available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

Significant variation in performance does not reflect the effectiveness of the cervical cancer screening programme in Finland

Laura Kotaniemi-Talonen[a,*], *Pekka Nieminen*[b], *Matti Hakama*[c], *Johanna Seppänen*[a], *Jorma Ikkala*[d], *Jorma Martikainen*[e], *Jussi Tarkkanen*[f], *Terttu Toivonen*[g], *Ahti Anttila*[a]

[a]*Finnish Cancer Registry, Liisankatu 21 B, FI-00170 Helsinki, Finland*
[b]*Helsinki University Central Hospital, Department of Obstetrics and Gynaecology, Box 140, FI-00029 HUS, Helsinki, Finland*
[c]*University of Tampere, School of Public Health, Box 607, FI-33101 Tampere, Finland*
[d]*Papa-Center, Box 157, FI-28101 Pori, Finland*
[e]*Pathology Laboratory of the Finnish Cancer Society, Uusikatu 35, FI-90100 Oulu, Finland*
[f]*Haartman Institute/HUSLAB, Department of Pathology, Box 400, FI-00029 HUS, Helsinki, Finland*
[g]*Pirkanmaa Hospital District, Laboratory Center, Department of Pathology, Box 2000, FI-33521 Tampere, Finland*

ARTICLE INFO

Article history:
Received 9 June 2006
Received in revised form
16 August 2006
Accepted 22 August 2006
Available online 16 October 2006

Keywords:
Cervical cancer
Screening
Variation
Performance indicators
Laboratory
Automation-assisted

ABSTRACT

Aim: To characterise the variation in performance indicators of the Finnish cervical screening programme by screening laboratory and to assess whether the performance affects cervical cancer incidence.

Methods: Cervical cancer screening data from 1999 to 2003 from six well-established laboratories were used to analyse rates for follow-up recommendations, referrals and histologically confirmed dysplastic lesions. Laboratory-specific cervical cancer incidences for 1954–2003 were assessed using the cancer registry files.

Results: Differences in follow-up recommendations were up to 3.1-fold and 2.2-fold in referrals; differences in the rates for CIN1, CIN2 and CIN3+ were up to 4.5-, 4.7-, and 1.5-fold, respectively. Pre-screening incidence rates for cervical cancer varied 1.5-fold by laboratory, with no major differences in the incidence trends since the onset of screening.

Conclusion: The performance of a cervical screening programme differs by screening laboratory but does not materially affect the overall programme effectiveness. This leads to variation in cost-effectiveness and probably in avoidable adverse effects. In cervical cancer screening studies, the outcome should be selected as closely as possible to the true measure of programme effectiveness, prevented invasive cervical cancers and subsequent deaths.

1. Introduction

The programme for cervical cancer screening in Finland is considered to be one of the most successful in the world. Since the implementation of organised population-based screening with the Papanicolaou (Pap) smear in the mid 1960s both cervical cancer incidence and mortality rates have reduced by 80% throughout the country.[1] Most other cervical

* *Corresponding author:* Tel.: +358 9 135 33 281; fax: +358 9 673 108.
E-mail address: laura.kotaniemi-talonen@cancer.fi (L. Kotaniemi-Talonen).
0959-8049/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejca.2006.08.026

screening programmes based on conventional cytology have reached a 20–80% reduction in the incidence of invasive cervical cancer.[2] The success of the Finnish programme is based on multiple organisational and societal factors: invitational coverage of the target population is high as a nationwide population registry based on unique personal identifiers is used for invitations, individual invitation letters with pre-fixed appointment times are widely used, screening visits are free of charge for the attendees, screening samples are collected by trained registered nurses or midwives, smears are analysed mainly in laboratories specialised in cervical cytology, histological confirmation and treatment procedures are conducted by highly qualified professional personnel, and reporting the screening results to a nationwide mass screening register within the Finnish Cancer Registry is mandatory for quality assurance purposes. Moreover, the nationwide cancer registry continuously produces information on the effectiveness of the programme.

Up to the 1990s the Finnish cervical screening programme was organised and monitored by the Finnish Cancer Organisations (FCO). Since then, many of the screening laboratories formerly owned by the FCO have been privatised, and the monitoring role of the FCO has diminished. Historically, no national guidelines for cytological and histological reporting or for quality assurance in cervical cancer screening have existed in Finland, and no certification has been required from the laboratories analysing cervical screening samples. Therefore, it is possible that the variation among screening laboratories for diagnostic criteria is great. The variation in cytological and histological criteria among laboratories may lead to local differences in the rates of follow-up recommendations, referrals, histologically detected lesions and treatments, and eventually to differences in costs and screening-related disadvantages.[3]

The aim of the current study was to characterise the variation in performance indicators of the Finnish cervical cancer screening programme by screening laboratories. We also assessed whether laboratory-related differences in performance affect cervical cancer incidence rates.

2. Materials and methods

2.1. *Study material*

The Finnish cervical cancer screening programme is targeted for 30- to 60-year-old women and, as a rule, invitations are five-yearly; however, in most municipalities women with Pap class II cytology result on the screening test are re-invited within the programme the following year (risk group screening). Since 1999 this screening programme has served as a basis for a randomised evaluation of automation-assisted cytology: in six well-established screening laboratories (five laboratories of the Finnish Cancer Organisations (FCO) and the laboratory of the Helsinki municipality, later of the Helsinki and Uusimaa Hospital District), one-third of the screening samples have been analysed with automation-assisted technology. Up to 2003, based on data with 777,144 screening invitations, we have observed only minor differences in cytological and histological detection rates between the automation-assisted and conventional screening arms (data not shown). Conventional and automation-assisted screening protocols have been described in previous articles.[4,5]

For the current analysis, we included all the data on the randomised evaluation trial, including information on invitations, screening visits, cytological diagnoses and histological confirmations. The performance indicators of interest were test positivity (recommendation for cytological follow-up, referral for further examinations), histological detection (CIN1, CIN2, CIN3+), and positive predictive value (PPV) for various histological cutoffs (CIN1+, CIN2+, CIN3+). The screening laboratories were coded with the letters A–F in descending order based on the total number of screening invitations.

During the current study period, all cytological samples taken within the Finnish screening programme were classified according to the Papanicolaou classification system, where class I is normal, class II atypical but non-malignant (corresponds to reactive changes or ASC-US or LSIL in the Bethesda 2001; borderline changes in British terminology), class III suggestive of malignancy (ASC-H, LSIL or HSIL; mild to moderate dyskaryosis), class IV strongly suggestive of malignancy (HSIL; moderate to severe dyskaryosis), and class V conclusive for malignancy. All women with classes III–V cytology were referred for colposcopy. Most women with Pap class II were recommended for cytological follow-up after 6–12 months, and those with persistent abnormal smears were referred for colposcopy. The histological classification of the biopsies was based on the dysplasia classification system (the WHO classification), in which condyloma acuminatum and condyloma plana were not included in the category of mild dysplastic lesions. In the present study, the classification for cervical intraepithelial neoplasia (CIN) was used.

Incidence rates for invasive cervical cancer (ICD-10 code C53) by laboratory area were drawn from the files of the Finnish Cancer Registry. The rates and trends were compared within the main target population of the programme, among women from 30 to 64 years.

3. Statistical analysis

Laboratory-specific rates for test positivity measured by recommendations for follow-up smears and referrals for further examinations, and rates for histological detection measured by CIN1+, CIN2+ and CIN3+ were analysed with Poisson regression.[6] Laboratory C, which had the most constant outcomes during the study period, was used as the reference. In addition to crude estimates (RR_{crude}), estimates adjusted (RR_{adj}) for randomisation group (automation-assisted, conventional), age group (20–39, 40–49, 50–72), invitational group (five-yearly or risk group screening), and invitational year were reported with 95% confidence intervals (CI).

A positive predictive value was defined as the proportion of women with histologically confirmed lesions among the women referred for colposcopy. PPVs were calculated for histological cutoff CIN1+, CIN2+ and CIN3+. The association between PPV proportions and referral category was tested for all three cutoffs with an extended Mantel–Haenszel test.[7]

Cervical cancer incidence for the screening period (1964–2003) among women aged 30–64 was modelled by Poisson

regression including year of diagnosis, laboratory area and their interaction as explanatory variables.

4. Results

Following the 777,144 invitations for organised cervical screening during the years 1999–2003, altogether 548,205 (70.5%) screening visits were recorded. By screening laboratory, the attendance rate ranged from 65.7% to 78.8%, being lowest in the biggest laboratories. The proportions of abnormal smears (Pap classes II–V) varied between laboratories from 3.6% to 10.8%. This was followed by variation in the proportions of follow-up recommendations (range from 3.2% to 10.2%) and referral rates (from 0.45% to 1.12%). Histologically confirmed dysplastic lesions were detected in 0.19–0.52% of the screened women (Table 1).

The variation by screening laboratory in the rates of follow-up recommendations and referral rates was statistically highly significant. Even when controlled for randomisation group, age group, invitational group and invitational year, up to 3.1-fold differences in relative risk estimates for follow-up recommendations were obtained; for referrals the differences in RR estimates were up to 2.2-fold (Table 2). Variation between laboratories in the rates of histologically confirmed lesions was also significant, but it diminished towards high-grade lesions: at the level of CIN1 the differences between adjusted RR estimates were up to 4.5-fold, at the level of CIN2 up to 4.7-fold, and at the level of CIN3+ 1.5-fold (Table 3).

PPV estimates for the histological cutoff of CIN1+ were between 32.8% and 52.0%, for CIN2+ between 23.2% and 36.9%, and for CIN3+ between 12.8% and 23.5%. Variation in the estimates by laboratory was statistically significant (Table 4).

In 1954–1958, well before organised screening started in Finland, the five-yearly average incidence rates for invasive cervical cancer varied within the target population by laboratory area from 44.4 to 30.2 per 100,000 woman-years, i.e. there was maximally a 1.5-fold variation in the observed rates (Fig. 1). Since the onset of screening, a proportionally uniform decrease in cervical cancer incidence has occurred in all laboratory areas: in the Poisson regression the main effects of laboratory area ($p < 0.0001$) and calendar year ($p < 0.0001$) were highly significant, whereas the interaction between these variables showed no statistical significance ($p = 0.20$).

5. Discussion

Overall variation in performance indicators was highly significant between the well-established screening laboratories within the Finnish screening programme. The differences in follow-up recommendations were up to 3.1-fold and 2.2-fold in referrals; differences in the rates for CIN1 and CIN2 were up to 4.5- and 4.7-fold, respectively. At the level of CIN3+, the variation by laboratory was the least, 1.5-fold. As historical differences in cervical cancer incidence between laboratory areas were also about 1.5-fold, observed variation in the CIN3+ rates is likely to be mostly due to differing background risks.

Previous trend analyses have suggested at least moderate variation in performance but only minor differences in the overall programme effectiveness between laboratories in Fin-

Table 1 – Description of the study data by screening laboratory

	Laboratory													
	A		B		C		D		E		F[a]		All	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Invitations	227,266	100.0	200,103	100.0	132,349	100.0	96,822	100.0	73,166	100.0	47,438	100.0	777,144	100.0
Screens	156,699	68.9	131,386	65.7	93,778	70.9	76,385	78.9	53,912	73.7	36,045	76.0	548,205	70.5
Screening test result	156,698	100.0	131,386	100.0	93,778	100.0	76,385	100.0	53,912	100.0	36,045	100.0	548,204	100.0
Pap class I	142,510	90.9	118,979	90.6	90,296	96.3	68,100	89.2	49,495	91.8	34,763	96.4	504,143	92.0
Pap class II	12,768	8.1	10,888	8.3	3017	3.2	7831	10.3	4007	7.4	1122	3.1	39,633	7.2
Pap class III	1270	0.81	1239	0.94	393	0.42	402	0.53	375	0.70	152	0.42	3831	0.70
Pap class IV	145	0.09	157	0.12	69	0.07	38	0.05	35	0.06	8	0.02	452	0.08
Pap class V	5	0.00	10	0.01	0	–	3	0.00	0	–	0	–	18	0.00
Inadequate	0	–	113	0.09[b]	3	0.00	11	0.01	0	–	0	–	127	0.02
Cytological follow-up	12, 563	8.0	10,835	8.2	2965	3.2	7823	10.2	4033	7.5	1381	3.8	39,600	7.2
Referral	1764	1.13	1487	1.13	538	0.57	452	0.59	531	0.98	166	0.46	4938	0.90
Colposcopy	1716	1.10	1471	1.12	518	0.55	447	0.59	500	0.93	162	0.45	4814	0.88
Other	48	0.03	16	0.01	20	0.02	5	0.01	31	0.06	4	0.01	124	0.02
Histologically detected lesion	807	0.52	704	0.54	241	0.26	235	0.31	174	0.32	70	0.19	2,231	0.41
CN1	323	0.21	177	0.13	67	0.07	68	0.09	51	0.09	15	0.04	701	0.13
CIN2	216	0.14	331	0.25	75	0.08	79	0.10	55	0.10	16	0.04	772	0.14
CIN3	250	0.16	162	0.12	92	0.10	82	0.11	62	0.12	36	0.10	684	0.12
Invasive cancer	18	0.01	34	0.03	7	0.01	6	0.01	6	0.01	3	0.01	74	0.01

a Participated in 1999–2002.

b All inadequate samples were collected in 2003, when they represented 0.4% of all the smears.

Table 2 – Crude and adjusted relative risks for the two levels of test positivity (cytological follow-up and referral) by screening laboratory, in comparison to laboratory C

Laboratory	Cytological follow-up			Referral		
	RR_{crude}	RR_{adj}	CI 95%	RR_{crude}	RR_{adj}	CI 95%
A	2.54	2.44	2.34–2.54	1.96	1.79	1.63–1.98
B	2.61	2.61	2.51–2.72	1.97	1.77	1.60–1.95
C	1.00	1.00	–	1.00	1.00	–
D	3.24	3.10	2.97–3.23	1.03	0.96	0.84–1.09
E	2.37	2.23	2.13–2.34	1.72	1.57	1.39–1.77
F[a]	1.21	1.16	1.09–1.24	0.80	0.79	0.66–0.94

Adjusted figures are controlled for randomisation group, age group, invitational group and invitational year.
For both levels of test positivity the p-value for RR_{adj} was <0.001.
a Participated in 1999–2002.

Table 3 – Crude and adjusted relative risks for CIN1, CIN2 and CIN3+ by screening laboratory, in comparison to laboratory C

Laboratory	CIN1			CIN2			CIN3+		
	RR_{crude}	RR_{adj}	CI 95%	RR_{crude}	RR_{adj}	CI 95%	RR_{crude}	RR_{adj}	CI 95%
A	2.89	2.65	2.04–3.45	1.72	1.58	1.21–2.05	1.62	1.49	1.18–1.87
B	1.89	1.69	1.27–2.24	3.15	2.60	2.02–3.35	1.41	1.18	0.92–1.50
C	1.00	1.00	–	1.00	1.00	–	1.00	1.00	–
D	1.25	1.16	0.83–1.63	1.29	1.24	0.90–1.70	1.09	1.05	0.79–1.40
E	1.32	1.22	0.85–1.76	1.28	1.19	0.84–1.69	1.20	1.11	0.81–1.51
F[a]	0.58	0.59	0.34–1.03	0.56	0.55	0.32–0.94	1.03	0.98	0.67–1.42

Adjusted figures are controlled for randomisation group, age group, invitational group and invitational year.
For CIN1 and CIN2 the p-value for RR_{adj} was <0.001, for CIN3+ $p = 0.002$.
a Participated in 1999–2002.

Table 4 – Positive predictive values for histologically detected lesions by screening laboratories, calculated for histological categories CIN1+, CIN2+ and CIN3+

Laboratory	Referral	CIN1+		CIN2+		CIN3+	
		n	%	n	%	n	%
A	1764	807	45.7	484	27.4	268	15.2
B	1487	704	47.3	527	35.4	196	13.2
C	538	241	44.8	174	32.3	99	18.4
D	452	235	52.0	167	36.9	88	19.5
E	531	174	32.8	123	23.2	68	12.8
F[a]	166	70	42.2	55	33.1	39	23.5

For all histological cutoffs the p-value for RR_{adj} was <0.001.
a Participated in 1999–2002.

land.[8,9] Even though the variation in performance was greater than reported earlier, the current results on effectiveness are in line with the previous analysis. However, as the laboratories included in this study have a long history of mutual collaboration and cooperation, the overall variation in the performance indicators between all laboratories analysing screening smears in Finland is probably even greater than now described.[1,3,10] Analogously with our study, wide variation in screening results, especially in low-grade cytology, has been reported between screening laboratories in England and between local screening programmes in Italy.[11,12] In a study comparing performance indicators between 18 European countries, CIN3 detection rates were observed to vary 8-fold, which was suggested to be mainly due to differences in diagnostic and registration criteria and only partly due to differences in background risks.[13] However, in that study the variation due to a background risk could not be clearly distinguished from the variation related to differing criteria. In general, monitoring the extent of variation in screening performance is useful, because it is likely to predict the effect of the programme.

Some of the cytological and histological variations can be explained by differences in diagnostic criteria: those screening laboratories that favour high specificity tend to be conservative with respect to the grading of cytological abnormalities, and laboratories putting focus on sensitivity have less stringent cytological criteria as they are not willing to miss any lesion. This kind of variation in cytological

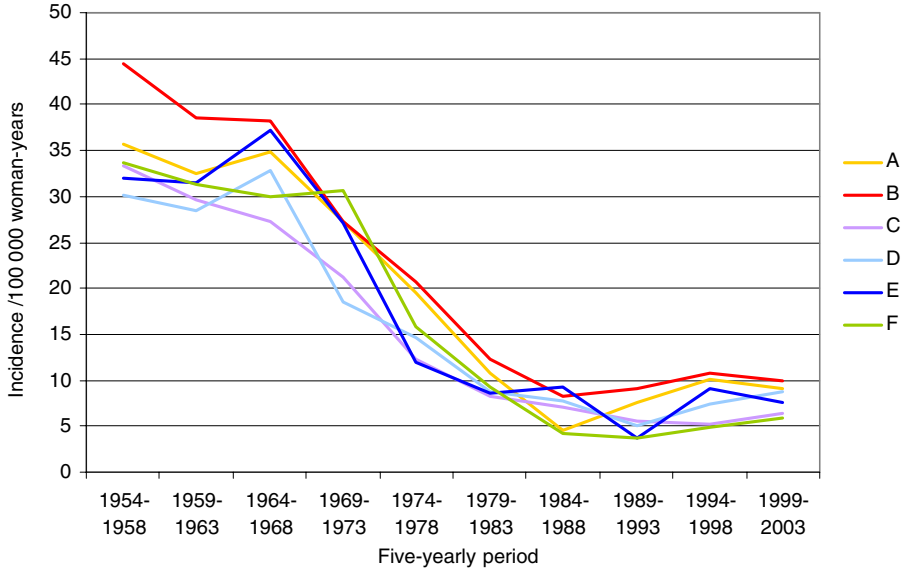

Fig. 1 – Cervical cancer incidence among 30- to 64-year-old women in 1954–2003 by laboratory area.

criteria results could lead to large laboratory-specific differences in CIN1 and CIN2 profiles, as is also seen in the current data. High numbers of histologically confirmed low- to moderate-grade lesions may represent overdiagnosis and overtreatment, whereas very low numbers may result from underdiagnosis, possibly causing delays in treatments. It is not evident whether a new screening method will eliminate these intimate problems in cervical cancer screening: in the current study the laboratory-related variation in performance was similar for both screening tests (data not shown).

Compared to many other countries, the results of the Finnish screening programme are excellent, i.e. the programme sensitivity for progressive cervical cancer precursors is high.[2,14] Average referral rates in Finland are markedly low – the specificity of the programme is high. Despite the 3-fold differences in the rates of follow-up smear recommendations, 2-fold differences in referral rates, and more than 4-fold differences in the rates of lesions requiring treatment, there is no evidence that any laboratory is inferior to the others in terms of cervical cancer incidence. Nevertheless, by strengthening the collaboration between individual screening laboratories and by committing the laboratories to a nationwide feedback or auditing programmes, the diagnostic criteria for cytology and histology could be further developed and, thus, the cost-effectiveness of the screening programme improved.

In the current study, we estimated the effectiveness of the screening programme by screening laboratory in terms of histologically confirmed CIN3+ lesions and by comparing the CIN3+ rates to laboratory-specific cervical cancer incidence trends. This is not an entirely optimal approach: historical differences in the screening coverage and compliance and in the uptake of opportunistic Pap smears, and potential differential fluctuations in the background risk could partially affect these trends. Therefore, we cannot rule out small differences in effectiveness between laboratories. Interval cancer incidence is a more reliable outcome indicator than detection of any precancerous lesion. Mortality due to cervical cancer could also be used as outcome indicator, if taken into account that not only the variation in diagnostic criteria for precancers, but also treatment procedures may affect these cause-specific mortality rates. In the literature on cervical screening, even the histologically confirmed CIN2+, which is more susceptible to variation due to differences in diagnostics and background risks than CIN3+, is widely accepted as a reasonable outcome indicator – which should be discouraged.

In conclusion, the performance of the cervical screening programme differs by screening laboratory, but the overall programme effectiveness is not materially affected. This leads to a variation in cost-effectiveness and probably in the avoidable adverse effects of the cervical cancer screening programme. For this, in cervical cancer screening studies the outcome should be selected as closely as possible to the true measure of programme effectiveness, which is the number of prevented invasive cervical cancers and subsequent deaths.

Conflict of interest statement

Jorma Ikkala, the medical director of the Papa-Center in Pori, has had 50% ownership in the Papa-Center since 2002. Four other authors have or have had leading positions, although not ownerships, in the laboratories included in the study: Pekka Nieminen is the medical director of the Policlinic of the Finnish Cancer Organisations in Helsinki, Jorma Martikainen is the medical director of the Pathology Laboratory of the Finnish Cancer Society in Oulu, Jussi Tarkkanen is the chief medical officer of the mass screening program in HUSLAB, and during the study period doctor Terttu Toivonen was responsible for cytological pathology in the Cancer Policlinic and the Laboratory of the Finnish Cancer Organisations in Kotka. Laura Kotaniemi-Talonen, Matti Hakama, Johanna Seppänen and Ahti Anttila do not have financial relations to any of these laboratories. None of the authors have any connection to the former NSI Incorporation, the manufacturer of the automation-assisted screening system used in the study.

Role of the funding source

The randomised study on automation-assisted versus conventional screening has been supported with research grants from the European Commission program Europe Against Cancer through the European Cervical Cancer Screening Network, the Academy of Finland and the Finnish Cancer Organisations.

REFERENCES

1. Finnish Cancer Registry, Cancer Statistics, 2006. Available at www.cancerregistry.fi.
2. The Working group of IARC. *IARC handbooks of cancer prevention: cervical cancer*. Lyon: IARC Press; 2005.
3. Nieminen P, Hakama M, Tarkkanen J, Anttila A. Effect of type of screening laboratory on population-based occurrence of cervical lesions in Finland. *Int J Cancer* 2002;**99**(5):732–6.
4. Nieminen P, Hakama M, Viikki M, Tarkkanen J, Anttila A. Prospective and randomised public-health trial on neural network-assisted screening for cervical cancer in Finland: results of the first year. *Int J Cancer* 2003;**103**(3):422–6.
5. Nieminen P, Kotaniemi L, Hakama M, et al. A randomised public-health trial on automation-assisted screening for cervical cancer in Finland: performance with 470,000 invitations. *Int J Cancer* 2005;**115**(2):307–11.
6. McCulloch CE, Searle SR. *Generalized, linear, and mixed models*. New York: Wiley; 2000.
7. Mantel N, Byar DP. Marginal homogeneity, symmetry and indipence. *Commun Stat Ser A* 1978(7):953–76.
8. Anttila A, Pukkala E, Soderman B, Kallio M, Nieminen P, Hakama M. Effect of organised screening on cervical cancer incidence and mortality in Finland, 1963–1995: recent increase in cervical cancer incidence. *Int J Cancer* 1999;**83**(1):59–65.
9. Tarkkanen J, Geagea A, Nieminen P, Anttila A. Quality improvement project in cervical cancer screening: practical measures for monitoring laboratory performance. *Acta Obstet Gynecol Scand* 2003;**82**(1):82–8.
10. Anttila A, Nieminen P. Cervical cancer screening programme in Finland. *Eur J Cancer* 2000;**36**(17):2209–14.
11. Statistical Bulletin. NHS Health and Social Care Information Centre, Health Services Community Statistics; 2005.
12. Ronco G, Giubilato P, Naldoni C, et al. Activity level and process indicators of organised programmes for cervical cancer screening in Italy. *Epidemiol Prev* 2006;**30**(1):25–38.
13. Anttila A, Ronco G, Clifford G, et al. Cervical cancer screening programmes and policies in 18 European countries. *Br J Cancer* 2004;**91**(5):935–41.
14. Viikki M, Pukkala E, Hakama M. Risk of cervical cancer after a negative Pap smear. *J Med Screen* 1999;**6**(2):103–7.