameters and rates of CIN3+ after a negative screen were analysed for interlaboratory heterogeneity.

Results: The rates of follow-up recommendations and referrals varied by up to 3.6- (2.8–10.2%) and 4.0-fold (0.03–0.12%), respectively. CIN1, CIN2 and CIN3+ screen detection rates differed by up to 8.5- (0.02–0.17%), 5.4- (0.05–0.25%) and 3.3-fold (0.05–0.18%). False negative rates determined by re-evaluation showed up to 2.1-fold differences (29–62%). Rates of CIN3+ after a negative screen (0.023–0.048%) and as a proportion of total CIN3+ (15–31%) in the screened population were low and did not vary significantly.

Conclusions: There were large variations in the sensitivity–specificity trade-off between laboratories, reflected in all performance indicators as well as in the test validity estimates of the re-evaluation phase, but not in screening effectiveness. Even though performance variations do not always have an impact on the effectiveness of screening, they lead to variations in cost, treatment and psychological burden, and should be addressed.

Keywords: quality control, cytology, uterine cervical neoplasms, mass screening, false negative tests, sensitivity and specificity

Correspondence:
Stefan Lönnberg, Finnish Cancer Registry,
Pieni Roobertinkatu 9, 00130 Helsinki, Finland
Tel.: +358-9-1353-3281; gsm +358-50-354-4910;
E-mail: stefan.lonnberg@cancer.fi

Introduction

Cervical screening based on conventional cytology has been effective in reducing incidence of and mortality from cervical cancer but requires a high level of

screening programme organization for good results.[1] Cytology itself often suffers from low reproducibility and interobserver agreement.[2–5] In order to achieve and maintain high effectiveness, continuous quality assurance of a screening programme is essential.[1] Key performance indicators, including attendance, referral and cervical intraepithelial neoplasia (CIN) detection rates, are readily available from screening records of good quality. The screening test itself can be evaluated by comparing results with a gold standard and calculating validity estimates. Finally, with linkage studies the association between screening history and incidence of disease can be analysed, yielding longitudinal test sensitivity estimates. Feedback produced by these evaluations and returned to the laboratories facilitates consistent diagnostic criteria and a reliable screening service.

The organized cervical cancer screening programme in Finland started in 1963 and became nationwide within a decade. Currently the target age group is 30–64 years, with nearly complete invitational coverage. The recommended interval between negative screens is 5 years. All women in the target age groups are invited with a personal letter regardless of their screening history. Significant opportunistic screening exists in parallel,[6] probably somewhat decreasing attendance for organized screening. The attendance rates have been around 70% for the past two decades, with lower, and decreasing, rates in the younger age groups. The screening test in use is conventional cytology, except for a currently running randomized controlled trial integrated into the programme, where a primary human papillomavirus (HPV) test with cytology triage is used in one study arm.[7–10] The screening programme has been successful, with a decrease of about 80% in cervical cancer incidence and mortality.[11]

In a study by Kotaniemi-Talonen *et al.*,[12] significant variation in some screening laboratory performance indicators was shown, but no corresponding variations in cancer incidence trends could be demonstrated. In the current study we assess the performance and validity of cytology screening and relate the results to screening outcome as defined by CIN3 and cancer case rates after a negative screen. This outcome measure may be more useful than incidence trends for quality assurance purposes because of a shorter response time. We have used clinical outcome to detect potential false negative smears, together with control smears from the healthy screening population, and blinded re-evaluations to establish the final cytology results.

Methods

The Mass Screening Registry (MSR) contains data on all screening invitations, visits and histologically confirmed findings. Cervical cancer, cervical intraepithelial neoplasia grade 3 (CIN3) and adenocarcinoma *in situ* (AIS) cases are reported by doctors and pathologists to the Cancer Registry. The current study is based on a linkage of these two registries over the period of 1990–1999, involving 2.3 million visits among 1.3 million women. The seven largest laboratories were enrolled in the study, providing access to smear archives and participating in the rereading of potentially false negative audit case smears. The study material amounted to a total of 953 610 screen samples.

In the 1990s the Papanicolaou classification was universally used in the Finnish screening programme for reporting cervical cytology, and was replaced in 2006 by Bethesda 2001 terminology (TBS2001).[13] Papanicolaou class II (Pap II) results were usually recommended for intensified follow-up, meaning re-invitation after 1 year instead of the normal screening interval of 5 years, and referral for colposcopy only if the finding was persistent over several tests, usually two or three. The Pap II category corresponds to reactive changes, atypical squamous cells of undetermined significance (ASC-US) and atypical glandular cells not otherwise specified (AGC-NOS) in the TBS2001 nomenclature. Pap III or higher corresponds to low-grade intraepithelial lesions (LSIL), atypical squamous cells where a high-grade lesion cannot be ruled out (ASC-H), high-grade intraepithelial lesions (HSIL) and carcinoma, warranting immediate referral. Under current guidelines, ASC-US results lead to intensified follow-up within 1 year and referral after two to three repeats, AGC-NOS results are followed-up within half a year and referred if persistent, whereas LSIL, ASC-H, HSIL and AGC-FN results lead to immediate referral. In this study, audit cases were defined as cervical cancers and CIN3/AIS lesions (CIN3+) diagnosed during 1990–1999 among women with a Papanicolaou class I or II result at previous screen without a referral for colposcopy. The audit case criteria were fulfilled in 474 cases. The participating laboratories were asked to locate the corresponding smears along with two control smears (the one preceding and the one following the case smear in sequence in their smear archives). Controls were thought to be particularly useful for the assessment of the specificity of

rereading. Of the requested case smears, 395 (83%) were traced and retrieved along with 787 controls.

The smears were arranged in a random sequence and blinded as to the status of case or control. The rereading was performed in the original laboratories as well as in a reference laboratory and recorded using TBS2001. The cytology laboratory at the Department of Obstetrics and Gynaecology at the Helsinki University Central Hospital was used as reference. One of two experienced cytotechnicians and one cytopathologist at the reference laboratory evaluated all smears in order to minimize interobserver variation in the reference readings. An expert panel resolved discrepancies in 366 out of 1182 smears (207 case smears and 159 controls) and provided final cytology to be used as the gold standard for test validity estimates. The panel consisted of five members: two from the original laboratory (a cytotechnician and a cytopathologist), two from the reference laboratory, and the last member was an external cytopathologist experienced in screening.

Statistical methods

For individual laboratories, exact, unconditional 95% confidence intervals for a difference of two related binomial proportions for pair-wise comparison of laboratory rereading and final cytology were calculated. Asymptotic confidence intervals were used for the whole material. Laboratories whose rereading results differed significantly from the final cytology using different cut-off points (ASC-US+ or LSIL+) as well as by case status were identified using these confidence intervals. StatXact 9 (Cytel Inc, Cambridge, MA, USA) was used for this analysis. Test validity estimates of the rereading phase were computed with reference to the final cytology result as the gold standard. Rereading sensitivity and 1-specificity were plotted and fitted using a simple logarithmic expression [$y = a - b^{*}\ln(x + c)$] in the OriginPro 7.5 SR1 (OriginLab Corporation, Northampton, MA, USA) software package. Interlaboratory heterogeneity of performance and outcome parameters were tested on material limited to the invitational years 1996–1999 as all laboratories were represented throughout this period leading to uniform follow-up time of audit cases across laboratories. Follow-up, referral, CIN detection rates and positive predictive values (PPVs) were modelled with logistic regression using the GENMOD procedure in SAS/STAT 9.1 (SAS Institute Inc, Cary, NC, USA) and the *P*-value of the likelihood ratio test of adding laboratory as a variable to the model of age and year is reported. For the proportion of audit cases out of all CIN3+ cases in the population, log-binomial regression models were used.

Results

In the study material, 1 312 139 invitations resulted in 953 662 visits, producing an overall attendance rate of 73%. The attendance rate ranged from 68% to 81% between laboratories. Variation in referral rates from 0.3% to 1.2% (4.0-fold) and in follow-up rates from 2.8% to 10.2% (3.6-fold) were followed by large variations in the CIN1, CIN2 and CIN3+ detection rates, up to 8.5-, 5.4- and 3.3-fold, respectively. PPVs for referral at CIN1+ were 24–64% and at CIN3+ 5–41% (Table 1). Screen-detected neoplasias (including invasive disease) are defined as diagnoses after referral from the invitational screening programme following an abnormal test. Symptomatic cases are not explicitly excluded, but usually they would be diagnosed outside the programme by way of patient initiative.

Out of a total of 474 identified CIN3+ audit cases, 395 (83%) were obtained for re-evaluation. One-third of these had originally been diagnosed as Pap II (Table 2). The definition of false negativity used here is based on the final cytology diagnosis established by concurrent independent rereading in two laboratories or, in the case of discrepancies, the expert panel. Because Pap II smears were included in the audit case smears, our primary cut-off for false negativity was LSIL+ (corresponding to Pap III or higher and including ASC-H). With these criteria, the false negative proportion of the audit case smears was 38% overall and ranged from 29% to 62% between laboratories (Figure 1). We also assessed the false negative proportion using a rereading threshold of ASC-US+ among the 252 case smears that were classified as normal (Pap I) by original cytology; the range was similar (29–59%), with an overall proportion of 40% (data not shown). According to rereading in the original screening laboratory, LSIL+ proportions of all audit cases ranged from 8% to 62% (Figure 1). In three laboratories (A, C and E) as well as overall, the difference between screening laboratory rereading and final cytology was significant.

Sensitivity and specificity of screening laboratory rereading was calculated with respect to the final cytology gold standard. The point estimates for test validity were widely dispersed (Figure 2). In general and as expected, a trade-off between sensitivity and

Table 1. Baseline screening data and cross-sectional performance parameters by laboratory

	A	B	C	D	E	F	G	All
Screening laboratory*	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)
Invitations	392 599 (100.0)	242 823 (100.0)	213 505 (100.0)	159 287 (100.0)	130 565 (100.0)	99 626 (100.0)	73 734 (100.0)	1 312 139 (100.0)
Visits	265 536 (67.6)	177 949 (73.3)	162 496 (76.1)	109 376 (68.7)	105 241 (80.6)	78 488 (78.8)	54 576 (74.0)	953 662 (72.7)
Smears[†]	265 536 (100.0)	177 948 (100.0)	162 478 (100.0)	109 354 (100.0)	105 241 (100.0)	78 481 (100.0)	54 572 (100.0)	953 610 (100.0)
Referral	3149 (1.2)	1711 (1.0)	1935 (1.2)	1241 (1.1)	393 (0.4)	472 (0.6)	161 (0.3)	9062 (1.0)
Intensified follow-up	18 402 (6.9)	5639 (3.2)	5250 (3.2)	9117 (8.3)	2936 (2.8)	7977 (10.2)	2419 (4.4)	51 740 (5.4)
Normal cytology	243 985 (91.9)	170 598 (95.9)	155 293 (95.6)	98 996 (90.5)	101 912 (96.8)	70 032 (89.2)	51 992 (95.3)	892 808 (93.6)
Screen-detected neoplasia	1195 (0.45)	540 (0.30)	461 (0.28)	592 (0.54)	229 (0.22)	224 (0.29)	102 (0.19)	3343 (0.35)
Cx Ca	54 (0.02)	13 (0.01)	6 (0.00)	12 (0.01)	6 (0.01)	9 (0.01)	4 (0.01)	104 (0.01)
CIN3/AIS	339 (0.13)	186 (0.10)	84 (0.05)	184 (0.17)	116 (0.11)	77 (0.10)	62 (0.11)	1048 (0.11)
CIN2	349 (0.13)	142 (0.08)	156 (0.10)	269 (0.25)	57 (0.05)	68 (0.09)	25 (0.05)	1066 (0.11)
CIN1	453 (0.17)	199 (0.11)	215 (0.13)	127 (0.12)	50 (0.05)	70 (0.09)	11 (0.02)	1125 (0.12)
CIN1/CIN3+ (ratio)	(1.15)	(1.00)	(2.39)	(0.65)	(0.41)	(0.81)	(0.17)	(0.98)
CIN1/CIN2+ (ratio)	(0.61)	(0.58)	(0.87)	(0.27)	(0.28)	(0.45)	(0.12)	(0.51)
PPV (CIN1+)	(37.9)	(31.6)	(23.8)	(47.7)	(58.3)	(47.5)	(63.4)	(36.9)
PPV (CIN2+)	(23.6)	(19.9)	(12.7)	(37.5)	(45.5)	(32.6)	(56.5)	(24.5)
PPV (CIN3+)	(12.5)	(11.6)	(4.7)	(15.8)	(31.0)	(18.2)	(41.0)	(12.7)
Test specificity (CIN1+)[‡]	(99.3)	(99.3)	(99.1)	(99.4)	(99.8)	(99.7)	(99.9)	(99.4)
Test specificity (CIN2+)[‡]	(99.1)	(99.2)	(99.0)	(99.3)	(99.8)	(99.6)	(99.9)	(99.3)
Test specificity (CIN3+)[‡]	(99.0)	(99.1)	(98.9)	(99.0)	(99.7)	(99.5)	(99.8)	(99.2)
Test specificity (CIN1+)[§]	(92.3)	(96.2)	(95.8)	(91.0)	(97.0)	(89.5)	(95.5)	(94.0)
Test specificity (CIN2+)[§]	(92.1)	(96.1)	(95.7)	(90.9)	(97.0)	(89.4)	(95.4)	(93.8)
Test specificity (CIN3+)[§]	(92.0)	(96.0)	(95.6)	(90.7)	(96.9)	(89.3)	(95.4)	(93.7)

Cx Ca, cervix carcinoma; CINx, cervical intraepithelial neoplasia, grade x; AIS, adenocarcinoma *in situ*; PPV, positive predictive value of referral; CIN1/3+, CIN1/3 or worse.
*Laboratories B and C participated with material from 1990 to 1999, A from 1993 to 1999, E and G from 1994 to 1999, F from 1995 to 1999 and D from 1996 to 1999.
[†]Some visits resulted in smears of unsatisfactory quality and are not reported here.
[‡]Referral as cut-off for test positivity.
[§]Intensified follow-up as cut-off for test positivity.

Table 2. The rereading material: cases and controls by laboratory, histology and original cytology

	A	B	C	D	E	F	G	All
Screening laboratory	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)
Smears without referral	262 387 (100.0)	176 237 (100.0)	160 543 (100.0)	108 113 (100.0)	104 848 (100.0)	78 009 (100.0)	54 411 (100.0)	944 548 (100.0)
Audit cases	181 (0.07)	42 (0.02)	74 (0.05)	53 (0.05)	70 (0.07)	29 (0.04)	25 (0.05)	474 (0.05)
Cx Ca	51 (0.02)	24 (0.01)	13 (0.01)	6 (0.01)	16 (0.02)	4 (0.01)	7 (0.01)	121 (0.01)
CIN3/AIS	130 (0.05)	18 (0.01)	61 (0.04)	47 (0.04)	54 (0.05)	25 (0.03)	18 (0.03)	353 (0.04)
Audit cases in rereading*	151 (83.4)	24 (57.1)	63 (85.1)	50 (94.3)	63 (90.0)	26 (89.7)	18 (72.0)	395 (83.3)
Cx Ca	29 (56.9)	9 (37.5)	5 (38.5)	4 (66.7)	9 (56.3)	2 (50.0)	0 (0.0)	58 (47.9)
CIN3/AIS	122 (93.8)	15 (83.3)	58 (95.1)	46 (97.9)	54 (100.0)	24 (96.0)	18 (100.0)	337 (95.5)
Cases	151 (100.0)	24 (100.0)	63 (100.0)	50 (100.0)	63 (100.0)	26 (100.0)	18 (100.0)	395 (100.0)
Pap II	36 (23.8)	7 (29.2)	20 (31.7)	28 (56.0)	25 (39.7)	16 (61.5)	11 (61.1)	143 (36.2)
Pap I	115 (76.2)	17 (70.8)	43 (68.3)	22 (44.0)	38 (60.3)	10 (38.5)	7 (38.9)	252 (63.8)
Controls	302 (100.0)	48 (100.0)	125 (100.0)	98 (100.0)	126 (100.0)	52 (100.0)	36 (100.0)	787 (100.0)
Pap II	12 (4.0)	4 (8.3)	4 (3.2)	9 (9.2)	4 (3.2)	9 (17.3)	2 (5.6)	44 (5.6)
Pap I	290 (96.0)	44 (91.7)	121 (96.8)	89 (90.8)	122 (96.8)	43 (82.7)	34 (94.4)	743 (94.4)
Total smears in rereading	453	72	188	148	189	78	54	1182

*Some identified audit case smears could not be traced for the rereading phase.

specificity can be seen in the distribution of test validity estimates. Specificity-orientated laboratories, notably laboratory E, fared less well for sensitivity compared with the final consensus cytology and *vice versa*. One outlier, laboratory F, seemed to achieve high sensitivity without loss of specificity with the threshold at referral (LSIL+).

In order to determine whether statistically significant variation existed in the performance parameters for the different laboratories, regression analysis was performed on the material from the years 1996–1999. During this period all laboratories are represented and the invitational year could be corrected for in the regression models. Also, the follow-up time for audit cases was uniform. Highly significant heterogeneity was found for follow-up recommendation rates, referral rates, CIN1+, CIN2+ and CIN3+ detection rates and PPV of referral (Table 3). The incidence rate of audit cases was used as a measure of screening efficiency and did not show heterogeneity. The same applies to the proportion of audit cases out of all diagnosed CIN3+ cases in the screened population.

Discussion

The purpose of the current audit study of the cytology laboratories operating in the population-based cervical cancer screening programme was to assess the performance and, in case of differences, to investigate whether screening programme effectiveness was influenced. The results enable us to produce feedback to the laboratories with the hope of harmonizing performance. Performance indicators were found to differ significantly between the laboratories. The differences in referral and CIN detection rates may partly be attributed to local variations in the prevalence of high-risk HPV (hrHPV) infection and resulting cervical lesions, but clearly laboratory-specific reporting criteria are also an issue. Recently, a report on European screening programmes found large variation in performance indicators between countries.[14] In this situation, national recommendations and education of personnel can play a large part. However, variations within countries have also been reported.[12,15–17]

Attendance rates below average were recorded for the capital area and other large urban centres in this study. It is plausible that differential use of private gynaecology services and opportunistic screening leads to lower programme attendance in urban areas. Another reason may be that women with low

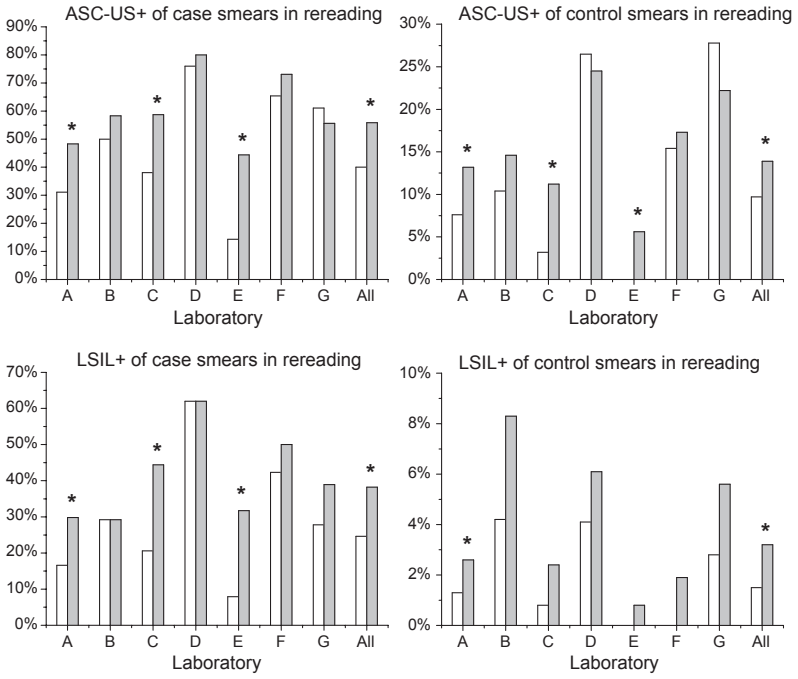

Figure 1. Rereading results expressed in cumulative proportions of ASC-US or worse and LSIL or worse, by screening laboratory and case status. □, screening laboratory result; ■, final cytology result. *Signifies a statistically significant difference between screening laboratory result and final cytology.

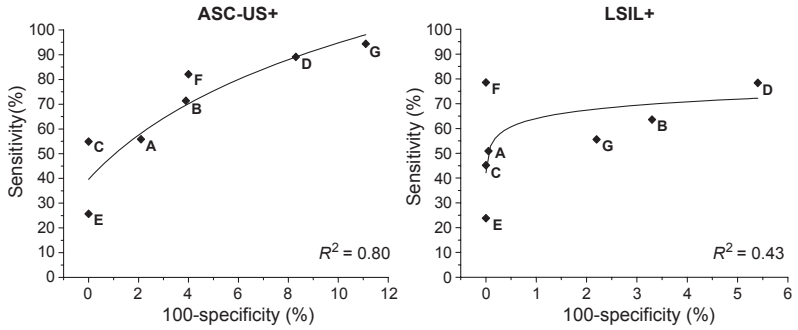

Figure 2. Test sensitivity and specificity in the rereading phase at cut-offs of A-SC-US+ (follow-up recommendation) and LSIL+ (immediate referral) by laboratory using final cytology as the g-old standard with fitted logarithmic cu-rve for trend visualization. R^2 was 0.78 for LSIL+ without laboratory F.

Table 3. Performance parameters and outcome indicators by laboratory (number of tests) over the period 1996–1999 (%)

Screening laboratory	A (113 026)	B (73 925)	C (66 859)	D (109 354)	E (70 441)	F (62 371)	G (39 535)	All (535 511)	*P*-value* (homogeneity)
Follow-up rate	7.2	2.9	4.6	8.3	2.8	10.3	4.1	6.1	<0.0001
Referral rate	1.1	0.7	1.4	1.1	0.4	0.6	0.3	0.9	<0.0001
CIN1+ rate	0.45	0.29	0.35	0.54	0.25	0.29	0.19	0.37	<0.0001
CIN2+ rate	0.27	0.19	0.17	0.43	0.19	0.20	0.17	0.25	<0.0001
CIN3+ rate	0.16	0.11	0.06	0.18	0.13	0.12	0.13	0.13	<0.0001
PPV (CIN1+)	40	43	24	48	57	47	63	42	<0.0001
PPV (CIN2+)	24	29	12	38	44	33	55	28	<0.0001
PPV (CIN3+)	14	17	4	15	29	20	43	15	<0.0001
Audit CIN3+ case rate	0.030	0.028	0.025	0.048	0.044	0.027	0.023	0.034	0.67
Audit CIN3+ /all CIN3+	16	20	31	21	25	19	15	20	0.29

**P*-value of the likelihood ratio test where laboratory was added to an age + year model.

socioeconomic and immigrant status are more likely to live in urban areas, and low socioeconomic status in turn predicts poorer health behaviour, including screening programme attendance.[18]

During the period under observation, the Papanicolaou classification system was used for reporting cytology in all laboratories. Criteria for making recommendations for intensified cytological follow-up and referral for colposcopy were also generally uniform. Hence, the variations seen in the follow-up and referral rates are due to variations in analysis and cytology reporting criteria as well as disease burden in the screened population. Overall, both follow-up and referral rates were low by international standards[11,14] even though the current estimate of the prevalence of disease-causing hrHPV infection (7.5%)[19] in the target population aged 30–64 years is average among European countries with data available (1.7–12.5%).[20] There are no data on historical hrHPV prevalence, however, which may have been lower in Finland.

The detection rates of CIN lesions of different grades reflect the cytology reporting policies of the respective laboratories. High specificity criteria lead to low detection rates of CIN1, high test specificity and PPV, especially for CIN3+ (laboratories E and G), whereas high sensitivity criteria lead to high CIN1 detection rates, and low test specificity and PPV for CIN3+ (laboratories A–C). One laboratory (F) defied this trend by displaying both high sensitivity and specificity in rereading, and high PPV and test specificity of referral combined with better than average longitudinal sensitivity (audit case rate and proportion) according to the point estimates. Interestingly, this laboratory also had the highest follow-up recommendation rate, which may account for these findings, as unclear cases tend to resolve with time and repetitive evaluations.

Of the 474 smears identified for rereading, 83% were traced and retrieved overall, but only 48% of those related to cases of invasive cancer. An explanation might be that on diagnosis of cervical cancer, the treating hospital often requests any previous smears, sometimes failing to return them to the screening laboratory. However, overall validity measures and false negative rates did not vary with the eventual histological diagnosis.[2]

The re-evaluation of audit smears was conducted in 2007 and 2008 and hence these test validity estimates do not necessarily reflect the situation during the 1990s due to personnel and policy changes since. However, the false negative rates of audit smears as defined by final cytology are valid, and the validity estimates of rereading are, at the least, an accurate reflection of the current situation. While the rereading was performed blinded as to the status of the smear, laboratories still knew that the material at hand had a high proportion of difficult abnormalities, potentially leading to higher sensitivity and lower specificity. Controls, randomization and blinded reading were used to minimize this effect, but loss of specificity can still be observed in the higher proportion of LSIL+ of controls compared with the referral rates in Table 1. The effect was variable but not unacceptable overall as the proportion of LSIL+ (warranting immediate referral for colposcopy) in the screening laboratory rereading was 1.5% (Figure 1) compared with an average referral rate of 1.0% in the programme. Laboratory E demonstrated especially good reproducibility and high specificity (no cellular abnormalities reported in the controls) but also low sensitivity in the case material compared with final cytology. Generally, the screening laboratory rereading found fewer false negative case smears than expected by final cytology.

The audit case proportions of all screening tests from the whole material (Table 2) cannot be directly compared between laboratories as the follow-up time was variable. To perform this comparison, a subset covering 1996–1999 was used to assess heterogeneity between laboratories. The most valid outcome measure for the effect of a screening programme is mortality; however, surrogate endpoints, such as the interval CIN3+ rate used in the current study, are justified in the setting of continuous quality assurance as the response time is far shorter and quality issues can be identified quickly. There was an up to two-fold variation in audit case rates between laboratories, but the variation was not statistically significant. Thus, no heterogeneity was detected in audit case rates or in the proportion of audit cases of all diagnosed CIN3+ cases in the screened population, despite significant variations in all performance indicators analysed (Table 3).

Because cervical cancer screening detects largely reversible precursors of disease, and because we do not yet have the means to distinguish between regressive and progressive lesions, determining the optimal CIN detection profile remains difficult. A study using cross-sectional data from the English screening programme estimated a desirable mean CIN score (a numerical value for the CIN detection profile) by relating it to the screen-detected cancer rate.[16] As the author cautions, the approach may be somewhat

problematic given that CIN1-3 and invasive cancer are parts of the same disease progression continuum. Hopefully some measure of the CIN detection profile, or in other words the sensitivity/specificity profile, can be shown to correlate with more robust outcome measures in the future. However, given the complexity of the screening chain from the target population and disease prevalence to quality of diagnosis and treatment, there is a risk that any recommended values may tend to be only relevant regionally.

Although reporting policies vary between laboratories, no significant variation in the rates of missed progressive lesions could be demonstrated. The results suggest that the cervical cancer screening programme functions with comparable, and good, effect across the different regions. The explanation may be that the differences in detection rates of progressive CIN lesions are smaller than those of non-progressive lesions and that in the interval between programme smears in many cases an opportunistic or clinical smear is also taken, which may rescue missed lesions before progression. Feedback concerning reporting criteria for cytology, using outcome information, is clearly indicated in order to harmonize the screening service and avoid unnecessary psychological, follow-up and treatment burdens. The findings are also important for the development of quality assurance systems and for the accreditation procedures for laboratories participating, or aspiring to participate, in organized screening.

Acknowledgments

We thank Kaija Halonen and Sirkku Vuorma from the Mass Screening Registry of the Finnish Cancer Registry for help with the preparation and handling of the register data. We also thank Kaija Knuuti, Merja Herttola-Hämäläinen and Tuomo Timonen from the Department of Pathology at the University Hospital of Helsinki and Anna Sankila from the Central Laboratory of the Finnish Cancer Organisations for providing facilities and participation in the rereading. We thank the cytotechnologists from the screening laboratories that participated in the panel sessions: Päivi Kelkka, Anna-Maija Korhonen, Päivi Häkkinen, Benita Nyqvist, Regina Lääkkö, Eila Hanelius and Berit Hällfors.

Contributors

PN and AA initiated this study and developed the study protocol. HK, JM, GG, MV, TP, AS, JA and PN were responsible for the various phases of smear reassessment. LK-T coordinated and supervised the processes of smear handling at the Mass Screening Registry of the Finnish Cancer Registry. LK-T, AA, SL and TL compiled and analysed the results. SL, AA and PN drafted the manuscript. All authors critically reviewed the manuscript and approved the final version for publication.

Funding

This study was financially supported by the European Commission and the Finnish Cancer Organisation.

Competing interests

None to declare.

Ethical approval

This study was approved by the ethical committee of obstetrics and gynaecology E8 of the Helsinki-Uusimaa hospital district (decision No 517/E8/01). The use of cytological samples originally collected for diagnostic purposes was authorized by the National Authority for Medicolegal Affairs (decision No 3670/32/300/01) and the use of confidential registries by the National Research and Development Centre for Welfare and Health (permit No 1087/900/2006).

References

1. Anttila A, Ronco G, Lynge E *et al.* Epidemiological guidelines for quality assurance in cervical cancer screening. In: *European Guidelines for Quality Assurance in Cervical Cancer Screening*, 2nd edn. Arbyn M, Anttila A, Jordan J *et al.* (eds). Luxembourg: Office for official publications of the European Communities; 2008: 11–52.
2. Lönnberg S, Anttila A, Kotaniemi-Talonen L *et al.* Low proportion of false-negative smears in the Finnish program for cervical cancer screening. *Cancer Epidemiol Biomarkers Prev* 2010;**19**:381–7.
3. Stoler MH, Schiffmann M, Atypical Squamous Cells of Undetermined Significance-Low-grade Squamous Intraepithelial Lesion Triage Study (ALTS) Group. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. *JAMA* 2001;**285**:1500–5.
4. O'Sullivan JP. Observer variation in gynaecological cytopathology. *Cytopathology* 1998;**9**:6–14.

5. Renshaw AA, Lee KR, Granter SR. Use of statistical analysis of cytologic interpretation to determine the causes of interobserver disagreement and in quality improvement. *Cancer* 1997;**81**:212–9.
6. Monto K, Nieminen P. The annual number of pap smears and the reporting system used in Finland. *Suomen Lääkärilehti* 2005;**18–19**:2093–7. Finnish.
7. 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.
8. Leinonen M, Nieminen P, Kotaniemi-Talonen L *et al.* Age-specific evaluation of primary human papillomavirus screening vs conventional cytology in a randomized setting. *J Natl Cancer Inst* 2009;**101**:1612–23.
9. Anttila A, Kotaniemi-Talonen L, Leinonen M *et al.* Rate of cervical cancer, severe intraepithelial neoplasia, and adenocarcinoma in situ in primary HPV DNA screening with cytology triage: randomised study within organised screening programme. *BMJ* 2010;**340**:c1804.
10. Anttila A, Hakama M, Kotaniemi-Talonen L, Nieminen P. Alternative technologies in cervical cancer screening: a randomised evaluation trial. *BMC Public Health* 2006; **6**:252.
11. Finnish Cancer Registry [Internet]. Cancer Statistics at http://www.cancerregistry.fi. [updated 11.8.2010].
12. Kotaniemi-Talonen L, Nieminen P, Hakama M *et al.* Significant variation in performance does not reflect the effectiveness of the cervical cancer screening programme in Finland. *Eur J Cancer* 2007;**43**:169–74.
13. Solomon D, Davey D, Kurman R *et al.* The 2001 Bethesda System: terminology for reporting results of cervical cytology. *JAMA* 2002;**287**:2114–9.
14. Ronco G, van Ballegooijen M, Becker N *et al.* Process performance of cervical screening programmes in Europe. *Eur J Cancer* 2009;**45**:2659–70.
15. Ronco G, Montanari G, Confortini M *et al.* Effect of circulation and discussion of cervical smears on agreement between laboratories. *Cytopathology* 2003;**14**:115–20.
16. Blanks RG. The development of the referral outcome diagram and an analysis of laboratory cancer detection rates in the English NHS cervical screening programme – is there an optimum level of detection of CIN 1 and CIN 2 lesions? *Cytopathology* 2008;**19**:244–53.
17. Blanks RG. Using a graph of the abnormal predictive value versus the positive predictive value for the determination of outlier laboratories in the National Health Service cervical screening programme. *Cytopathology* 2010; **21**:379–88.
18. van der Aa MA, Siesling S, Louwman MW *et al.* Geographical relationships between sociodemographic factors and incidence of cervical cancer in the Netherlands 1989–2003. *Eur J Cancer Prev* 2008;**17**:453–9.
19. Leinonen M, Kotaniemi-Talonen L, Anttila A *et al.* Prevalence of oncogenic human papillomavirus infection in an organised screening population in Finland. *Int J Cancer* 2008;**123**:1344–9.
20. De Vuyst H, Clifford G, Li N, Franceschi S. HPV infection in Europe. *Eur J Cancer* 2009;**45**:2632–9.