Chapter 4
Efficacy of screening

Methodology and analytical issues in assessment of efficacy

The core concept of screening is that detection of early disease offers the opportunity to change its prognosis. Earlier diagnosis may improve prospects for survival because early intervention permits treatment at a more tractable stage (Morrison, 1992). However, as experience with screening has accumulated and understanding of cancer biology has evolved, it has become apparent that there is much heterogeneity among cancers at particular sites, and that this heterogeneity can influence the impact of screening. Models of screening should take account of this heterogeneity.

General definitions

A simplified model of screening is presented in Figure 47. Several definitions are needed to understand this model. First, the model assumes that there is a period in which there is no detectable disease, but early malignant changes may have taken place and a clone of cells is dividing and de-differentiating until it attains a size that can be detected by screening. The point at which a lesion can be found by screening is the beginning of the sojourn time (Zelen & Feinleib, 1969) or 'detectable preclinical phase' (DPCP) (Cole & Morrison, 1980). For cervical cancer screening, lesions during the DPCP are mainly preinvasive, but also include some early invasive and microinvasive lesions. 'Lead time' refers to the period between the moment a lesion is found by screening and the time of diagnosis of the invasive cancer that would have developed (Morrison, 1992). Sojourn time is a combined function of the lesions and of the screening test. Lead time will in addition usually be affected by the frequency of screening, depending on the distribution of the sojourn time. Both sojourn time and lead time will vary widely in a population. Neither is directly observable for an individual, unless a screening test is repeated at frequent intervals, the results of a positive screening test are ignored and the woman is observed until she becomes symptomatic. Such a situation is clearly not tenable. However, in a population that has undergone screening, the distribution of lead time and sojourn time can be estimated (Walter & Day, 1983).

Sensitivity of the screening test for early detection of invasive lesions

In addition to sojourn time and lead time, two parameters traditionally of importance in screening are sensitivity and specificity. For a condition which either exists or does not, such as Tay–Sachs disease, these two parameters are defined in terms of a 2 x 2 table:

Result of screening test	'True' disease state	
	Positive	Negative
Positive	*a*	*b*
Negative	*c*	*d*

Sensitivity = $a/(a + c)$, specificity = $d/(b + d)$

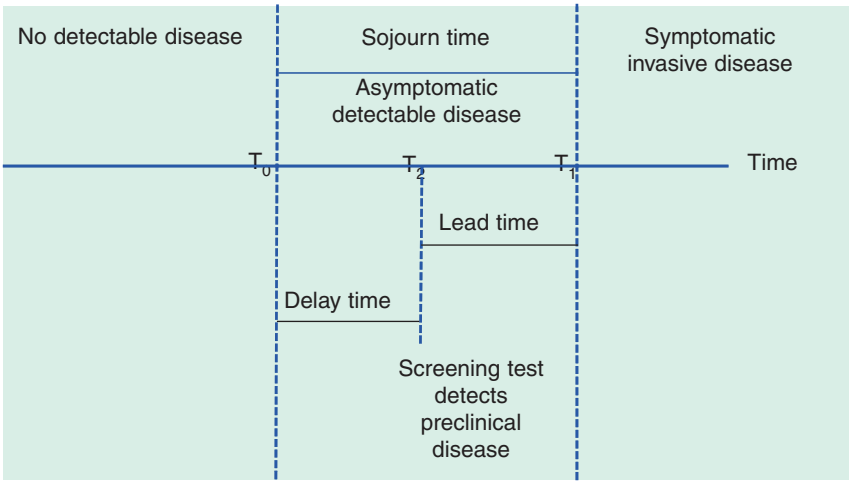

Figure 47 Scheme of progression of a chronic disease, with the intervention of an early-detection screening test

The situation is more complex for screening for cervical cancer, because it is a progressive condition. At the time at which screening is performed, there is no 'gold standard' diagnostic test for the disease: the condition being screened for is a future invasive disease. The 'true' disease state being sought at the time of screening is a lesion that will progress into an invasive cancer. This state can be determined for an individual only by following her forward in time. Since, however, a positive result at screening should lead to an intervention to prevent the development of a clinical cancer, much of the information required for direct estimation of sensitivity and specificity will be missing, so that there is no direct measure of the quantity *a* in the table. If one follows forward in time a group of individuals who showed no lesion on the screening test, some will develop invasive disease. There is a tradition to estimate the unknown quantity *c* by follow-up after screening: the women who develop invasive disease after a negative screening result thus constitute the cell entry *c* in the table above.

The length of time after screening that is used to define this group of 'screen-negative' and 'disease-positive' individuals is variable, and is a somewhat arbitrary interval. The interval between screening rounds is a natural choice if an organized rescreening programme is being evaluated.

In actual screening programmes, a variety of screening intervals are used and a relatively long interval may be used to define sensitivity. This has the advantage that it is less subject to statistical variation due to small numbers and less dependent on the exact date of diagnosis, although more affected by bias due to new cancers. Clearly, the longer the interval used to define sensitivity, the lower will be the resulting estimate (as follows from the discussions below and Figure 48).

To estimate sensitivity, one must then identify the individuals, or indirectly estimate their number, who constitute the cell entry *a*. The 'true' disease state is agreed to be invasive cancer appearing after a positive screening episode. Thus one needs to estimate the number of lesions, detected at screening and treated, which in the absence of screening would have progressed to an invasive cancer. This group forms the screen-positive, disease-positive group. The quantity $a + c$ is the number of cancers that would have presented as frankly invasive cases in the screened group if no screening had taken place. If one has a directly comparable unscreened population, as in a randomized trial, the quantity $a + c$ is observable. In the absence of a comparison group strictly defined by randomization, other approaches are needed, but for any general population sample, estimates based on age-adjusted cancer incidence data from a comparable population or a time when screening was not practised should provide a good approximation, if used judiciously. The quantity *a* is then obtained by subtraction, and the episode sensitivity estimate is given as before (Day, 1985):

$$a/(a + c).$$

This approach to the estimation of sensitivity, called the 'incidence method', can be expressed graphically as in Figure 48 and can be used to estimate sensitivity by means of the proportionate incidence of interval cancers (see below).

Since screening for cervical cancer is aimed at detection of preclinical lesions, most of which are preinvasive, the estimate of $a/(a + c)$ (for invasive disease) is to be interpreted as the proportion of screen-detected preinvasive lesions that would have progressed to invasion among all lesions in the DPCP that would have progressed to invasion. This is equal to one minus the proportion of observed invasive cancer post-screening to the expected invasive cancer rate in the absence of screening.

An additional issue is the categorization of microinvasive (stage IA1) disease. This is almost always screen-detected and can be effectively treated with minimal morbidity. While this category is usually included among the invasive cancers, for the evaluation of screening, it should be considered as a success and is best included with the carcinoma *in situ* or CIN 3 cases as a less than fully invasive lesion.

It should be realized that diagnostic issues related to disease verification as discussed in Chapter 2 are fundamental to this discussion. The frequency of disease that is identified in a screening

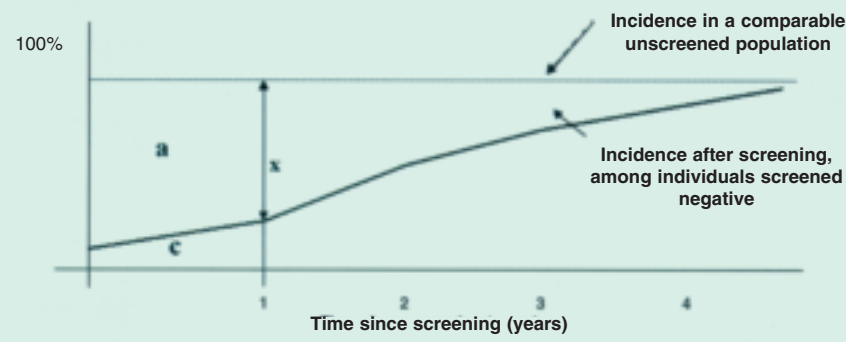

Figure 48 Sensitivity defined in terms of one-year proportionate incidence: incidence of interval cancers as a proportion of the incidence in a comparable unscreened population

programme is also dependent on the diagnostic methods that are used to augment colposcopy and biopsy (e.g., LEEP investigation) and on how aggressively one looks for disease.

Positive predictive value and specificity

A similar approach can be taken to the definition of specificity and positive predictive value, as, if one has estimates of the values of *a* and *c*, and *a* + *b* and *c* + *d* are known from the results of screening, then clearly one has estimates of *b* and *d*. For positive predictive value, for example, it is of special interest to estimate the proportion of lesions detected at screening that would have progressed to clinical cancers (i.e., $a/(a + b)$) before the next round in a periodic screening programme. For specificity, or rather its complement, one might be interested in the proportion of individuals who had a positive screening result among those who would not have developed a clinical cancer in the interval between screening tests (i.e., $b/(b + d)$). The positive tests that were confirmed negative should be included for the test specificity. The test validity indicators correspond to each other like those of episode validity. In particular, it is deficient to report (only) episode specificity and only test sensitivity.

Attention should also be paid to the definition of the screening test. Screening for cervix cancer is essentially a multiple-step process, with the initial screening test leading, if positive, to more detailed investigations, culminating in a biopsy for a definitive diagnosis. The definitions of sensitivity and specificity discussed in this section refer to the complete screening episode, the final assessment of positivity or negativity being based on the results of the screening test and all further assessment. It is a common experience that women with a positive test but classified as negative (i.e., disease-free) on further assessment are at higher risk of subsequent disease than the general population. The implication is that if only the screening test is considered, it will be more sensitive than the overall screening episode, although of course with less specificity. The sensitivity of screening tests could be estimated in analogous fashion to the sensitivity of the complete screening episode, but in practice such estimation is rarely attempted.

Relative sensitivity of different screening tests

The preceding paragraph considered sensitivity as the capacity of a screening episode and test to detect future invasive disease. This measure might be termed absolute sensitivity. Often, however, one requires a more rapid and direct method of comparing the sensitivity of two screening tests, probably based on cross-sectional rather than follow-up data. In this situation, it is useful to define the relative sensitivity of the two tests in terms of a surrogate measure of future invasive disease. In the context of cervical screening, this surrogate is usually taken as histological diagnosis of the screen-detected lesions.

It should be noted that not all lesions diagnosed histologically as malignant would have progressed to invasive cancers. It is known that many preinvasive lesions, including CIN 3 and carcinoma *in situ,* will regress. Furthermore, colposcopically directed biopsies are known to miss significant lesions. It is therefore clear that relative sensitivity and absolute sensitivity are measures of different quantities. Relative sensitivity would normally be larger than absolute sensitivity, due to overdiagnosis at histology and the length bias of cytologically detected lesions. Hence, if relative and absolute sensitivities differ, due to inherent bias in the relative sensitivity, the absolute sensitivity will give the more correct estimate.

To determine relative sensitivity, a sample of women would undergo both screening tests (e.g., split-sample studies; see Chapter 2). In some studies, all women then undergo colposcopy as well, and the histological diagnosis is obtained on the entire sample. The results would then be summarized as in the table below.

The sensitivity of test 1 relative to histology is then $w_1/(w_1 + x_1)$, and of test 2 is $w_2/(w_2 + x_2)$. The specificity of test 1 relative to histology is $z_1/(y_1 + z_1)$, and of test 2 is $z_2/(y_2 + z_2)$.

Since the same women undergo both tests, $w_1 + x_1 = w_2 + x_2$ (the number positive by histology), so the comparison of the relative sensitivities of the two tests is given simply by the ratio w_1/w_2.

Thus to compare the relative sensitivities of the two screening tests, it is only necessary to obtain histological diagnosis on all women positive on at least one of the two tests.

Many studies of comparative sensitivity and specificity in fact only have

		Histological diagnosis	
		Positive	Negative
Screening test 1	Positive	w_1	y_1
	Negative	x_1	z_1
Screening test 2	Positive	w_2	y_2
	Negative	x_2	z_2

Detection rate of Test 1 is $w_1/(w_1 + x_1 + y_1 + z_1)$
Test positivity rate of Test 1 is $(w_1 + y_1)/(w_1 + x_1 + y_1 + z_1)$

histological diagnosis available on those positive on one of the screening tests to be compared. These studies give an unbiased estimate of the ratio of their respective relative sensitivities, provided all those positive on one of the two screening tests have a histological diagnosis available. However, they give a biased overestimate of the sensitivity relative to histology, as they exclude from the denominator women positive by histology but negative on the screening tests under consideration.

Estimates of specificity can also be obtained from such studies, since there will be women negative on one of the screening tests among the women who are positive on at least one of the tests. This estimate will be severely biased (underestimated), since a major component of a correct specificity estimate will be women negative on both screening tests and by histology.

This bias in sensitivity and specificity is known as the verification bias (Franco, 2000, 2003). Some studies have chosen a sub-sample of women negative on the screening test to undergo a histological diagnosis, in order to attempt to correct for verification bias. This attempt may be more or less successful, depending on the size of the sample, the comparability of the colposcopy and perhaps other factors in reducing verification bias. An example of correction for verification bias is given in Table 58 (taken from Ratnam *et al.,* 2000). The effects are clearly large. [The ratio of the uncorrected sensitivities differs slightly from the ratio of the corrected sensitivities since not all women positive on one or both of the two tests were included in the uncorrected estimates, although included in the corrected estimates.]

One must also be alert to potential bias in the interpretation of the net efficacy of screening when two tests are used in series or in parallel in the same women (Macaskill *et al*., 2002). A nominal increase in relative sensitivity always occurs by chance whenever an adjunct test (e.g., HPV DNA testing) is used in parallel with a conventional one (e.g., cytology), even if the new test gives totally random results with respect to the disease being evaluated. This increase in sensitivity can be misleading, even if deemed significant by a statistical test. Combined testing prevents a loss in specificity but may offer no real sensitivity gain in certain conditions (Franco & Ferenczy, 1999). An empirically valuable adjunct test, such as the HPV assay, should complement cytological testing so that the net combined sensitivity and specificity will be truly superior to those of cytology alone. In practice, we can compute the "expected null values" for sensitivity and specificity using the following formulae (Franco, 2000):

$$S_{exp} = S_{cyt} + P\,(1 - S_{cyt})$$

for the expected null sensitivity, and

$$W_{exp} = W_{cyt} - P\,(W_{cyt})$$

for the expected null specificity,

where S = sensitivity, W = specificity, S_{cyt} and W_{cyt} represent, respectively, the sensitivity and specificity for the original cytological test, S_{exp} and W_{exp} denote the adjusted (for the addition of the new test) sensitivity and specificity, and P is the expected positivity rate for HPV testing or any other test used as an adjunct to cytology in the same population. These expected values should then be used in a comparison with the equivalent indices from the combined testing approach.

An analogous issue to the latter bias that stems from the application of adjunctive testing may also appear in randomized controlled trials as an asymmetry problem. Figure 49 illustrates this problem using a generic example with two screening tests, one (test A) that serves as the paradigm (e.g., cytology) and the other (test B) that serves as the adjunctive, experimental technology (e.g., HPV DNA testing), whose benefit is to be evaluated. If a trial is designed without long-term follow-up, it will probably misinterpret the difference between arms in detection rates of preinvasive lesions as being indicative of the putative efficacy attributable to the intervention. One needs to exercise caution and avoid inadvertently claiming that the combination of cytological and HPV DNA testing in such trials is superior simply because it detected more prevalent (or short-term incident) high-grade lesions than the cytology-only arm, as a measure of the greater sensitivity of the combined testing approach. The asymmetry bias virtually guarantees that this would happen even if the HPV test performed randomly with respect to

Table 58. The effect of verification bias on estimate of sensitivity and specificity
Newfoundland Study: screening performance after correction for verification bias (HSIL or worse)

Screening	Definition of	Uncorrected		Corrected	
		Sensitivity	Specificity	Sensitivity	Specificity
Cytology	LSIL or worse	38.2	80.5	26.8	96.2
HPV	Positive	85.3	58.0	68.1	90.6
Combination	HPV+ or ≥ LSIL	97.1	51.3	76.3	89.3

From Ratnam *et al.* (2000)

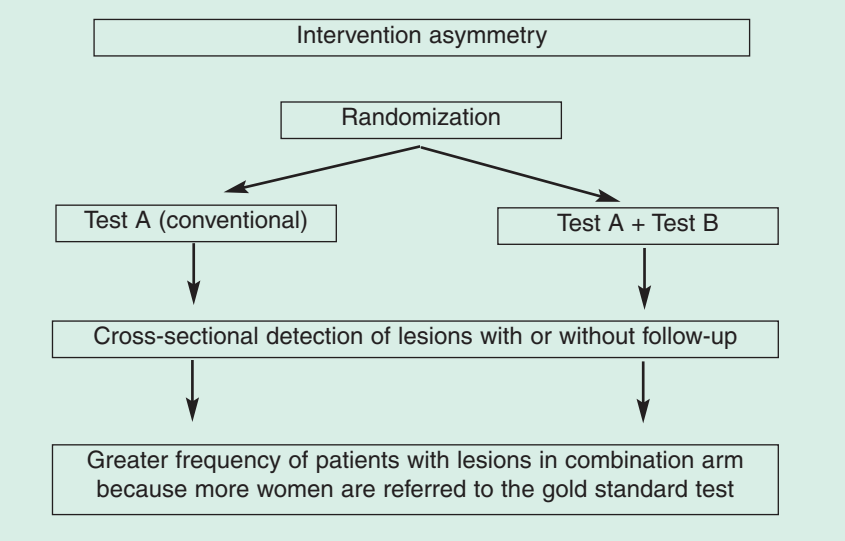

Figure 49 Intervention asymmetry bias in randomized controlled trials of an adjunctive screening test with short-term follow-up

lesions and cytology. This is because there will be more women selected for triage, which will increase the probability of detecting incipient lesions that would never be found were it not for the contribution of the adjunctive test, however inadequate it may be (Franco, 2004).

A more complex view of cancer

The model shown in Figure 47 describes the operational process of screening, incorporating no information on the biology of the carcinogenic process. Current knowledge of the neoplastic process allows us to distinguish a number of steps, which may begin with mutation at specific genetic loci and other cellular events and continue until cells divide and disseminate throughout the organism. Cancer development is a long process, and not all the steps are necessarily irreversible. In the future, screening modalities may be developed to target these early molecular changes. In that case, more complex models of the screening process will be required.

If all women do not undergo each test, $w_1 + x_1$ is no longer equal to $w_2 + x_2$ (as in the direct-to-vial studies mentioned in Chapter 2). While cross-sectional analysis mimics the relationship between relative sensitivities, follow-up may allow estimation of the absolute sensitivities, depending on the validity of controls.

Methods used to evaluate the efficacy of screening

Reductions in mortality and/or incidence of invasive disease are fundamental measures of the efficacy of screening. A reduction in the incidence of invasive disease as a consequence of the treatment of disease precursors is a predictor of a reduction in mortality from cervical cancer.

Because screening for cervical cancer results in the detection and treatment of precursors, reduction in the incidence of the disease is an appropriate outcome measure. Reduction in mortality was used in some early studies to evaluate cervix cancer screening. It is accepted that case survival is not an appropriate outcome, because of lead time, length bias, selection bias and overdiagnosis bias.

Randomized controlled trials are the most valid method to assess reductions in mortality or incidence consequent on screening (Prorok *et al.*, 1984), but until recently have not been employed in the evaluation of screening for cervical cancer.

Use of observational studies in evaluation of screening

Observational studies can be used to evaluate the efficacy of screening, provided that the programme was introduced sufficiently long before the study that an effect can be expected to have occurred. The major bias that potentially affects observational studies in the evaluation of screening is selection bias, as the health-conscious may select themselves for screening and are likely to have different underlying (lower) risks of developing cancer of the cervix from those who refuse screening. Thus any comparison that essentially compares the incidence of cancer of the cervix in those who accept invitations to attend for screening with those who decline such invitations is potentially affected by selection bias.

In the cohort study design, the incidence of cancer of the cervix in an individually identified and followed screened group (the cohort) is compared with the incidence in a control population, often derived from the general population, sometimes using data from before screening started, or from another study population in which screening has not been used. The incidence in the controls should be adjusted for the incidence in the refusers (Cuzick *et al.*, 1997) for a valid estimate of efficacy. An estimate of efficiency is obtained by comparing the incidence in those offered screening (attenders and refusers) with that among controls. A historical comparison could be biased if changes in risk factors in the population are affecting the incidence of the disease. If mortality

were to be used as the end-point, it must be recognized that those recruited into a screening programme are initially free of the disease of interest, so that it is not appropriate to apply population mortality rates for the disease to the person-years experience of the study cohort. Rather, as is required in estimating the sample size required for a controlled trial of screening, it is necessary first to determine the expected incidence of the cases of interest, then apply to that expectation the expected case-fatality rate from the disease to derive the expectation for the deaths (Moss *et al.*, 1987). In practice, it is difficult in a cohort study of screening to adjust for selection, so the results have to be interpreted with caution.

A case–control study of screening is another approach that can be used to evaluate screening. Case–control studies should be designed to mimic randomized controlled trials as far as possible, especially in terms of the cases used (ideally, in this instance, cases of invasive cancer of the cervix). They depend on comparing the screening histories of the cases with the histories of comparable controls drawn from the population from which the cases arose. Individuals with early-stage disease if sampled would be eligible as a control, providing the date of diagnosis was not earlier than that of the case, as diagnosis of disease truncates the screening history. However, a bias would arise if advanced disease were compared only with early-stage disease, as the latter is likely to be screen-detected, though this is just a function of the screening process, not its efficacy (Weiss, 1983). Cases have to reflect the end-points used to evaluate screening, i.e., those that would be expected to be reduced by screening.

Selection bias may be difficult to adjust for in the analysis, though this should be attempted if relevant data on risk factors for the disease (confounders) are available. Such bias may not be a problem, however, if it can be demonstrated that the incidence of cancer in those who declined the invitation to the screening programme is similar to that expected in an unscreened population. This is only seldom true, however (see Chapter 3).

Even if data are available on risk factors for disease, control for them may not result in avoiding the effect of selection bias. Experience in breast cancer screening studies in Sweden and the United Kingdom, where case–control studies were performed within trials, shows that although breast cancer incidence among those who refuse invitations for screening is similar to that of controls, their breast cancer mortality experience is worse than that of controls. This means that the estimate of the effect of screening in such case–control studies will be greater than could be expected in the total population (Miller *et al.*, 1990). This differential could also arise because the case–control analysis directly measures the effect in those screened.

There are other problems with case–control studies of screening, notably two relating to the exposure measure (Weiss, 1994, 1998). One is the issue of excluding tests that are done because the disease is present. These are so-called diagnostic tests, often performed in women with symptoms, or suspected to be at high risk. The second is that of bias due to counting only negative tests as exposure to screening. This will bias the result by omitting positive tests, although these are the only tests that can possibly influence risk of disease.

Assessing the efficacy of a new screening test

When a new screening test becomes available as a supplement or even a replacement for a screening test that is known to be effective in reducing incidence of invasive cancer, a comparative assessment can be made in terms of absolute sensitivity. Since absolute sensitivity, as defined earlier, quantifies the reduction in the incidence of invasive disease following screening, a comparison between two tests of their absolute sensitivity should be predictive of the comparative efficacy of the two tests. Often, however, it will initially be easier to compare relative sensitivity, and specificity, in cross-sectional studies using histological diagnosis of preinvasive lesions as a surrogate. These results need to be interpreted with caution, since a proportion of such preinvasive lesions will have little potential to progress to invasion. However, if the new test identifies both the lesions detected by the existing test (or the great majority of such lesions) together with additional lesions, the problem of differential regression of lesions identified by the new test is of less concern. Nevertheless, confirmation, preferably by follow-up studies to establish absolute sensitivity, would normally be required before the new test can be considered of at least equal efficacy to the existing test.

The Working Group concluded that at present no surrogate marker suitable for evaluating the efficacy of a new screening test for cervical cancer has been fully established.

Cytological screening

Randomized trials

Efficacy of cervical cancer screening programmes using cytological testing was never tested in a randomized trial. Evidence has therefore been derived from observational studies (cohort and case–control). Studies relating trends in cervical cancer incidence or mortality to screening have also provided very convincing evidence in support of the effectiveness of these programmes, as described in Chapter 5.

Cohort studies

In the evaluation of cervical cancer screening, the 'exposure' has been defined in cohort studies either by invitation to screening or by participation in screening. In efficacy trials (randomized), there is no bias, but in observational studies several biases must be assumed. In the early studies, the rates expected in the absence of screening were usually calculated from the population rates during the latest period before screening was implemented or from concurrent rates in areas without a screening programme. More recently, reference rates were obtained either from those among non-screened or non-participants, or from the average rates in the general population during the screening period. In studies that followed incidence among participants and compared it with that among non-participants, there is potential for selection of a more healthy group to participate compared with non-participants (see Chapter 3), which would introduce a bias towards overestimation of the impact of screening. In studies using the general population rates as the reference, a large part of the general population had been screened in the programme. Therefore, only the early studies with a non-targeted population (not intended to be screened) as controls will in principle give estimates on efficiency with less bias, i.e., on the effect if screened albeit not in ideal but in routine conditions. This, however, assumes that the risk in the control population was adjusted for the risk among non-responders to screening (Cuzick *et al.,* 1997). Such adjustment was done by the Working Group and is indicated in brackets in the following and called an efficacy estimate. Not all cohort studies provided the necessary information for such a correction, however, and none of the studies with non-responders as controls or incidence in the total population as reference value can yield such an estimate.

Several of the cohort studies reviewed below, which use invasive cervical cancer as the main outcome parameter, were included in a review by Lynge (2000), upon which this chapter is based. Table 59 summarizes the characteristics and main findings of the cohort studies.

British Columbia, Canada

The British Columbia cervical cancer screening project started in 1949. In 1959, about 8% of women above the age of 20 years were screened, and about 44% in 1971. The first cohort study was based on the population incidence data and individual screening records from 1958 to 1966. In 1965, 13 clinical invasive carcinomas were detected among the screened women, with 81.2 cases expected based on the 1955–57 incidence rates (standardized incidence ratio (SIR) = 0.16). Among the unscreened women, the numbers were 67 and 62.1 respectively (SIR = 1.08). The effectiveness of the programme was estimated for the total population at 80 observed versus 143.4 expected (SIR = 0.56). [The efficacy estimate of the SIR corrected for selective participation (Cuzick *et al.,* 1997) was 0.17.] Pre-clinical invasive cancers were not included among the observed cases (Fidler *et al.,* 1968). A later cohort study from the same programme reported cancer incidence rates after negative tests (van Oortmassen & Habbema, 1986) (included in the IARC study (IARC, 1986), see below).

Finland

One of the first indications of the magnitude of the effect of screening for cervical cancer in the Nordic countries was provided by a cohort study (Hakama & Räsänen-Virtanen, 1976). In Finland, an organized screening programme started in 1963 and it gradually developed to become nationwide. A cytological test was offered every fifth year to all women aged 30–55 years. Data on 407 000 women screened at least once during 1963–71 and followed up from their first screening until the end of 1972 revealed that the (relative) risks of invasive cervical cancer were 0.2 after a negative test and 1.6 among non-attenders, in terms of the unit risk of reference from the overall Finnish incidence in years preceding the start of the programme. There were 1.4 million woman-years in the follow-up of invasive carcinoma among screened women, and their average follow-up time was thus 3.5 years. The effect of the screening test applied was 80% (100 x (1.0 – 0.2)), without correction for the selective attendance. The attendance rate was 85%, and the effectiveness of the public health service was estimated at 60%, indicating the result of outcome evaluation of the programme among the whole population in that early follow-up phase. [The efficacy estimate for SIR corrected for selective attendance was 0.22].

A further cohort follow-up study in Finland was based on the follow-up of a sample of 45 572 women with a Papanicolaou group I test result in the mass screening programme during 1971–76, when up to the third invitational round was in action (Viikki *et al.,* 1999). The follow-up was performed using the files of the cancer registry up to 1994, and the reference risk was obtained from expected numbers calculated from the general population rates, including women screened in the programme. Overall, 48 invasive cancers were observed (SIR = 0.5; 95% CI 0.4–0.7). In the five-year follow-up since screening, the SIR estimates were, respectively, 0.3 and 6.5 among those with Papanicolaou group I or II–V results but no malignancy confirmed in the screening episode. Follow-up of those with a positive test result at entry was also reported by Viikki *et al.* (2000). The SIR estimate

Table 59. Characteristics and main findings from the cohort follow-up studies on screening impact on cervical cancer

Location (reference)	Cohort description: Numbers of women, screening period, source of screening data, follow-up period and source of follow-up data[a], screening recommendation	Cervical cancer end-point	Screening assessment	Observed (expected) rate per 100 000	No. of cases/ deaths observed (expected)	Relative risk	Comments
British Columbia, Canada (Fidler *et al.*, 1968)	310 000 screened and 233 000 unscreened women, screening in 1958–65, screening laboratory database, incidence follow-up to 1965, laboratory database and pathological files in the province plus a mortality registry, screening recommendation in 20+ years old women once a year	Incidence, clinical squamous carcinoma (invasive) in 1965	Screened previously at least once in the programme	4.2 (26.2)	13 (81.2)	0.16	Pre-clinical 'occult' invasive carcinoma not included
			Unscreened in the programme	28.8 (26.6)	67 (62.1)	1.08	
Finland (Hakama & Räsänen-Virtanen, 1976)	406 358 screened and 35 279 unscreened women, screening 1963–71, screening registry, incidence follow-up in 1963–72, cancer registry, screening 30–55-year women every 5 years	Incidence, invasive carcinoma after the first programme test	Screened at least once in the programme	7.7 [38.5]	109 [545]	0.2	Average follow-up time 3.5 years among screened women (1.4 million person-years). Expectation was drawn from time before screening.
			Unscreened in the programme	NR	NR	1.6	
Iceland (Johannesson *et al.*, 1978)	Not reported [percentage of women in 1974 who had ever been screened was approx. 89% in women aged 30–54], screening 1964–74, screening registry, mortality follow-up in 1965–74, cancer and mortality registry, screening women of 25–59 (25–70 from 1969) every 2–3 years	Mortality from cervical cancer	Women with an initial negative screening	2.6 (NR)			
			Never-screened, follow-up in 1965–69	29.5 (NR)			
			Never-screened, follow-up 1970–74	23.5 (NR)			
Maribo, Denmark (Berget, 1979; Mellemgaard *et al.*, 1990; Lynge, 2000)	16 187 women invited to the first screening round, screening 1967–70, not described, incidence and mortality follow-up to the end of 1984, not described, not described	Incidence of cervical cancer	Participated in screening (after invitation)		115 (217)	0.53	Comparison group was the whole Danish female population
			Not participated (after invitation)		63 (35.96)	1.75	
Østfjold, Norway (Magnus *et al.*, 1987)	45 960 women without previous invasive or preinvasive lesions of the cervix and invited to screening, screening 1959–77, not described, incidence and mortality follow-up in 1959–82, cancer registry, screening women of 25–59 every 2–4 years	Incidence and mortality, invasive carcinoma [after invitation]	Whole study population	NR	267 (341.5) incidence 103 (123.9) mortality	0.78 0.83	Expectation without screening was calculated from same period and age groups among women in 5 neighbouring regions
			Screened		178 (286.1) incidence 55 (102.6) mortality	0.62 0.54	
			Unscreened		89 (55.4) incidence 48 (21.3) mortality	1.61 2.25	

Table 59 (contd)

Location (reference)	Cohort description: Numbers of women, screening period, source of screening data, follow-up period and source of follow-up data[a], screening recommendation	Cervical cancer end-point	Screening assessment	Observed (expected) rate per 100 000	No of cases/ deaths observed (expected)	Relative risk	Comments
			Women with any negative tests in the programme		125 (259.4) incidence	0.48	
					41 (92.6) mortality	0.44	
			Women with 5 negative tests in the programme		6 (34.2) incidence	0.18	
			Women referred to gynaecologist with other diagnosis than severe dysplasia, ca. *in situ* or invasive carcinoma		11 (26.6) incidence	0.41	
					5 (10.0) mortality	0.50	
Sweden (Sparén, 1996)	386 990 women, screening and population registries, incidence follow-up in 1968–92, cancer registry, any screenings	Screened ever vs never	NR	NR	438 (500 among unscreened)	0.55	
Finland (Viikki *et al.*, 1999)	A sample of 45 572 women screened negative, screening in 1971–76, screening registry, incidence follow-up in 1971–94, cancer registry, screening 30–55 y old women every 5 y	Incidence, invasive carcinoma after 1st, 2nd or 3rd programme smear	Screened negative	NR	48 (94)	0.5	SIR estimate from population rates during the screening period (including the screened population).
			Screened negative, follow-up 5 years			0.3	
			Screened positive, follow-up 5 years			6.5	Follow-up of women with positive results reported also in Viikki *et al.* (2000)
			All attenders, follow-up 5 years			0.7	

[a] Vital status or losses from follow-up were not reported.
NR, not reported
If not available in the original publication, confidence intervals were estimated based on an assumption that the observed number of cases followed a Poisson distribution (indicated in square brackets).

for all attenders combined was 0.7. Among non-attenders, the estimated SIR was 1.6. The approximate relative risk (RSIR) between attenders and non-attenders was 0.7/1.6 = 0.44; and between test negatives and non-attenders 0.3/1.6 = 0.19 (confidence intervals not available). The long-term protection provided by screening scheduled every five years was evaluated. The SIR for invasive cancer remained less than unity (with a 3% annual increase) during all the 23-year follow-up. The SIR of preinvasive lesions exceeded unity at follow-up year 10, i.e., at the second screening round. The results confirmed the appropriateness of the five-year screening interval used in Finland.

Iceland

Cervical cancer screening started in Iceland in 1964 and became nationwide in 1969. Women aged 25–59 were invited every 2–3 years (later extended to women aged 25–70). In 1974, approximately 89% of women aged 30–54 had had at least one test; the age-group-specific coverage proportions varied from 81% to 95%. Among women aged 25–29 and 55–59 years, the proportions were lower (47% and 77%). In all Icelandic women aged 25–59, the mortality from cervical cancer changed from 20 per 100 000 in 1955–59, to 21 in 1960–64, 32 in 1965–69 and 15 in 1970–74 (Johannesson *et al.,* 1978). The rates in never-screened women 25–59 years old were 30 in 1965–69 and 23 in 1970–74. The average mortality rate among women with an initial negative screening result was 2.6 per 100 000 in ten years of follow-up. The population mortality rate had decreased in a later study by 60% between 1959–70 and 1975–78, and the mortality rates among the unscreened population were more than ten-fold greater than among the screened (Johannesson *et al.,* 1982).

Maribo, Denmark

An organized screening programme was started in 1967 in Maribo County, Denmark, among women aged 30–49 years. The 16 187 women who were invited to the first round in 1967–70 were followed up for incidence of cervical cancer to the end of 1984 and the observed numbers were compared with the expected number based on rates for all Danish women. In the 87% of the invited women who participated, 115 cervical cancer cases were observed compared with 217 expected (SIR 0.53), whereas the numbers were 63 and 35.96 (SIR 1.75) among the 13% of invited women who did not participate (Berget, 1979; Mellemgaard *et al.,* 1990; Lynge, 2000). The effectiveness of the programme estimated by the SIR for the total population was 0.70 (178/253). The comparison group was the total population of Denmark, where there was extensive spontaneous screening, and some other organized programmes were operating. The Maribo cohort was later extended to include all women screened in the area in 1967–82; the results on the follow-up for cervical cancer incidence after negative tests were included in the IARC study (see below; IARC, 1986). [The efficacy estimate of SIR corrected for selective attendance was 0.60, but, as mentioned above, it was affected by contamination in the control population].

Manitoba, Canada

A province-wide cervical cytology screening programme was initiated in Manitoba in 1963, and included a screening registry. Cases of cervical cancer were recorded at the Manitoba Cancer Registry (Choi & Nelson, 1986). The data on cancer incidence after a negative screening test were included in the IARC study (see below; IARC, 1986).

Sweden

From 1964 onwards, several counties in Sweden gradually introduced organized cytological screening programmes for cervical cancer. Women aged 30–49 years were targeted, with a four-year screening interval. The programme covered all of Sweden except the municipality of Gothenburg by 1973. All tests within the organized programme were reported to the National Board of Health and Welfare, where 930 127 women were registered with at least one test during the period 1967–75. This cohort was followed up for incidence of invasive cervical cancer to the end of 1980 (Pettersson *et al.,* 1986). The data for women with a negative result at entry were included in the IARC study (see below; IARC, 1986).

In a later study, a cohort of 386 990 women resident in Uppsala and Gävleborg counties was followed up for invasive squamous-cell cancer of the cervix (Sparén, 1996). The screening histories were derived from computerized registers including any cytological tests performed in the area. Record linkage allowed complete follow-up with regard to cancer incidence, migration and deaths during 1968–92. The relative risk of squamous-cell cervical cancer incidence among ever- versus never-screened women was 0.55 (95% CI 0.51–0.61). The lowest age-specific relative risks among the screened women were in the age group 40–59 years (RRs from 0.27 to 0.38).

Østfold, Norway

A regional cervical cancer screening programme was organized in Østfold County, Norway. The first round took place in 1959–65 and the last (fifth) round in 1974–77. A cohort follow-up study included all 45 960 women invited to the first screening in the age-group 25–59 years and not previously diagnosed with cervical cancer. The cohort was followed up to the end of 1982, and the observed incidence and mortality were compared with those of women in five neighbouring counties

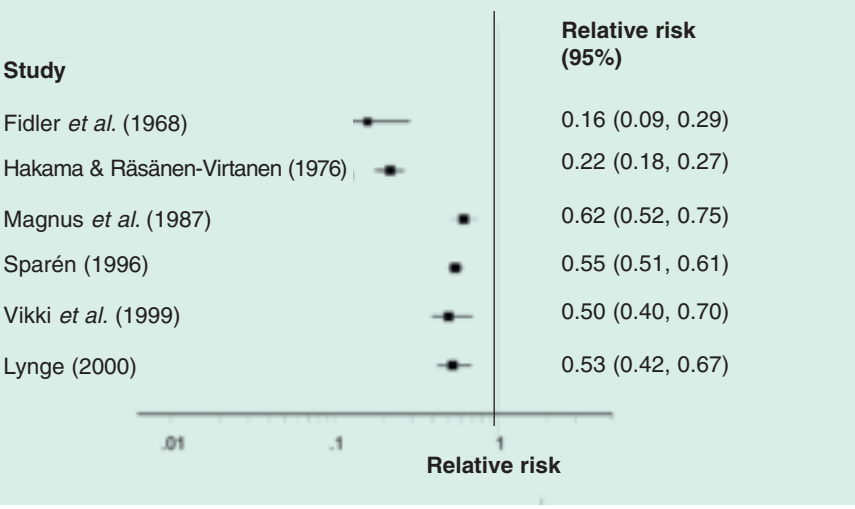

Figure 50 Forest plot of risk ratio estimates of incidence in cohort studies with invitational screening

which did not offer organized screening. During the period 1959–82, 267 new cases of invasive cervical cancer were observed in the Østfold cohort compared with 341.5 expected cases (SIR = 0.78), and 103 deaths from cervical cancer were observed and 124 expected (SMR = 0.83) (Magnus *et al.,* 1987). Women not participating in the screening programme had a 61% higher incidence of cervical cancer than that observed in the reference population and a more than two-fold excess in the mortality rate. Women with any negative test in the programme had an incidence of cervical cancer of 48% and those with five negative tests 18%, compared with the incidence expected from the counties without organized screening. [The efficacy estimate of SIR corrected for selective attendance was 0.77.]

The quantitative results from the cohort studies are illustrated in Figure 50. Four of the cohort studies (in Canada, Denmark, Finland and Norway) allow an estimate of efficacy with presumably only small bias. The results show large variation in effect, with RRs from 0.17 in the British Columbia study to 0.77 in the Østfold county study. Only part of the variation can be accounted for by bias or random variation; most of it is likely to be true. The screening programmes have an effect that varies from close to eradication of invasive disease to a marginal one, which further emphasizes the need for organization and quality assurance, as outlined in other chapters of this volume.

Case–control studies

Case–control studies do not measure the impact of screening in relation to the situation that would be expected in the absence of screening in the population subjected to the screening programme. Instead they compare the risks among the screened to that among non-screened or never-screened groups. As the absolute risks remain unknown, it is not possible to adjust for selective attendance. Despite the inherent biases in their design, resulting in overestimation of efficacy, case–control studies, like cohort studies, have been crucial in the assessment of the efficacy of screening. In 1986, the conclusions of the IARC Working Group on Cervical Cancer Screening were based on a review of studies performed in a number of countries with widely different approaches (Hakama *et al.,* 1986). Among these, the results of five case–control studies were analysed; two (Macgregor *et al.,* 1985; Geirsson *et al.,* 1986) were nested case–control studies within organized programmes (in Aberdeen, Scotland, and in Iceland) and three (Clarke & Anderson, 1979; Berrino *et al.,* 1986; Raymond *et al.,* 1984) in areas where screening was not centrally organized (Toronto, Canada; Milan, Italy; Geneva, Switzerland).

The studies in Iceland and Aberdeen were designed to determine the reduction in risk of invasive cervical cancer among women with a previous negative test, in terms of time elapsed since the smear was taken (Macgregor *et al.,* 1985; Geirsson *et al.,* 1986). A combined analysis of the two studies showed a relative protection (RP) of more than ten-fold (RP = 11.1; 95% CI 2.4–52.2) for women with their last negative test performed 0–11 months before the diagnosis of the case, compared with women who had had their last negative test ten or more years before (Moss, 1986). The other three case–control studies were designed to evaluate the effects of cytological screening for invasive cervical cancer. They differed in the criteria used for case and control selection and in the definition of screening history, but the odds ratios (OR) observed for ever versus never screened were similar, ranging from 0.26 to 0.37. Furthermore, the effects were similar to those observed in the other studies when time elapsed since the last screening and number of previous tests were taken into account.

Since the publication of the IARC monograph (Hakama *et al.,* 1986) on the efficacy of cytological screening, several case–control studies have been carried out to evaluate screening programmes and activities. Fourteen studies published between 1986 and 1998 were reviewed by Zappa and

Ciatto (2000): three of these had been carried out in North America, two in Central America, four in Asia and five in Europe. Since then, another four case–control studies have reported effects of cytological screening by screening status (screened versus non-screened), in Mexico (Jiménez-Pérez & Thomas, 1999), Finland (Nieminen *et al.,* 1999), Sweden (Andersson-Ellström *et al.,* 2000) and South Africa (Hoffman *et al.,* 2003). Table 60 summarizes the main characteristics and findings of these studies.

About half of the studies were carried out within organized programmes or with active invitation of women. Exposure was usually defined by the screening (attendance) status, not by screening invitation. It is also possible that there had been some screening activity among the 'non-screened' groups, diluting the exposure contrast and affecting the rates. This might also have been the case in studies using smear archives and registers. In all studies, cervical cancer incidence was used as the outcome. Two studies also considered cervical cancer mortality in separate analyses. Many studies limited cervical cancers to the squamous subtype, whereas two studies made separate analyses for squamous and adenocarcinomas (Herrero *et al.,* 1992; Sato *et al.,* 1997), while in others the histological type of cervical cancer was not specified. In certain studies, attention was paid to the stage of cervical cancer and/or only advanced stages were considered (van der Graaf *et al.,* 1988; Zhang *et al.,* 1989) or separate analyses were carried out for different grades of invasion (Herrero *et al.,* 1992). About half of the studies were population-based and the others were hospital-based. In some studies, controls were selected from subjects with a negative test at the date of diagnosis of matched case (Macgregor *et al.,* 1994); in another (Sobue *et al.,* 1990), the controls for screen-detected cases were selected among subjects with a negative test in the same year as the diagnosis of the respective case. Two of the studies were nested within cohorts of women invited to be screened (Zhang *et al.,* 1989; Sato *et al.,* 1997).

The proportion of controls 'ever screened' may give an idea of the coverage of cytological testing in the general population, although in most studies controls were matched to cases for several co-variables, so that the actual coverage cannot be directly estimated. Proportions tested differ substantially from one study to another. The proportion of ever-screened controls ranged from 20% in Osaka and 37% in Bangkok to 88% in Miyagi and 93% in Maryland. Most studies tried to identify and exclude tests performed because of symptoms by excluding those performed within 6 or 12 months before the index date.

In spite of the differences mentioned in relation to eligibility criteria for cases and/or controls, methods of collection of screening history and adjustment for confounding variables, the results from the review by Zappa and Ciatto were quite similar, ORs ranging from 0.27 in the Danish study to 0.43 in the Canadian study. Some results fall outside this range, with lower risks in the studies in Miyagi, Japan and Jingan, China (OR = 0.16) and a higher risk observed in Mexico City (OR = 0.76). In the latter study, the OR fell to 0.38 (95% CI 0.28–0.52) when only tests performed in the absence of gynaecological symptoms were considered.

Two fairly recent studies show smaller impact of screening, with ORs in the range 0.5–0.8 (Nieminen *et al.,* 1999 for opportunistic screening; Andersson-Ellström *et al.,* 2000). In the Finnish study (Nieminen *et al.,* 1999), there was a clear difference between self-reported screening in the organized programme and that in opportunistic screening. The OR of cervical cancer was 0.25 (95% CI 0.13–0.48; all ages included) among those who participated in the organized screening only (but who were not screened in the spontaneous screening modality) in comparison with the non-screened; the corresponding OR was 0.57 (95% CI 0.30–1.06) for those who had at least one test in the spontaneous modality only (and did not participate in the organized programme). Most women had had tests in both screening modalities (OR = 0.27; 95% CI 0.15–0.49). The OR for women in both screening modalities (organized and spontaneous) versus those with spontaneous screening alone was 0.47 (95% CI 0.29–0.75). This difference in effect was obtained with less resources in the organized screening (see Chapter 3).

In the Swedish study (Andersson-Ellström *et al.,* 2000), the effect of screening was estimated in comparison with women who had not been tested during the last six years before the index date. Some of these might have been tested earlier, diminishing the screening contrast. There was a large proportion of cases, compared with controls, in whom carcinoma *in situ* or another (milder) lesion had been previously treated (see Table 60). In only 18 cases (16%) had all the previous tests been negative (the corresponding figure for controls was not given). [These findings suggest that inadequacies in the management of treatment may have, at least partly, accounted for the rather modest effect of screening; or that a fraction of the case women had been tested in the course of management follow-up activity.]

Figure 51 presents a summary of results from the case–control studies on incidence included in Table 60. For the study by Hernandez-Avila *et al.* (1998), the results excluding tests performed on account of symptoms were considered most relevant for the purpose of estimating efficacy in this

diagram. From the study by Nieminen *et al.* (1999), only the results on organized screening were included. The study by Andersson-Ellstrom *et al.* (2000) was not included, as symptomatic women may have been included. There was a strong indication of heterogeneity in the results over all studies reported in the table, reflecting the many differences in studies mentioned above, whereas in the results selected for Figure 51, the heterogeneity is very much less (p = 0.288, calculated after Der Simonian & Laird, 1986). There was indication of publication bias, however (p = 0.023, calculated after Begg & Mazumdar, 1994). Most studies reported large decreases in the cervical cancer risk attributable to screening, although because of the limitations of and selection in the individual studies, one needs to be cautious when interpreting the pooled point estimate (0.34) of the impact from the case–control studies. A crude attempt to adjust for bias can be made by assuming that selection was the most important source of bias and that it was relatively constant over the studies. The relatively high homogeneity over the case–control studies in the estimated ORs is another justification for such asn assumption. The ORs in the cohort studies in risk between non-responders and controls (not non-responders) was about 1.5, which indicates that the non-responders may have an inherent risk up to two times that among responders. Such an adjustment would imply a true protective effect in the populations subjected to case–control studies of screening of about 0.7. If the estimates of ORs in the case–control studies were 0.6 or larger, there is a possibility that the programme was practically without effect.

When death from cervical cancer is taken as the end-point, the protective effect of screening tends to be slightly higher than estimated from the incidence studies. In the Scottish study (Macgregor *et al.,* 1994), the OR for mortality was 0.25 (95% CI 0.11–0.48) and for incidence 0.35 (95% CI 0.25–0.50), while in the Osaka study (Sobue *et al.,* 1990) the corresponding figures were 0.22 (95% CI 0.03–1.95) and 0.41 (95% CI 0.13–1.29). Most of the studies did not report impact on mortality.

Cervical cancer incidence after screening negative

The IARC joint study on the incidence of invasive cervical cancer by number of previous negative tests (Day, 1986; IARC, 1986) was based on data from ten centres worldwide from which individual screening histories were available and could be linked to cancer registry data. Five were cohort studies, two nested case–control studies and three population-based case–control studies. The cohort studies were those listed above from British Columbia (van Oortmarssen & Habbema, 1986), Manitoba (Choi & Nelson, 1986), Sweden (Petterson *et al.,* 1986), Norway (Magnus & Langmark, 1986) and Denmark (Lynge & Poll, 1986a, b). The nested case–control studies (Macgregor *et al.,* 1985; Geirsson *et al.,* 1986) were carried out within organized screening programmes in Aberdeen and Iceland. The population-based case–control studies (Clarke & Anderson, 1979; Berrino *et al.,* 1986; Raymond *et al.,* 1984) were from areas where screening was not centrally organized (in Toronto, Milan and Geneva, respectively). In the case–control studies and in one of the cohort studies (British Columbia), the reference population was the potentially selected group of unscreened women; in the other cohort studies, the expected incidence in the absence of screening was derived from corresponding population incidence rates in a period before mass screening was started. A negative result was defined in the IARC study as either a Papanicolaou group I result or one or two suspicious (group II) results followed by a group I result. The relative risk of squamous-cell carcinoma of the cervix among women aged 35–64 years, whose second negative test occurred at age 35, by time since the index negative smear is given in Table 61. The risk estimates were 0.07 (95% CI 0.04–0.10) during the first year (12 months), 0.08 (95% CI 0.05–0.13) during the second year, 0.13 (95% CI 0.08–0.19) during the third year and 0.36 (95% CI 0.25–0.53) during the fifth year since screening negative. These risks were used to calculate the cumulative percentage reduction in risk of squamous-cell carcinoma of the cervix assuming different screening intervals (see below).

One recent cohort follow-up study (Van den Akker-van Marle *et al.,* 2003a) and two case–control studies (Miller *et al.,* 2003; Sasieni *et al.,* 2003) have also reported cervical cancer incidence rates after negative screening results. These studies collected data on screening history from archive sources and followed cervical cancer incidence since time from the index negative smear.

Van den Akker-van Marle *et al.* (2003a) followed invasive cervical cancer incidence among women who tested negative in the Dutch screening programme during 1975–97. Data on screening were derived from a national pathological archive, and information on cervical cancers was obtained from the same source for the period 1994–97. Incidence rates were calculated for women aged 35–64 years with one and with two previous negative tests. A negative screen was defined as an episode consisting of a cytological or histological examination with a negative result, or a cytological examination with a positive result but without histological confirmation of invasive cervical cancer or a precursor.

Table 60. Main characteristics and results of case–control studies on cervical cancer screening published after 1986 (modified and updated from Zappa & Ciatto, 2000)

Country (reference)	Outcome, period of observation, number and source of cases and controls	Screening modality	Proportions of cases/ controls ever screened (%)	OR ever vs never screened	95% CI	Data source for screening information. Notes
Bangkok, Thailand (Wangsuphachart *et al.*, 1987)	Incidence (all histological types, ages 15–54 y), 1979–83, 189/1023, hospital records	Not invitational	30/37	0.39 (screened every 2–5 y vs never)	0.21–0.74	Questionnaire
Denmark (Olesen, 1988)	Incidence (all histological types, mean age 52.6 y), 1983, 428/428, cancer registry	Invitational	45/67	0.27	0.18–0.42	Questionnaire to general practitioners
Nijmegen, Netherlands (van der Graaf *et al.,* 1988)	Incidence (FIGO >1A, age <70 y), 1979–85, 36/120, cancer registry and registrar's office	Invitational	47/68	0.22	0.1–0.81	Questionnaire
Maryland, USA (Celentano *et al.,* 1988)	Incidence (age 22–84 y), 1982–84, 153/153, hospital admission records	Not invitational	72/93	0.29 (screened within 3 y vs never)	0.15–0.58	Interview
Washington, USA (Shy *et al.,* 1989)	Incidence (FIGO >1B-occult, ages 31–75 y), 1979–83, 92/178, cancer registry	Not invitational	85/93	0.21	0.09–0.50	Telephone interview. Smears collected in the follow-up of an abnormal test or at the cancer diagnosis were excluded. OR estimate a re-calculation by Zappa & Ciatto (2000).
Jingan, China (Zhang *et al.,* 1989)	Incidence (FIGO >1A, squamous), 1965–74, 109/545, screening archive	Invitational	Not available	0.16 (smears performed within last 2 y vs smears performed 6 or more y earlier)	0.05–0.58	Archive
Osaka, Japan (Sobue *et al.,* 1990)	Incidence (ages 30–79 y), 1965–87, 28/272, cancer registry and dwelling history	Invitational	25/39 (within 10 years)	0.41 (screened within 10 y vs not screened within 10 y)	0.13–1.29	Archive. Only negative tests included.
Osaka, Japan (Sobue *et al.,* 1990)	Mortality (age <80 y), 1965–87, 15/150, cancer registry and dwelling history	Invitational	7/20 (within 10 years)	0.22 (screened within 10 y vs. not screened within 10 y)	0.03–1.95	Archive. Includes diagnostic smears
Florence, Italy (Palli *et al.*, 1990)	Incidence (age <75 y), 1982–85, 191/540, cancer registry and residents list	Invitational	19/48	0.29	0.15–0.55	Archive
Bogota, Mexico City, Panama, Costa Rica (Herrero *et al.,* 1992)	Incidence (age <70 y), 1986–87, 759/1433, cancer treatment centres, hospital admission list and partly from census list.	Not invitational	50/72	0.40	0.31–0.48	Interview

Table 60 (contd)

Country (reference)	Outcome, period of observation, number and source of cases and controls	Screening modality	Proportions of cases/ controls ever screened (%)	OR ever vs never screened	95% CI	Data source for screening information. Notes
Manitoba, Canada (Cohen, 1993)	Incidence (ages 25–64 y), 1981–84, 415/29269, cancer registry and residents list	Not invitational	76/87 (within 10 y)	0.43	0.32–0.57	Health care files
South-east Scotland, (Macgregor *et al.,*1994)	Incidence (squamous CC), 1982–91, 282/564, screening records	Invitational	45/73	0.35	0.25–0.50	Cytopathology database
South-east Scotland, (Macgregor *et al.,* 1994)	Mortality (squamous CC), 1982–91, 108/216, screening records	Invitational	35/73	0.25	0.11–0.48	Cytopathology database
UK (Sasieni *et al.,* 1996)	Incidence (age >20 y), 1992, 348/677, pathology laboratories and registry of local health authority	Invitational	73/85	0.26 (tests performed 24–35 months before, vs not screened or screened >66 months before)	0.14–0.47	Archive. These data were included in the later study (Sasieni *et al.*, 2003)
Miyagi, Japan (Sato *et al.*, 1997)	Incidence (ages 35–79 y), 1984–89, 119/218, screening archive	Invitational	55/88	0.16 (screened within 5 y vs not screened within 5 y)	0.09–0.28	Interview and archive
Mexico City, Mexico (Hernandez-Avila *et al.*, 1998)	Incidence, 1990–92, 397/1005, hospital admissions records and sample of residents	Not invitational	42/51	0.76	0.59–0.98	Interview. OR = 0.38 (95% CI 0.28–0.52) when tests due to gynaecological symptoms were excluded.
Guadalajara, Mexico (Jiménez-Pérez & Thomas, 1999)	Incidence (age <70 y), 1991–94, 143/311, hospital records	Not invitational	54/82	0.3	0.2–0.4	Interview
Finland (Nieminen *et al.*, 1999)	Incidence, 1987–94, 147/1098, hospital records and population files	Invitational Not invitational	56/72 all ages 68/88 ages 30–59 64/66 all ages 80/80 ages 30–59	0.36 0.32 0.73 0.85	0.25–0.53 0.19–0.57 0.49–1.07 0.45–1.60	Questionnaire
Värmland, Sweden (Andersson-Ellström *et al.*, 2000)	Incidence (ages 20+ y), 1990–97, 112/112, pathology and population files	Any smears (about 50% of the tests were after invitation)	61/65 (within 6 y) 83/88 in ages 20–59 (within 6 y)	[0.83] [0.62]		Pathology database. 16 cases (14%) and 4 (4%) controls had been previously treated for carcinoma *in situ* of the cervix ($p < 0.01$); 32 cases (29%) and 6 controls (5%) had former atypia ($p < 0.001$)
Western Cape, South Africa (Hoffman *et al.*, 2003)	Incidence (stage >IA), 524/1540, hospital records	Not invitational	50/73	0.3	0.3–0.4	Interview. OR 0.2 among those with at least 3 tests; and 0.3 among those with <10 y since the last screen

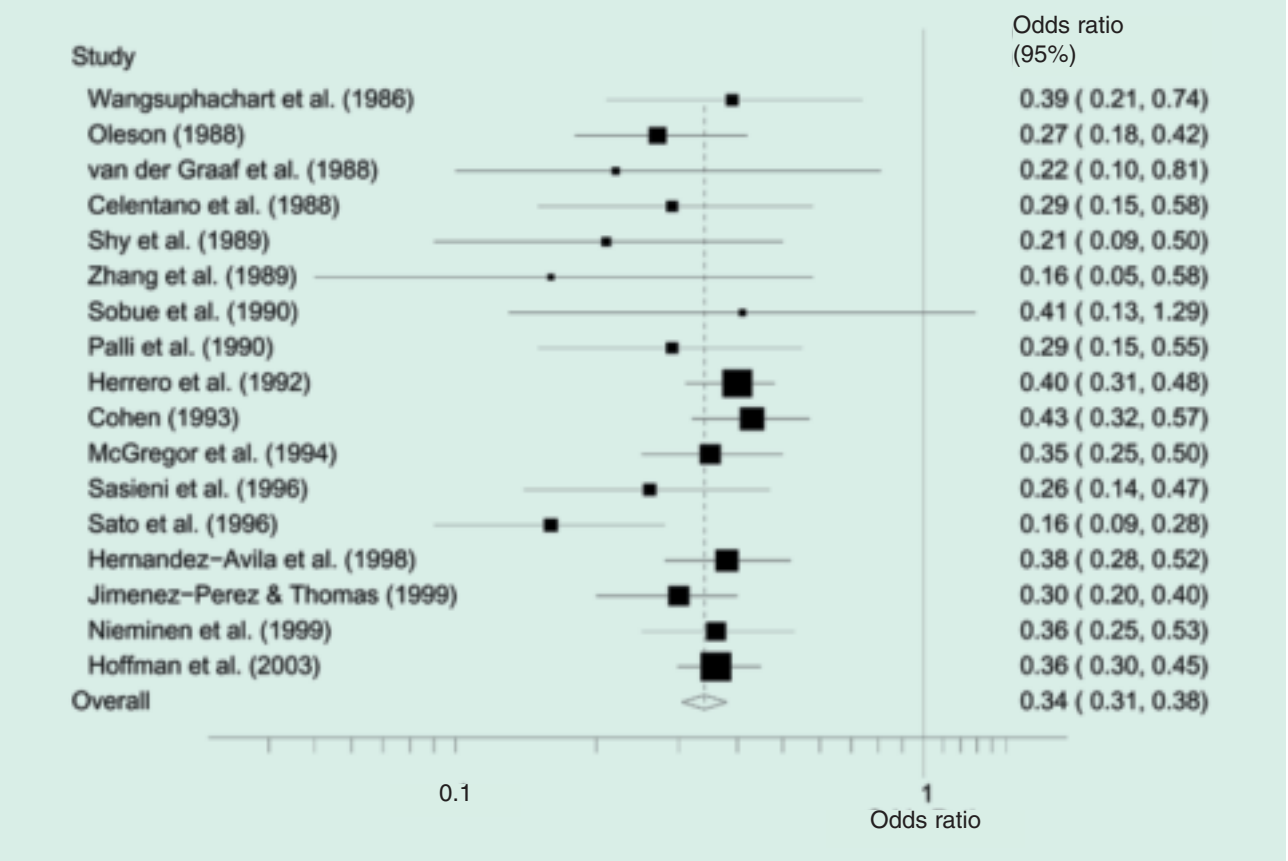

Figure 51 Forest plot of results from case–control studies with invitational and non-invitational screening, including a pooled odds ratio estimate (incidence) using a random effects model (Der Simonian & Laird, 1986)

The incidence expected in the absence of screening was estimated using incidence data from three regions during 1965–69 (the latest period before the screening programme was started) covering 8% of women in the whole country. In addition, age-period-cohort (APC) modelling was used to refine the expected incidence without screening, using the same pre-screening period incidence data as the input. The relative risk of invasive cervical cancer increased from 0.13 in the first year after screening to 0.24 after more than six years from screening for women with one previous negative screening (confidence intervals not available). These figures decreased to 0.12 (95% CI 0.08–0.17) and 0.06 (95% CI 0.03–0.10) for 0–6 and 7–12 months since the last negative screening and 0.18 (95% CI 0.11–0.30), respectively, for more than six years among women with two or more previous negative screening results. The identification and linking method used in the pathology register was not perfect (i.e., the identification code consisted of the same characters for two or more women); this was considered to have produced an upward bias in the incidence rate after a negative test. As a consequence, the true reduction in relative risk might have been somewhat larger than reported. On the other hand, the analysis using APC modelling suggested overestimation of the background risk.

Miller *et al.* (2003) analysed negative cytological histories within a selected group of women having continuous participation in the Kaiser Permanente medical care programme in northern California, USA, for at least 30 months before the diagnosis date of the cervical cancer cases. The cases (N = 482), diagnosed between 1983 and 1995, were drawn from the files of the medical care programme, SEER and the California cancer registry; controls (N = 934) were matched for age, length of membership and race. About 92% of women aged 20 years or more

Table 61. Relative risk of invasive carcinoma of the cervix within different follow-up windows since screening negative, in comparison with expectation in the absence of screening

Time interval since screening years (months)		Relative risk (95% confidence interval) IARC (1986) [a] Ages 35–64		
1	(0–11)	0.07 (0.04-0.10)		
2	(12–23)	0.08 (0.05-0.13)		
3	(24–35)	0.13 (0.08–0.19)		
4	(36–47)	0.19 (0.13–0.28)		
5	(48–59)	0.36 (0.25–0.53)		
6	(60–71)	0.28 (0.17–0.48)		
7–10	(72–119)	0.63 (0.30–1.67)		
		Van den Akker-van Marle *et al.* (2003a) [b] Ages 35–64		
1	(0–6)	0.12 (0.08–0.17)		
1	(7–12)	0.06 (0.03–0.10)		
1–2		0.08 (0.06–0.12)		
2–4		0.15 (0.11–0.19)		
4–6		0.20 (0.14–0.29)		
6–10		0.18 (0.11–0.30)		
		Sasieni *et al.* (2003)[c]		
		Ages 20–39	**Ages 40–59**	**Ages 55–69**
1	(0–18)	0.24 (0.16–0.37)	0.12 (0.08–0.18)	0.13 (0.08–0.22)
2	(18–30)	0.33 (0.21–0.51)	0.14 (0.08–0.22)	0.13 (0.07–0.23)
3	(30–42)	0.67 (0.43–1.04)	0.25 (0.16–0.40)	0.15 (0.08–0.26)
4	(42–54)	1.06 (0.65–1.72)	0.30 (0.18–0.50)	0.18 (0.09–0.34)
5	(54–66)	1.40 (0.75–2.62)	0.61 (0.34–1.09)	0.28 (0.14–0.57)
6	(66–78)	1.86 (0.88–3.93)	0.72 (0.36–1.43)	0.33 (0.14–0.79)
>6	(>78)	2.37 (1.16–4.85)	0.69 (0.36–1.34)	0.55 (0.27–1.10)

[a] Including invasive squamous-cell carcinoma of the cervix uteri, since the last negative test at ages 35–64 years, in comparison with expectation in the absence of screening. Assuming that a woman is screened negative at age 35 and that she had at least one negative screen previously.
[b] Invasive cervical cancer for 35–64-year old women since two or more previous negative screenings, in comparison with expectation without screening.
[c] Invasive cervical cancer in various age groups since the last operationally negative smear.

had been screened at least once and 89% within the last three years. A test was defined as negative if the cytological result did not require a change in the follow-up interval (i.e., no referral or control test was required). In addition, there was a group of 'other smears' including those for which the result was missing or unrelated to invasive cancer (e.g., atypical or dysplastic endometrial cells, atrophic changes, *Trichomonas* infection); and a group of 'abnormal' results. For 32% of the cases and 10% of the controls, no negative results were available (indicating either that the women had been screened elsewhere or had not been screened at all, or that they had had positive or other smears). In the follow-up of the last negative test (irrespective of the other two groups in earlier screenings), the OR for a two-year (19–30 months) versus one-year (0–18 months) follow-up interval was 1.72 (95% CI 1.12–2.64) and for a three-year (31–42 months) versus one-year follow-up interval 2.06 (95% CI 1.21–3.50). Adjustment for ever having had an abnormal result before the index test and for having at least one previous consecutive negative result within 36 months before the index test did not essentially change the results. For the sub-sample of women with at least two consecutive negative results, the OR was 2.15 (95% CI 1.12–4.11) for the two-year follow-up and 3.60 (95% CI 1.50–8.68) for the three-year follow-up, as compared with the one-year follow-up. [The study did not quantify the overall reductions in cervical cancer attributable to screening. The Working Group noted that the baseline risk with one-year follow-up was difficult to estimate and could be subject to bias; therefore the results of this study were not included in the table.]

A study in the United Kingdom (Sasieni *et al.,* 2003) used screening data on women registered within a group practice drawn from a computerized database of the screening programme; information on cervical cancer cases was obtained from pathology laboratories. There were 1305 women aged 20–69 years, diagnosed between 1990 and 2001 with frankly invasive cervical cancer, and 2532 age-matched controls. It was not possible to identify which cancers were screen-detected, because some 50% of the women screened in England in the mid-1990s did not attend in response to an invitation to the group practice. In all analyses, the date of

diagnosis for a case was used as the index date and in each case–control stratum only the registered tests performed before that date were considered. An operationally negative result was defined as a negative one not preceded by an abnormal one (borderline or worse) within the previous 12 months. Overall, 66% of the cases and 80% of the controls had at least one recorded test (with any result); the figures were 48% and 71% in age group 55–69; 68% and 85% in age group 40–54; and 80% and 83% in age group 20–39 years. Compared with those who never had a negative test, the ORs for invasive cervical cancer varied among women aged 55–69 years from 0.13 (95% CI 0.08–0.22) in the follow-up window of one year (0–18 months) since the last negative test to 0.28 (95% CI 0.14–0.57) in that of five years. The corresponding ORs among women aged 40–54 years were from 0.12 (95% CI 0.08–0.18) to 0.61 (95% CI 0.34–1.09) and among women aged 20–39 years from 0.24 (95% CI 0.16–0.37) to 1.40 (95% CI 0.75–2.62). The higher risk estimate of women screened negative as compared with non-screened among the youngest age group might be related to selection among those who attended regular screening.

Age to start screening

The incidence of carcinoma of the cervix is very low in women aged less than 25 years, but then begins to climb. However, in an extension of the British Columbia cohort study, the incidence of carcinoma *in situ* at age 20–24 was of the order of 16 per 100 000 (Miller *et al.,* 1991b), encouraging a national workshop in Canada to recommend that screening should start at the age of 20 years (Miller *et al.,* 1991a). Similar conclusions have been drawn by other North American advisory committees (e.g., Saslow *et al.,* 2002).

Other countries have taken a different view. They have noted that although young women below the age of 25 or 30 have much higher rates of cervical abnormality than older women, the rise in cervical cancer incidence does not take place until the next decade. While treatments are very successful and have very low rates of complications, the consequences for a young woman can be much greater than for an older woman, for whom preservation of fertility is not an issue. For younger women, the risk of harm may be greater than the risk of benefit.

In Europe, the age to start screening varies widely, with women in Finland and the Netherlands invited to the organized programmes from the age of 30 years, while some countries start screening at much younger ages (Miller, 2002b). Sasieni *et al.* (2003), on the basis of a study principally designed to determine the frequency of re-screening, found that the effectiveness of cytological screening was relatively low in young women, but rose in older women (lower part of Table 61). This led to the decision in England to move from a recommended age of 20 years for starting screening to the age of 25 years.

For developing and some middle-income countries, in order to maximize use of resources, and given the infrequency of cervical cancer below the age of 35 years, it is generally recommended to start screening at 35 years and only extend screening to younger ages when resources permit (WHO, 1986). It has been pointed out that age is the most important risk factor forcervical cancer and that screening should aim to target high-risk women. A good guide would be to take the age at the beginning of the rise in incidence of cervical cancer and begin screening five years before this age. In most countries, this would be at about 30–35 years of age (Miller *et al.,* 2000).

Frequency of re-screening

Screening programmes seek to maximize the reduction in incidence of and mortality from disease, for a given level of resources. The optimal screening interval is one that provides the most favourable ratio between degree of disease control and cost of screening. The design of a screening programme defines two key parameters for achieving these objectives, the target population and the screening interval. Compliance with these parameters is crucial in maintaining the effectiveness of the programme and in measuring its cost-effectiveness in order that resources can be used to increase population coverage (see Chapter 3). Significant deviation from the recommended screening interval or target population may reduce the programme efficiency either by using excessive resources, as in the case of annual re-screening for cervical cancer, or by allowing the disease to 'escape' the period at which early intervention can lead to treatment and/or cure. Models can facilitate decisions on the optimal periodicity of screening.

Determining the frequency of screening is helped by understanding the natural history of the condition to be screened for, especially the duration of the asymptomatic (latent) phase. A high frequency of screening will result in a low number of cases per screen and thus a low predictive value. The reason for this is that the prevalence of asymptomatic disease will be low in the population if the screening frequency is high. On the other hand, screening too infrequently will leave much of the disease uncontrolled (Cole & Morrison, 1980).

An early evaluation of cervical cancer screening in British Columbia using a Markov–Chain model supported a prolonged natural history of carcinoma in situ (an average sojourn time of at least nine years) and suggested that cytologically negative women should be rescreened every five years (Shun-

Zhang *et al.,* 1982).

The IARC Working Group on Cervical Cancer Screening Programmes (IARC, 1986) established a proper approach to re-screening. This study showed that there was very little evidence to support annual screening and largely provided the basis for international recommendations for three-yearly or even less frequent screening. It underlined the importance of concentrating screening between the ages of 35 and 64 years, with almost as much benefit expected from three-yearly screening as from annual re-screening (Tables 61 and 62). These findings have been reinforced by a study in the Netherlands (Van den Akker-van Marle *et al.,* 2003a; see Table 61).

It is important to note that the greatest percentage reduction in cumulative incidence can be obtained only if a high proportion of the population complies with screening. However, even in the best of circumstances, experience in highly efficient cytology screening programmes of many countries shows that no realistic screening schedule results in the abolition of invasive cervical cancer.

Part of the reason for the imperfect outcome of screening programmes is failure of an essential component of the programme, which can occur at the level of the woman, her physician or the laboratory examining the cytological smears (Miller, 1995). However, another reason is likely to be the variability of the natural history in different women. The models are based on averages of transition probabilities, each with a different distribution or range of time periods during which some lesions progress from one stage to the next, while others regress to normal, and still others remain stable for long periods of time. Some lesions may progress so rapidly that they cannot be found in a curable stage even with annual screening, and it seems unlikely that the majority of such lesions would be detected by more frequent screening. This does not mean there are different types of cancer of the cervix, as suggested many years ago (Ashley, 1966), just that the fast-growing lesions represent one extreme of the distribution of progression (sojourn) times.

Celentano *et al.* (1989) conducted a case–control study of 153 cases of invasive cancer, 153 case-nominated controls and 392 randomly selected controls. The results were largely congruent for the two sets of controls. The relative protection after a self-reported negative test was significant for 2–3 years from the last negative test (OR = 8.28; 95% CI 3.44–19.9 for case-nominated controls, OR 4.62; 95% 2.04–10.5 for randomly selected controls) and some degree of protection was seen for 4–6 years (OR = 4.30; 95% CI 1.46–12.7; 3.63; 95% CI 1.38–9.57, respectively) after adjustment for a variety of confounders.

Herbert *et al.* (1996) studied the incidence of cervical cancer in a region of the UK after the introduction of the UK computerized call-and-recall system. The incidence of invasive cancer was significantly higher in women who had not been screened in the previous five years than in those who had (RR = 2.6; 95% CI 1.6–4.3); the incidence was higher in those with an interval of 3.5–5.5 years compared to 0.5–3.5 years (RR = 2.2; 95% CI 1.3–3.8). The RRs were higher when screen-detected cancers were excluded. The authors concluded that a five-year interval is too long.

Viikki *et al.* (1999) studied the risk of cervical cancer after a negative test in the context of the five-yearly organized screening programme in Finland (see above). They found that the SIR was low initially after screening and increased gradually until the time the next test was due. There was an estimated 3% annual increase in risk of invasive disease. The risk did not reach the national average within the more than 20 years of follow-up. They also found that the relative risk of a preinvasive lesion after an initial negative result was decreased up to the second rescreening round at 10 years and concluded that the five-year screening interval applied in Finland was appropriate.

Goldie *et al.* (2001) modelled the natural history of cervical cancer using published data on transition and regression rates, and data from a study in Cape Town, South Africa. They concluded that in developing countries, if the limited resources are such as to allow three screenings in a lifetime, it may be more cost-effective to give these tests every five years from the age of 35 or 40 years, rather than every 10 years from the age of 35, as had

Table 62. Percentage reduction in the cumulative rate of invasive cervical cancer over the age range 35–64 years, with different frequencies of screening

Screening frequency	% reduction in the cumulative rate*	Number of tests
1 year	93.5	30
2 years	92.5	15
3 years	90.8	10
5 years	83.6	6
10 years	64.1	3

* Assuming a screen occurs at age 35 years, and that a previous negative screen had been performed
From IARC (1986)

been modelled by the IARC (1986) study, and subsequently adopted as a suggested policy for South Africa (Provincial Administration Western Cape: Department of Health, 1995).

Miller *et al.* (2003) conducted a case–control study of cases of invasive cancer diagnosed between 1983 and 1995 within the Kaiser Permanente medical care programme (see above). The ORs for various intervals between screens, with a one-year interval as the referent, adjusted for ever having had an abnormal cytological finding before the last negative result and for having at least one negative result within 36 months before the last negative one, increased to 2.24 (95% CI 1.28–3.92) at 2 years, 3.37 (1.97–5.76) at 3–5 years and 5.72 (3.48–9.41) at 5–10 years, although there was a low absolute risk of developing invasive cervical cancer within three years of a previous negative result. [The Working Group noted that the cited odds ratios were entirely dependent on the validity of the estimated referent risk level.]

Sawaya *et al.* (2003) studied the prevalence of biopsy-proven cervical neoplasia among 938 576 women under 65 years of age. The prevalence of all grades of CIN was highest in women aged less than 30 years and much higher in those with no previous negative cytological result than in those with one or more. No invasive cervical cancers were detected in those who had had three or more previous negative tests. Using a Markov model, and various rates of progression and regression from the literature, they estimated that for women aged 30–64 years who had had three or more consecutive negative tests, extending the re-screening interval from one year to every three years would result in an average excess risk of about 3 per 100 000.

Sasieni *et al.* (2003) conducted a case–control study in the United Kingdom based on the screening histories of 1305 women age 20–69 with stage 1B cancer of the cervix and 2532 age-matched controls from the records of the screening programme (Table 61). The OR for occurrence of cervical cancer increased with time from last negative result; it reached 1.0 (no protection) at three years for women aged 20–39, approached 1.0 at six years for women aged 40–54 and was still ~0.5 at six years for women aged 55–69 years. The authors estimated the proportion of cervical cancer that would be prevented by different schedules of re-screening. These proportions varied by age. For women aged 20–39, 30% would have been prevented by five-yearly, 61% by three-yearly and 76% by annual screening. The corresponding percentages for women aged 40–54 were 73%, 84% and 88%, and for women aged 55–69 were 83%, 87% and 87%, respectively. On the basis of these results, the authors recommended three-yearly screening for women aged 25–49, five-yearly screening for women aged 50–64 and, for women aged 65 or over, screening only of those who had not been screened since age 50.

Age to stop screening

Many countries recommend stopping screening or inviting women at around age 60 or 65 years, for a number of reasons. For example, older women have tended to be poor attenders for screening, and good-quality smears are difficult to obtain in women so far past the menopause. In addition, if they have had regular tests with a normal outcome in the past, they are considered to be at low risk of developing cervical cancer. However, in view of the relatively high age-specific incidence rates of invasive cervix cancer in all countries in older women, there is a consensus in developed countries that women over the age of 60 years who have never been screened or have not been screened for many years should be encouraged to have at least two tests, and only if both are negative should they stop screening (e.g., Miller *et al.,* 1991a; Sasieni *et al.,* 2003). However, in developing countries where resources are limited but available for some screening in older women, it has been recommended that women who have never previously been screened and are older than 60 years of age should have one test only (Miller *et al.,* 2000).

Cecchini *et al.* (1996) reviewed data for women aged 60–70 years from the Florence screening programme and the Tuscany Cancer Registry. Only five of 242 women with invasive cervical cancer had had two or more negative results between 50 and 60 years of age. However, of 11 342 women aged 58–60 who had a negative test between 1980 and 1987 and were followed to December 1990, only one invasive cancer was diagnosed, compared with 13.95 expected from age-specific incidence data from the cancer registry (OR = 0.07; 95% CI 0.002–0.39). The authors recommended reconsideration of continuing screening after 60–64 years of age.

In North America, it is generally recommended that women who have been actively screened and always been negative should cease screening at 69–70 (Miller *et al.,* 1991a; Saslow *et al.,* 2002). In Europe, the guidelines recommend 64 years as the upper age limit for active invitation for screening (Coleman *et al.,* 1993).

There have been suggestions that women who have been active participants in screening but never had a cytological abnormality could stop screening at younger ages (e.g., 55 or even 50 years) (Cruickshank *et al.,* 1997). Flannelly *et al.* (2004) analysed screening data for women aged 50 years and over who had had a satisfactory result between 1988 and 1996 (N = 36 512) from five regions in

England and Scotland. Women with prior dyskaryosis or borderline nuclear abnormalities had RRs for a positive test after the age of 50 of 4.39 and 3.08, respectively, compared with women whose screening history before the age of 50 was negative. However, 1.8% of women with a negative screen history before the age of 50 had dyskaryosis detected after the age of 50 during a median duration of follow-up of 33.2 months.

Visual inspection

Four screening techniques based on visual inspection have been assessed for early detection of cervical neoplasia, mostly in low-resource settings:

- Unaided visual inspection (also-known as downstaging)
- Visual inspection with 3–5% acetic acid (VIA)
- Visual inspection with acetic acid using low-level magnification (VIAM)
- Visual inspection with Lugol's iodine (VILI)

Unaided visual inspection involves naked-eye visualization of the cervix, without application of acetic acid, to identify abnormal tissue harbouring cervical neoplasia, particularly invasive cancer. Cross-sectional studies in India have shown low sensitivity (30–50%) for unaided visual inspection to detect cervical cancer precursors and it is no longer considered a suitable screening test (Sankaranarayanan *et al.,* 1997; Basu *et al.,* 2002). VIAM involves the use of low-level magnification (2–4 x) in visualizing acetowhite lesions after application of acetic acid. The test characteristics of VIA and VIAM have been evaluated in cross-sectional studies in India and South Africa (Denny *et al.,* 2000a, 2002; Sankaranarayanan *et al.,* 2004e). The results from these studies indicate that magnification did not improve the test performance over and above that of naked-eye visualization. Low-level magnification is no longer widely used for visualization after application of acetic acid.

VIA

VIA involves naked-eye inspection of the cervix one minute after application of 3–5% dilute acetic acid. VIA has been widely investigated for its test characteristics in detecting CIN 2–3 lesions and invasive cancer in several cross-sectional studies, mostly in developing countries (Slawson *et al.,* 1992; Cecchini *et al.,* 1993; Megevand *et al.,* 1996; Londhe *et al.,* 1997; Sankaranarayanan *et al.,* 1998b, 1999; University of Zimbabwe/JHPIEGO Cervical Cancer Project, 1999; Denny *et al.,* 2000a; Cronje *et al.,* 2001; Belinson *et al.,* 2001; Denny *et al.,* 2002; Rodriguez-Reyes *et al.,* 2002; Ngelangel *et al.,* 2003; Tayyeb *et al.,* 2003; Cronje *et al.,* 2003; Sankaranarayanan *et al.,* 2004a). The relative sensitivity of VIA to detect high-grade precancerous lesions and invasive cervical cancer varied from 29% to 95% and the specificity varied from 68% to 98% in cross-sectional studies suffering from verification bias (Slawson *et al.,* 1992; Cecchini *et al.,* 1993; Megevand *et al.,* 1996; Londhe *et al.,* 1997; Sankaranarayanan *et al.,* 1998b, 1999; Cronje *et al.,* 2001; Tayyeb *et al.,* 2003) (see Chapter 2, Table 26). In cross-sectional studies with minimal verification bias, the sensitivity of VIA to detect CIN 2–3 lesions varied from 37% to 92% and the specificity from 49% to 91% (University of Zimbabwe/JHPIEGO Cervical Cancer Project, 1999; Denny *et al.,* 2000a; Belinson *et al.,* 2001; Singh *et al.,* 2001; Denny *et al.,* 2002; Rodriguez-Reyes *et al.,* 2002; Ngelangel *et al.,* 2003; Cronje *et al.,* 2003; Sankaranarayanan *et al.,* 2004a) (see Chapter 2, Table 26). Conventional cytology was concurrently evaluated in most of the above studies and the sensitivity of VIA was found to be similar to or higher than that of cytology as provided in the respective study settings, but the specificity of VIA was consistenly lower than that of cytology (Londhe *et al.,* 1997; Cecchini *et al.,* 1993; Sankaranarayanan *et al.,* 1998b, 1999; University of Zimbabwe/JHPIEGO Cervical Cancer Project, 1999; Denny *et al.,* 2000a; Cronje *et al.,* 2001; Denny *et al.,* 2002; Cronje *et al.,* 2003; Sankaranarayanan *et al.,* 2004f) (see Chapter 2, Table 27). HPV testing was concurrently evaluated in cross-sectional studies in India, South Africa and Zimbabwe and was found to have sensitivity similar to that of VIA (Denny *et al.,* 2000a; Womack *et al.,* 2000; Sankaranarayanan *et al.,* 2004b) but similar (Womack *et al.,* 2000) or higher specificity than VIA (Denny *et al.,* 2000a; Sankaranarayanan *et al.,* 2004b).

VIA is being evaluated in three randomized intervention trials in India, to assess the reduction in incidence of and mortality from cervical cancer as compared to a control group with no screening (Sankaranarayanan *et al.,* 2003a, b, 2004c, d). Early results in terms of participation, detection rates of cervical neoplasia and stage distribution of invasive cancers detected have been reported from two of these studies.

The impact of a single round of screening with VIA provided by trained nurses on cervical cancer incidence and mortality as compared to a control group with no screening is being investigated in a cluster-randomized trial in Dindigul district, south India(Sankaranarayanan *et al.,* 2003a, 2004d). Women aged 30–59 years living in 507 villages grouped into 113 clusters were randomized to VIA screening (57 clusters, 48 225 women) by nurses and to a control group (56 clusters, 30 167 women). The early results from the study are given in Table 63. All the screen-positive women were investigated with colposcopy by nurses and most had biopsies taken. The detection

rates of lesions among screened women were 5.8% for CIN 1, 0.7% for CIN 2–3 and 0.2 for invasive cancer. 71% of women with CIN 1 and 80% of those with CIN 2–3 lesions accepted cryotherapy provided by nurses and excisional treatment by mid-level clinicians. Overall, 97 and 34 incident cervical cancer cases were observed in the intervention and control arms, respectively, giving age-standardized incidence rates of 92.4 and 43.1 per 100 000 person-years, respectively, during 2000–03, the screening phase of the study. One third of the cases in the VIA group were diagnosed in stage I, while three quarters of those in the control arm were diagnosed in stage III; no stage I cases were detected in the control group. The study groups are being followed up to monitor cervical cancer incidence and mortality.

The impact of screening by VIA, cervical cytology or HPV testing (using the Hybrid Capture® 2 (HC 2) probe B assay; see Chapter 2) on cervical cancer incidence and mortality, as compared to a control group, is being investigated in a cluster-randomized controlled trial in Osmanabad district, India (Sankaranarayanan *et al.,* 2003a; 2004c). Women aged 30–59 years living in 52 clusters of 497 villages in rural Osmanabad District, were randomized to a single round of screening by either VIA (13 clusters, 34 149 women) or cytology (13 clusters, 32 136 women) or HPV testing (13 clusters, 34 515 women) or to a control group (13 clusters, 30 378 women). The early results are given in Table 64. Participation of eligible women in screening was 78.4% in the VIA group, 79.5% in the cytology group and 78.7% in the HPV group. The test-positive rates were 14.0% for VIA, 7.0% for cytology and 10.4% for HPV testing. Test-positive women were investigated with colposcopy and biopsy based on colposcopy findings. Biopsies were taken from 9.9% of screened women in the VIA, 3.5% in the cytology and 4.3% in the HPV groups.

Low-grade lesions were detected in 1068 (4.0%) screened women in the VIA group, 304 (1.2%) in the cytology group and 327 (1.2%) in the HPV group. VIA had a significantly higher rate for detection of low-grade lesions than cytology or HPV testing ($p < 0.001$). The detection rates of CIN 2–3 were 0.7% in the VIA, 1.0% in the cytology and 0.9% in the HPV groups. The detection rates of CIN 2–3 were significantly different between arms ($p < 0.001$); after adjustment for socioeconomic factors affecting detection rates, the detection rate of CIN 2–3 in the VIA arm was significantly lower than in the cytology arm (OR = 0.7, $p = 0.005$). During 2000–03, 121 women in the VIA group, 131 in the cytology group, 100 in the HPV group and 59 in the control group were diagnosed with invasive cancer. In the intervention groups, 70–74% of the cancers were screen-detected, and 48–60% were diagnosed in stage I as opposed to 24% in the control group. The preliminary findings from this study indicate satisfactory participation rates for screening, diagnosis and treatment. VIA detected significantly fewer CIN 2–3 cases than did cytology. The trial participants are being followed up to document cervical cancer incidence and mortality in the four groups.

The preliminary findings from the above trials indicate that a VIA-based screening programme is feasible, safe and acceptable for a population in rural settings, and that it results in early detection of cervical neoplasia. VIA is associated with high detection of low-grade CIN. The detection rates of CIN 2–3 lesions by VIA were similar in both the trials. While the detection rate of CIN 2–3 lesions for VIA was constant in the Dindigul trial throughout recruitment, it declined from 1.0% at the beginning to 0.5% at the end of recruit-

Table 63. Initial results after the screening phase of the cluster-randomized controlled trial of visual inspection for cervical cancer with acetic acid in Dindigul district, India

	VIA-screened group		Control group	
Number of women	48 225		30 167	
Received screening	30 577		–	
Screened positive	2939	(9.6%)[a]	–	
Number of screen-positive women who had colposcopy	2939			
Number of women who received biopsy	2777			
CIN 1	1778	(5.8%)[b]	–	
CIN 2–3	222	(0.7%)[b]		
Number with invasive cancer	97		34	
Number with stage I cancer	34	(35.0%)[c]	0	(0.0%)[c]
Number with stage II cancer	18	(18.6%)[c]	6	(17.6%)[c]
Number with stage III cancer	45	(46.4%)[c]	26	(76.5%)[c]
Number with stage IV cancer	0	(0.0%)	2	(5.9%)

[a] Percentage of screened women
[b] Indicates detection rate of CIN per 100 screened women
[c] Percentage of all cancers

From Sankaranarayanan *et al.* (2004d)

ment in the Osmanabad trial; with cytology, the rate remained constant in the latter study. A high proportion of invasive cancers were diagnosed in stage I in women screened with VIA. The ultimate efficacy of VIA in reducing cancer incidence and mortality will become clearer with follow-up for cancer incidence and mortality in these studies.

An innovative option taking advantage of the immediate availability of test results with VIA is the screen-and-treat or single-visit approach, to ensure high treatment compliance among screen-positive women. This approach is based on the following premises: studies have reported high sensitivity for VIA to identify precancerous lesions; the lack of or inadequate infrastructure and resources for diagnostic facilities such as colposcopy and histopathology in many low-resource settings; the possibility of high rates of loss to follow-up associated with multiple visits; and the potential for protection against cervical neoplasia among women who had ablation (with electrocoagulation or cryotherapy) of the ectopic cervical epithelium and the transformation zone (Vonka *et al.,* 1984).

In the screen-and-treat approach, screen-positive women without clinical evidence of invasive cancer and satisfying the criteria for ablative therapy are immediately treated by cryotherapy, without confirmatory colposcopic or histological investigations. The safety, acceptability and feasibility of such a single-visit approach combining VIA and cryotherapy was assessed in a recent study in rural Thailand (RTCOG/JHPIEGO, 2003). Trained nurses tested 5999 women with VIA and 798 (13.3%) women were VIA-positive. Overall, 756 women received cryotherapy (either immediately or postponed). No major complications were recorded following cryotherapy; only 33 women (4.4%) of treated women returned for a perceived problem. At a one-year follow-up visit, the VIA test negative rate among treated women was 94.3%.

The efficacy of the screen-and-treat approach with VIA as compared to HPV testing and treatment in reducing the frequency of high-grade CIN is being assessed in a randomized clinical trial in South Africa, which has not yet published any results.

In order to assess how screen-and-treat with VIA will perform in a routine health service setting, a large demonstration project to screen women aged 30–49 years has been launched in the St Martin province of Peru. This programme aims to cover 80 000 women in three years; no results have yet been published.

A greater proportion of cervical cancers detected by VIA were in stage I than

Table 64. Initial results after the screening phase of the cluster-randomized controlled trial of visual inspection for cervical cancer with acetic acid, cytology and HPV DNA testing in Osmanabad district, India

	Group screened with VIA	Group screened with cytology	Group screened with HPV testing	Control group
Number of women	34 149	32 136	34 515	30 378
Received screening	26 755	25 535	27 159	–
Screened positive	3731 (14.0%)[a]	1790 (7.0%)[a]	2812 (10.4%)[a]	–
Number of screen positive who had colposcopy	3682	1559	2475	–
Number of women who received biopsy	2528	828	1114	–
Low-grade lesions	1068 (4.0%)[b]	304 (1.2%)[b]	327 (1.2%)[b]	–
CIN 2 lesions	84 (0.3%)[b]	103 (0.4%)[b]	105 (0.4%)	–
CIN 3 lesions	112 (0.4%)[b]	162 (0.6%)[b]	138 (0.5%)[b]	–
Number with invasive cancer	121	131	100	59
Number with stage I cancer	58 (47.9%)[c]	67 (51.2%)[c]	60 (60.0%)[c]	14 (23.7%)[c]
Number with stage II cancer	18 (14.8%)[c]	12 (9.2%)[c]	9 (0.0%)[c]	8 (14.9%)[c]
Number with stage III cancer	30 (24.8%)[c]	29 (22.1%)[c]	10 (10.0%)[c]	30 (50.9%)[c]
Number with stage IV cancer	5 (4.1%)[c]	1 (1.0%) [c]	1 (1.0%)[c]	3 (5.6%)[c]

[a] Percentage of screened women
[b] Indicates detection rate of CIN per 100 screened women
[c] Percentage of all cancers

From Sankaranarayanan *et al.* (2004c)

among cancers occurring in unscreened controls. The long-term impact of VIA screening in reducing cervical cancer incidence remains to be established.

VILI

Visual inspection with Lugol's iodine (VILI) involves naked-eye examination of the cervix to identify mustard-yellow iodine-non-uptake areas after application of Lugol's iodine. The test characteristics of VILI provided by nurses, midwives and trained non-medical workers have been studied in a set of ten cross-sectional studies, with a similar protocol, involving 49 080 women aged 25–65 years conducted in Burkina Faso, Republic of Congo, Guinea, India, Mali and Niger (Sankaranarayanan *et al.,* 2004a). VIA was also simultaneously evaluated in all 10 studies, conventional cytology in five and HPV testing in three studies. No untoward reaction to iodine was observed. VILI had a significantly higher pooled sensitivity than VIA (91.7% versus 76.8%) but similar specificity (85.4% versus 85.5%) in detecting CIN 2–3 lesions. VILI had significantly higher pooled sensitivity than cytology (83.9% versus 45.4%) but lower specificity (82.5% versus 99.2%) in detecting histologically confirmed CIN 2–3 lesions (Sankaranarayanan *et al.,* 2004f). VILI had similar sensitivity to that of HPV testing, but lower specificity, to detect histologically confirmed CIN 3 lesions in the pooled analysis of three cross-sectional studies (Sankaranarayanan *et al.,* 2004b). There are no randomized trials evaluating the efficacy of VILI in reducing cervical cancer incidence and mortality.

Human papillomavirus testing

Almost all of the studies on HPV testing have focused on the sensitivity and specificity of the test under various conditions. No studies have prospectively investigated its impact on subsequent cancer rates, but a few have retrospectively studied the detectability of HPV in archival smears which were negative on cytology some years before a diagnosis of cancer. Issues of the persistence of HPV in high-grade lesions and the length of protection following a negative HPV test have been addressed in a few studies. A comprehensive review of the role of HPV testing in cervical screening appeared in 1999 (Cuzick *et al.,* 1999b) and several important studies have been reported since then.

Relative sensitivity for CIN 2 or 3 compared with cytology

Most of the recent screening studies have used the Hybrid Capture™ 2 test for high-risk HPV types, which is the only test now commercially available. It is clear that this test is more sensitive than cytology for CIN 2 or 3 and for CIN 3, but it also has lower specificity (Table 65). The specificity improves if testing is restricted to women over the age of 30 years.

Typically, HPV testing has a sensitivity of 95% for detecting CIN 2 or worse lesions compared with 75% for cytology at the borderline (ASCUS) or above level and 70% for cytology at the mild dyskaryosis (LSIL) level (i.e., when the cytology threshold (or cut-off) is ASCUS or worse, or when it is LSIL or worse). Thus, virtually all of the lesions detected by cytology were HPV-positive, as were an additional 25% which were negative on cytology. In women over the age of 30 years, specificity is about 93%, compared with 95% for cytology at the borderline level and 98% at the mild level. For younger women, both tests have poorer specificity. For example, in the English screening programme, for cytology at the borderline cut-off, specificity is about 89% in women aged less than 30 and 96% for older women (NHS, 2003a). For HPV, the specificity is about 85% for women aged less than 30 and 93% for older women. Where available, the studies show even greater sensitivity for detecting CIN 3 (Table 65b)

Lower sensitivities of both HPV and cytological testing are seen in developing countries. In the three large European studies (Clavel *et al.,* 2001; Petry *et al.,* 2003; Cuzick *et al.,* 2003), the sensitivity of HPV testing was uniformly high (97% or higher), whereas the sensitivity of cytology was lower and highly variable between countries.

Retrospective studies of HPV evaluation

Eleven published studies have evaluated HPV infection in stored material (Table 66). Seven used archival smears (de Roda Husman *et al.,* 1995; Walboomers *et al.,* 1995; Chua & Hjerpe, 1996; Wallin *et al.,* 1999; Carozzi *et al.,* 2000; Ylitalo *et al.,* 2000a; Zielinski *et al.,* 2001a), one used previous biopsy specimens (Konno *et al.,* 1992) and three tested for HPV antibodies in stored serum samples (Chua *et al.,* 1996; Lehtinen *et al.,* 1996; Dillner *et al.,* 1997).

Two of these studies (Konno *et al.,* 1992; de Roda Husman *et al.,* 1995) did not include controls. They looked at a total of 15 women with invasive cervical cancer and five with CIN 3, and examined smears and biopsies taken up to 10 years previously. All stored specimens tested positive for HPV16, 18 or an unknown type. Chua and Hjerpe (1996) analysing archival smears, with two matched controls per case, obtained odds ratios of 16, 11 and 18 for invasive squamous, adenocarcinoma and carcinoma *in situ* of the cervix based on 18, 12 and 58 cases, respectively. Walboomers *et al.* (1995) used as controls women from a gynaecological clinic, some of whom were being treated for CIN. They used general

Table 65. Relative sensitivity and specificity of Hybrid Capture (HC) 2 compared with cytology on biopsy in cross-sectional screening studies

(a) for CIN2+

Reference	Sensitivity		Specificity		Comments
	Cytology ≥ LSIL	HPV	Cytology < LSIL	HPV	
Clavel *et al.* (1999)	85	100	95	85	
Cuzick *et al.* (1999a)	79	95	99	95	Age ≥ 35 years
Schiffman *et al.* (2000)	78	88	94	89	Costa Rica
Ratnam *et al.* (2000)	40	90	77	51	69% HC-I, 31% HC-II
Denny *et al.* (2000a)	78	73	85	76	South Africa
Denny *et al.* (2000b)	82	72	93	86	South Africa
Schneider *et al.* (2000)	20	89	99	94	PCR with GP5+/6+
Womack *et al.* (2000)	44	81	91	62	Zimbabwe – high HIV rate
Clavel *et al.* (2001)	68	100	95	86	Conventional
	88		93		LBC
Petry *et al.* (2003)	37	98	99	95	Age ≥ 30 years
Kulasingam *et al.* (2002)	36	63	96	83	Age ≥ 30 years
Cuzick *et al.* (2003)	70	97	99	93	Age ≥ 30 years
Sankaranarayanan *et al.* (2004b)	37–72	46–81	87–98	92–95	3 centres with variable results
Salmeron *et al.* (2003)	59	93	98	92	Mexico. Cut-point for cytology was ≥ ASCUS
Nieminen *et al.* (2004)	83	98	94	78	Hospital population

(b) for CIN3+

Reference	Sensitivity		Specificity		Comments
	Cytology ≥ LSIL	HPV	Cytology < LSIL	HPV	
Cuzick *et al.* (1999a)	79	100	99	95	Age ≥ 35 years
Ferreccio *et al.* (2003)	63	85	94	88	Conventional
	86		88		LBC
Salmeron *et al.* (2003)	60	94	98	90	Mexico. Cut-point for cytology was ≥ ASCUS
Petry *et al.* (2003)	40	97	99	95	Age ≥ 30 years
Cuzick *et al.* (2003)	77	98	99	93	Age ≥ 30 years
Kulasingam *et al.* (2002)	57	91	90	73	All ages
Sankaranarayanan *et al.* (2004b)	80	77–89	95	92–95	3 centres with variable results
Nieminen *et al.* (2004)	86	100	93	78	Hospital population

primers to probe archival smears and, consistent with other studies from this group, found a very strong association with high-risk HPV types. Sixteen of the 17 women with invasive carcinoma had HPV in archival smears compared with seven of the 50 controls, giving an odds ratio of 49. Further, all nine cases with two archival smears had the same type of HPV detected on both. The smears were taken between two months and six years before cancer diagnosis (median 1 year). By design, all smears were originally classed as normal. On reanalysis, four of the 26 archival smears from the cases were deemed inadequate, and the rest showed severe dyskaryosis or worse. Wallin *et al.* (1999) compared archival smears, all of which had normal cytology, from 118 women with subsequent cervical cancer with those from 118 controls. The average duration between smears and cancer was 5.6 years (range, 0–26 years). HPV was detected in 30% of the cases but only 3% of the controls (OR = 16.4; 95% CI 4.4–75). The PCR in this study used both MY09/MY11 and GP5/6 consensus primers.

Ylitalo *et al.* (2000a) reviewed all previous smears in 484 cases of carcinoma *in situ* and 619 matched controls

in Uppsala, Sweden. Smears were available for up to 26 years before diagnosis. Only HPV16 was tested for. The case smears from 16–18 years before diagnosis were HPV-positive for about 10% of women, which was similar to controls, but the proportion rose linearly to 56% 2.3 years before diagnosis, which was highly significant. A positive HPV16 result on the two last smears before diagnosis was associated with an odds ratio of 31.2 (95% CI 10.6–91.8) for carcinoma *in situ.* The mean time from HPV positivity to diagnosis was estimated to be between 7 and 12 years.

Carozzi *et al.* (2000) assessed archival smears classified as normal from 79 cases of CIN 2 or worse and matched controls. They used a consensus system which detected HPV types 16, 18, 31, 33, 52 and 58. An odds ratio of 64 (95% CI 31–133) for CIN 2 or worse was found for HPV-positivity in all smears, which rose to 103 (95% CI 43–251) in smears taken less than four years before the last cytological test; overall 77% of case smears were HPV-positive compared with 5.1% of control smears.

Zielinski *et al.* (2001a) examined the last normal archival smear from 57 women who later developed cancer and 114 control women, using GP5+/6+ primers. HPV was detected in 65% (37) of the smears from cases compared with 6% (7) of the control smears (OR = 28; 95% CI 11–72). Positivity was only slightly higher in smears taken within three years of diagnosis (76%) than in smears taken 4–20 years previously (65%).

Three studies (Chua *et al.,* 1996; Lehtinen *et al.,* 1996; Dillner *et al.,* 1997) looked for HPV16 (or HPV16 and 18) antibodies in stored serum samples, using a nested case–control design. All three found increased risk of cancer or carcinoma *in situ* in women with prior seropositivity to HPV16. The odds ratios associated with HPV antibodies in these studies ranged from 3 to 13. A longer lag time from sampling of sera to diagnosis was associated with greater relative risk. Chua *et al.* (1996) estimated the progression rates to cancer or carcinoma *in situ* in women of different ages with and without HPV16 antibodies. The incidence of cancer or carcinoma *in situ* decreased with age, as did the relative risk associated with HPV16 antibodies, whereas seropositivity increased with age in the controls. A possible explanation of this finding is that CIN is associated with an active HPV infection and women who developed antibodies some years earlier no longer necessarily carry the virus. The study also looked at antibodies for HPV18 and 33, but these were not significantly associated with disease. The largest of the studies (Dillner *et al.,* 1997) combined cohorts from Finland, Norway and Sweden and included 182 cases of invasive carcinoma. Overall, the relative risks were 2.7 for HPV16 antibodies and 2.2 for HPV16, 18 or 33 antibodies. The relative risk associated with HPV16 antibodies increased to 3.9 in those women with a lag time of over five years. The third study measuring antibodies (Lehtinen *et al.,* 1996) included 27 cases of invasive cancer and 25 of carcinoma *in situ.* Overall, the odds ratio for risk of developing cervical carcinoma according to the presence of HPV antibodies was 13.2. It was greater for invasive cancer (OR = infinite; 95% CI 2.0–infinite) than for carcinoma *in situ* (OR = 6.0; 95% CI 1.2–29.7) and for lag times of over five years (OR = 18; 95% CI 2.3–142) compared with under five years (OR = 8.6; 95% CI 1.0–75). [Due to the lack of sensitivity of serological tests (50% at best for some high-risk HPV types), odds ratios from these studies are inevitably much lower than for direct DNA testing in cervical specimens.]

Duration of protection

The higher relative sensitivity of HPV testing for CIN 2 or 3 compared with cytology suggests that it might be safe to lengthen screening intervals if HPV testing were used. Two studies of this issue have been reported (Bory *et al.,* 2002; Sherman *et al.,* 2003a) and others are in progress. Bory *et al.* (2002) found that among 2432 women who were negative for high-risk HPV, only two (0.08%) developed high-grade CIN after a median follow-up of 27 months. Both cases were HPV-positive at the time of diagnosis (after 18 and 24 months). This was compared with 21.2% developing CIN 2 or 3 among women who were initially HPV-positive but cytologically normal.

Sherman *et al.* (2003a) reported a 10-year follow-up of 20 810 women screened by cytology and HPV DNA testing at Kaiser Permanente in Portland, Oregon. Cervical lavage specimens were used for HPV testing. A total of 171 women were diagnosed with CIN 3 or cancer on follow-up. HPV positivity was more sensitive for detecting CIN 3 on follow-up than cytology (89 versus 58 in the first 45 months and 110 cases versus 59 cases overall). Conversely, there were fewer cases in HPV-negative women (29 versus 60 in the first 45 months and 61 versus 112 overall). Detection rates were similar between tests in the first nine months (15 cases in HPV-positive women, 15 cases in HPV-negative women) but HPV negativity was much more protective than cytology after that (14 versus 45 cases in months 10–45; 46 versus 97 in months 10–112).

In addition, modelling studies have suggested that the use of HPV testing could safely allow the screening interval to be lengthened. Goldie *et al.* (2004) modelled the US data and recommended that women whose results are negative by both HPV DNA testing and cytology should not be

Table 66. Studies retrospectively analysing stored samples for HPV DNA or antibodies

Reference	Assay	Follow-up	Setting	Age (years)	Case	Control	Material
Konno *et al.* (1992)	In situ hybridization PCR of negatives	< 10 y	Japan		CIN 3 (5) Microinvasive (2) Invasive carcinoma (1) All with HPV16/18 on hysterectomy section		Hysterectomy sections
	HPV 16 found in all previous biopsy specimens						
De Roda Husman *et al.* (1995)	General primer PCR	2–9 y; mean, 5.8 y	Netherlands Screening programme		Cervical cancer (12)	None	Archival smears
	HPV found in archival smears of all 12 cases. Same type found in smear and biopsy of tumour.						
Walboomers *et al.* (1995)	GP5/6	2 mo–6 y; median, 1 y	Netherlands 3-yearly screening programme		17 cancers with normal archival smears	50 controls from gynaecology clinic including women with CIN	Archival smears
	All nine cases with two archival smears had the same viral type on both On rescreening 26 archival smears from the cases, four were inadequate and the rest (22) were all severe dyskaryosis or worse						
Chua & Hjerpe (1996)	PCR (nested): (i) MY09/11 (ii) GP5+/6+	2–7 y mean 3 y	Sweden	17–68	Adenocarcinoma (12) Squamous-cell (18) Carcinoma *in situ* (58)	No carcinoma *in situ* for 5 y post-smear (but some history of abnormaly) (age matched)	Archival smears
Chua *et al.* (1996)	Serology L1 and L2 capsids for HPV16 antibodies	< 6.5 y; mean 3 y	Sweden	27–61	CIN 2/3 (41) CIN 1 (10) CIN not otherwise specified (23)	Population-based Match on age, date of blood Probability of CIN within 3 months given HPV seropositivity	Sera

	Case		Control	
	No.	HPV + (%)	No.	HPV+ (%)
Carcinoma *in situ*	58	71	58	12
Invasive squamous-cell	18	67	18	11
Adenocarcinoma	12	58	12	8

	Archival HPV_{hr}		Total	
	+	-	Women	Smears
Case	16	1	17	26
Control	17	43	50	88

Odds ratio, 49

	No.	Seropositive (%)	Age (years)	+	–
CIN 2–3	41	37	25–34	0.034	0.005
CIN 1	10	20	35–44	0.016	0.006
CIN not otherwise specified	23	43	45–64	0.002	0.002
Controls	148	16			

Table 66 (contd)

Reference	Assay	Follow-up	Setting	Age (years)	Case	Control	Material
Lehitinen *et al.* (1996)	HPV16 L1 and L2 capsid antibodies	0.7–22.8 y; mean, 10 y	Finland 18 814 women flagged to cancer registry	Mean, 39 at baseline, 49 at diagnosis	27 invasive carcinomas 25 carcinomas *in situ*	143 individually matched	Stored sera
Dillner *et al.* (1997)	L1 and L2 capsids to HPV16, 18 and 33 antibodies	14% under 1 y; 33% over 5 y	Finland Norway and Sweden population-based serum	< 40 (45%) > 45 (26%)	182 invasive carcinomas	~ 3 matched controls per case (total, 538) No. HPV 16+ HPV 16, 18, 33+ Case 182 16% 37% Control 538 7% 19% Relative risk 2.7 2.2 For HPV16, relative risk increased with increasing lag time: 3.9 (95% CI 1.6–9.6) over 5 years	Stored sera
Wallin *et al.* (1999)	MY09/11 and GP5/6	0–26 y, median, 5.6 y	Sweden	19–74, mean 44	Cancers with prior negative smears (118)	118 HPV + HPV– Case 35 83 Control 3 115 OR 16.4 (95% CI 4.4–75)	Archival smears (and biopsies)
Ylitalo *et al.* (2000a)	PCR HPV 16	0.26 y; median, 8 y	Sweden	20–70 median; 35	484 CIS (2228 smears)	619 matched controls (1806 smears) HPV+ HPV– Case 130 353 Control 36 581 First smear: OR 13 y prior, 3.6 (95% CI 1.2–11) 1 y prior, 11.1 (95% CI 5.5–22.2) Last 2 smears postive: OR 31.2 (95% CI 10.6–91.8) Average time to CIS, 7–12 years	Archival smears
Carozzi *et al.* (2000)	PCR (consensus primers HR HPV)	0–6 y; mean, 2.5 y	Florence, Italy	25–64	92 total smears 15 CIN 2 59 CIN 3 5 invasive	3 controls per case; total number of control smears, 332 All smears cytologically negative HPV+ HPV– Case 71 21 Control 17 315 OR 64 (95% CI 31–133) For slides with latency < 4 years: OR 103 (95% CI 43–251)	Archival smears
Zielinski *et al.* (2001a)	GP5+/6+ PCR	0–18 y	Netherlands 3-yearly screening programme		57 cancers with normal smear	114 age matched controls HPV+ HPV– Total Case 37 20 57 Control 7 107 114 OR 28 (95% CI 11–72)	Archival smears

CIS, carcinoma *in situ*; HR, high-risk; OR, odds ratio. Modifed and updated from Cuzick *et al.* (1999a)

re-screened before three years. Women with normal cytological results, but who are positive for high-risk HPV DNA, are at relatively low risk of having high-grade cervical neoplasia, and colposcopy should not be performed routinely in this setting. Instead, HPV DNA testing along with cervical cytology should be repeated in these women after 6–12 months. If test results of either are abnormal, colposcopy should then be performed. As a result of these recommendations, the American Cancer Society and the American College of Obstetricians and Gynecol-ogists have approved the extension of screening intervals from annually to three-yearly in women aged over 30 years, when HPV is added to cytology. [The Working Group noted that modelling results are no substitute for direct evaluation of this question.]

Cuzick *et al.* (2003) showed that women who originally had negative or borderline cytology but were subsequently found to have CIN 2 or worse lesions remained HPV-positive. [The Working Group noted the importance of persistence in determining the likelihood of underlying high-grade disease, but that currently this could only be assessed by two tests less than one year apart.]

Other screening methods

Liquid-based cytology systems

Introduction

Several literature reviews and meta-analyses have been published comparing the relative sensitivity and specificity of liquid-based cytology systems with those of conventional cervical testing (see Chapter 2).

The comprehensive systematic review and meta-analysis by Arbyn *et al.* (2004a) includes split-sample and more recent direct-to-vial studies. The authors noted that early studies often yielded favourable results for LBC when comparing test positivity rates for low-grade abnormalities on cytology only, whereas in studies using detection rates for biopsy-confirmed CIN 2 or 3, no significant differences between conventional and liquid-based cytology were found.

This meta-analysis showed that test positivity was higher in direct-to-vial studies than in traditional testing, suggesting a bias in previous split-sample studies to the disadvantage of LBC. In direct-to-vial studies, more LSIL and HSIL lesions were identified in both ThinPrep and SurePath LBC systems (Table 67) (Arbyn *et al.,* 2004a). Identification of equivocal samples (ASCUS) was similar. The authors were concerned that increased identification of cytological abnormalities (ratios >1) alone provides insufficient evidence for improved sensitivity of LBC in a screening programme and that verification with a valid gold standard is needed. Figures 52 and 53 show the test positivity ratios from studies available to the Working Group.

Only a few studies verified all cases (test positives and negatives) with a gold standard, allowing evaluation of sensitivity and specificy of both LBC and conventional cytology without verification bias (see Table 68). All these studies used the split sample design and none showed a statistically significant difference. In fact, the relative sensitivity was somewhat lower for LBC than for conventional testing.

Consideration of the positive predictive value (PPV) would allow determination of whether higher cytological positivity rates with LBC are due to an increase in false-positive tests. As positive predictive values for presence of CIN 2+ pooled from studies with at least an 80% gold standard verification of test-positives by colposcopically directed biopsy did not significantly differ from that of conventional cytology, this was probably true.

Evidence of efficacy from LBC in routine cervical screening programmes

Experience with LBC as the routine test in screening programmes is relatively recent and there is no long-term follow-up in terms of effects on incidence and mortality in the populations served. Thus the evaluation of efficacy has been made in terms of short-term surrogate markers such as relative sensitivity and specificity of LBC compared with those of high-quality conventional cytology. Short-term evaluations of LBC as the routine test in cervical screening in England, Scotland and Canada are described below.

A systematic review of the literature and modelling of cost-effectiveness, commissioned in the United Kingdom by the National Institute for Clinical Excellence (NICE) (Payne *et al.,* 2000), concluded that, despite the lack of published studies providing direct evidence regarding cost-effectiveness of LBC for cervical screening, it was likely that LBC would reduce the number of inadequate samples, reduce the number of false negative results and decrease the time required for examination of specimens by cytologists. NICE immediately commissioned a full cost-effectiveness trial of LBC in a low-prevalence population for routine screening. LBC was introduced, after a learning transition period of 3 to 6 months in three selected laboratories, as part of a 12-month pilot project aiming at 100 000 routine screening tests. In two laboratories, the ThinPrep system was introduced and SurePath in another. The cytological results of the first six months of the pilot period were compared with the four previous years where exclusively conventional cytology was used (Moss *et al.,* 2003). It was noted that different sampling devices were used before (wooden Aylebury spatula) and after (Cervex broom) the introduction of LBC, but there is published evidence that there is no statistical difference in sensitivity

Table 67. Meta-analysis: pooled ratios of test positivity rates for liquid-based cytology (direct-to-vial studies) versus conventional testing

	ThinPrep			SurePath		
Test threshold	**Pooled estimate**	**95% CI**	**No. of studies**	**Pooled estimate**	**95% CI**	**No. of studies**
HSIL+	1.72	1.42–2.08	21	1.47	1.14–1.89	7
LSIL+	1.74	1.47–2.06	21	1.52	1.24–1.86	7
LSIL	1.80	1.52–2.12	21	1.54	1.25–1.90	7
ASC+	1.23	1.07–1.40	19	1.19	0.96–1.46	7
ASC	0.95	0.84–1.09	19	0.93	0.67–1.31	7

From Arbyn *et al.* (2004a)

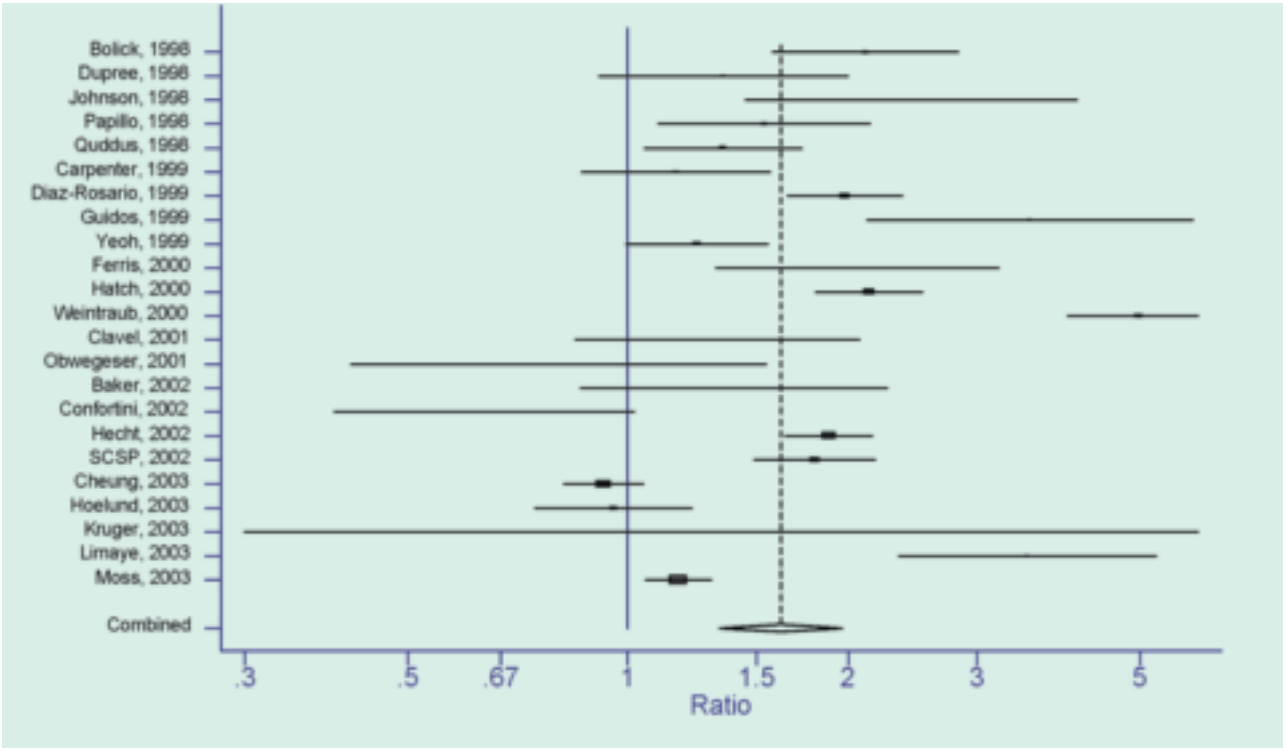

Figure 52 Ratios of test positivity for HSIL+ in direct-to-vial studies with ThinPrep LBC, compared with conventional Pap cytology
< > indicates the combined effect estimated by random-effects model

between these two devices (Buntinx & Brouwers, 1996).

The proportions of inadequate tests before and after conversion to LBC are shown in Table 69. The evaluation showed an 80% reduction in the rate of inadequate samples in all laboratories and in all age groups after introduction of LBC. The rate was lowest in the SurePath laboratory. The quality of conventional smears increased with age, but no age differential was observed with LBC. The inadequate rate in English laboratories is higher than in many other countries (average 9%).

The relative changes in the identification of cytological abnormalities with LBC versus conventional testing by

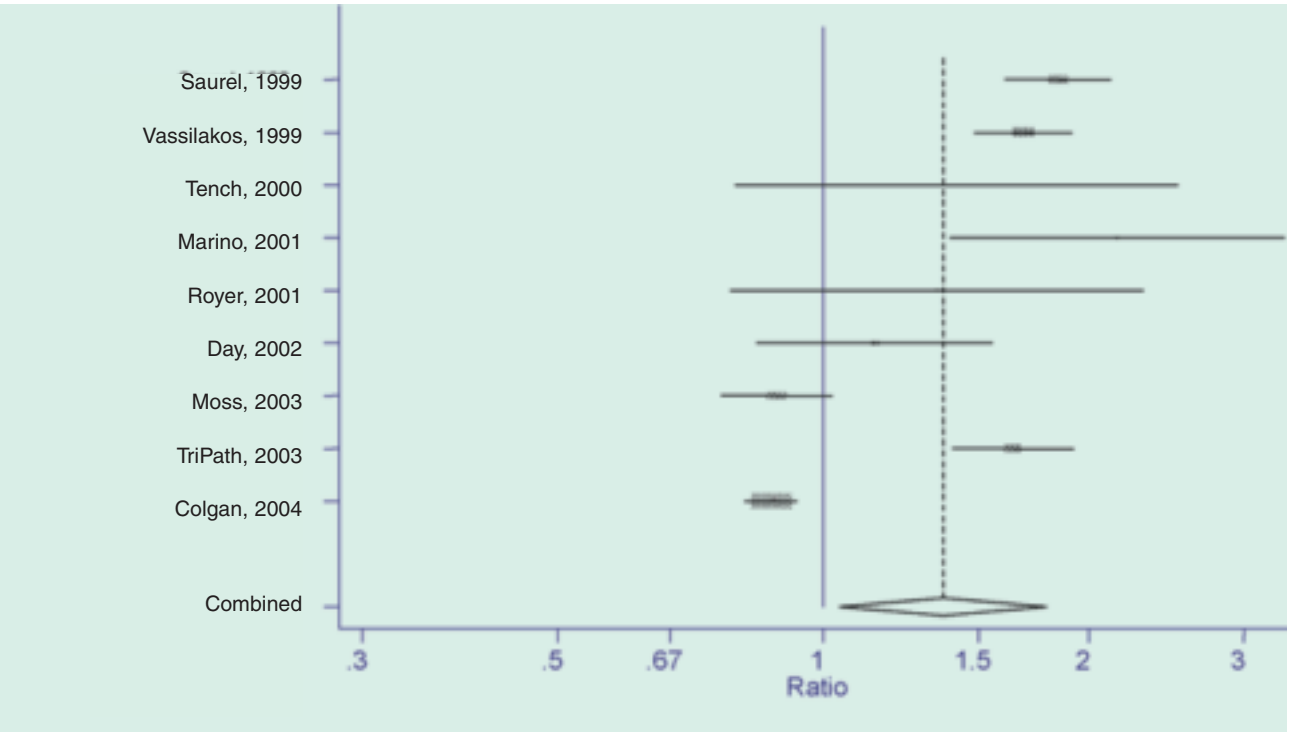

Figure 53 Ratios of test positivity for HSIL+ in direct-to-vial studies with AutoCyte Surepath LBC, compared with conventional Pap cytology
< > indicates the combined effect estimated by random-effects model

age group and by laboratory are shown in Table 70. There was no significant increase in the rates of HSIL when averaged across the three sites. However, in the ThinPrep laboratories, significantly more SIL and HSIL lesions were found, and one of them (lab C) also found more borderline lesions. In the laboratory where SurePath was used, less HSIL and borderline lesions were detected. The reason for this difference is not known. The increased identification of LSIL or worse lesions by SurePath was concentrated in the 20–34-year age group.

The reduction in inadequate rate should lead to fewer tests being performed, with a resulting decrease in workload for laboratories and primary care as well as recall systems. Referrals to colposcopy are likely to be affected only if the overall reporting of high-grade lesions increases. Comparing the running costs of LBC with those of conventional testing was complex since they utilize different amounts of laboratory resources.

There was some debate in the United Kingdom on the extent to which published differences between LBC and conventional cytology represented a true improvement (Herbert & Johnson, 2001, Moseley & Paget, 2002); a more recent article (Coste *et al.,* 2003) contradicted the National Health Service pilot findings.

In another LBC pilot study, conducted by the Scottish Cervical Screening Programme (SCSP, 2002), four regional groups of smear-takers were randomized into two groups, collecting respectively conventional cervical smears and ThinPrep preparations from women attending for routine or follow-up screening tests. ThinPrep alone was chosen as the LBC system studies, since Cytyc had a more established infrastructure to support the laboratories at the time and the numbers in the study (30 000 LBC) were insufficient for evaluation of LBC with two systems. Smears and ThinPrep LBC vials were sent to four selected laboratories where cytotechnologists had received training in the interpretation of ThinPrep slides. The LBC samples were collected using the plastic cervix broom, while conventional smears continued to be collected by the wooden Aylesbury spatula. The results, summarized in Table 71, showed a sharp reduction in the rate of unsatis-

Table 68. Ratio of sensitivity and specificity for CIN 2+ of two liquid-based cytology systems (LBC) relative to the conventional Pap smear (CP), pooled from studies with complete verification by colposcopy and/or biopsy

	ThinPrep			SurePath		
		95% CI	No. of studies		95% CI	No. of studies
Ratio of sensitivities (LBC/CP)	0.95	0.88–1.03	2	0.95	0.81–1.11	1
Ratio of specificities (LBC/CP)	1.08	0.90–1.30	2	0.94	0.87–1.01	1

Test threshold HSIL+
Adapted from Arbyn *et al.* (2004a)

factory samples and a significant improvement in the identification of high-grade lesions (between 3 and 9 women per 1000 tested). Reduced workload and increased productivity were also demonstrated in laboratories.

In Ontario, Canada, SurePath was adopted for routine cervical screening in large screening laboratories in 2001, after training of large numbers of cytotechnologists. Almost one million routine cervical screening results were reviewed using the Ontario Provincial database. The results for 445 011 SurePath samples reported between January and June 2002 were compared with 445 225 conventional smear results from the same period in 2001 (Colgan *et al.,* 2004; McLachlin *et al.,* 2004). All slides had been screened manually. The SurePath cases showed 21% higher reporting of LSIL+ but a 15% decrease in HSIL detection. Assessment with the addition of colposcopic diagnostic rates is in progress to determine the relative sensitivity and specificity of the SurePath technology as a routine screening test.

Table 69. Prevalence of inadequate specimens by preparation system

Preparation system	% inadequate	95% CI
Conventional Pap smear	9.7	9.4–10.0
ThinPrep LBC	2.0	1.8–2.2
SurePath LBC	0.9	0.8–1.1

Computed over four years of conventional cytology and the first six months' use of LBC
From Arbyn *et al.* (2004a), adapted from Moss *et al.* (2003)

Table 70. Relative change in test positivity for abnormal results, SIL and HSIL in liquid-based cytology in comparison with conventional cytology, by laboratory and by age (crude and weighted)

	Abnormal (≥ borderline)		SIL (≥ mild dyskaryosis)		HSIL (≥ moderate dyskaryosis)	
Laboratory	**RR**	**95% CI**	**RR**	**95% CI**	**RR**	**95% CI**
A	0.99	0.94–1.04	1.25	1.19–1.31	1.10	0.98–1.24
B	0.94	0.90–0.99	1.00	0.93–1.08	0.85	0.75–0.96
C	1.37	1.30–1.44	1.73	1.60–1.86	1.55	1.36–1.76
Age	**RR**	**95% CI**	**RR**	**95% CI**	**RR**	**95% CI**
20–34	1.15	1.11–1.19	1.25	1.19–1.31	1.18	1.08–1.28
35–49	0.96	0.91–1.02	1.03	0.94–1.13	1.01	0.87–1.16
50–64	0.93	0.85–1.01	1.03	0.87–1.23	0.75	0.55–1.01
20–64 (MH)	1.06	1.03–1.09	1.18	1.13–1.23	1.10	1.02–1.18

MH: overall Mantel–Haenzel-adjusted relative risk
Laboratories A and C used Thin Prep; B used Sure Path
From Arbyn *et al.* (2004a), adapted from Moss *et al.* (2003)

Table 71. Proportion of inadequate specimens and cytological abnormalities and ratios observed in two randomized groups of women from four Scottish areas

	Proportion		Ratio[a]	
Cytological category	ThinPrep	Conventional	(ThinPrep/ conventional)	95% CI
Unsatisfactory	1.86%	8.00%	0.23	0.20–0.26
Borderline	3.67%	4.35%	0.84	0.76–0.94
Mild dyskaryosis	2.10%	1.09%	1.93	1.60–2.32
Moderate dyskaryosis	0.97%	0.48%	2.01	1.52–2.66
Severe dyskaryosis	1.09%	0.59%	1.84	1.42–2.38

[a] Confidence intervals are approximate. The SCSP report stated that in half of the total of 30 288 women a conventional Pap smear was taken and in the other half a ThinPrep, so the calculation of confidence intervals was made assuming exactly 15 144 individuals in each group.
From Arbyn *et al.* (2004a), adapted from SCSP (2002)

Use of liquid-based cytology systems in cervical screening programmes

In May 1996, the ThinPrep T2000 processor was approved by the US Food and Drug Administration (FDA) for use in cervical screening on the basis of lower inadequacy rates and higher identification of LSIL and HSIL in comparison with conventional cytology and this was extended to the fully automated ThinPrep T3000 processor in May 2000. The SurePath approval by FDA in June 1999 indicated improved specimen quality and equivalent identification of cytological abnormalities to conventional cytology and in May 2003 a claim for increased HSIL identification was approved by the FDA. These two LBC systems now account for over 80% of cervical screening tests in the USA.

The Scottish Department of Health decided in 2002 that LBC should be implemented as the routine screening test throughout Scotland, following the Scottish LBC pilot project (see above). Scottish laboratories opted to use the ThinPrep system and all laboratories are now fully converted. In England and Wales, the NICE (2003) recommended to the National Health Service that LBC be introduced as the primary means of processing samples in its cervical screening programme. Health technology assessments in several other countries have not yet led to approval of LBC, although it is already widely used in the private sector in other parts of the world.

Other liquid-based systems are marketed, but there is little or no evidence in peer-reviewed literature for their efficacy (Johnson *et al.,* 2000; Bergeron & Fagnani, 2003; Alves *et al.,* 2004).

Automation-assisted devices

Automation-assisted screening is aimed at enhancing the performance of manual microscopic screening by excluding some of the normal slides from manual screening or by relocating the most suspicious cells down the microscope or collecting images into a gallery for review on computer screens. These technologies have the potential to decrease the fatigue of the user, allow 50% more slides to be reviewed per day, decrease the screening false negative rate due to human error, with appropriate decision support, identify morphological features that are not apparent in routine human review and, in particular, identifying small numbers of small abnormal cells, known to be very difficult to find in conventional screening.

Most of the automated scanning devices are capable of processing either conventional or liquid-based smears, potentially allowing their use in different kinds of screening programmes. Since LBC systems deposit cells onto a thin layer on a microscope slide, problems of cell overlap and obscuring debris are mitigated and thus LBC facilitates the performance of computer-assisted imaging (see Chapter 2)

Technological developments are very rapid in this area and several new approaches are emerging. A development of the FDA-approved AutoPap® system is the FocalPoint™ system, which is designed for use with SurePath (LBC) slides. The few published studies on the FocalPoint system suggest a performance equivalent to that of AutoPap (Cengel *et al.,* 2003, Parker *et al.,* 2004).

Another system that has become commercially available recently is the Cytyc ThinPrep Imager. Again, only a few studies on the performance of this system have been reported, but these show statistically significant improvement in sensitivity of the Imager review method over the conventional manual review for HSIL+. Specimen adequacy can be determined with the Imager

review method. Cytotechnicians were able to double their daily work output while maintaining the same quality (Biscotti *et al.,* 2003; McKee *et al.,* 2003).

A few randomized prospective studies using the obsolete Papnet system (Nieminen *et al.,* 2003) (see Chapter 2) show that automation-assisted screening is feasible in routine primary screening and that it performs in organized screening programmes at least as well as conventional manual microscopy. It is suggested that automation-assisted screening would not improve the outcome of an optimal cervical cytology service.

Efficacy of screening among HIV-positive women

There are no data specifically on the efficacy of screening in HIV-positive women. In considering this group of women, three aspects of screening need to be taken into account.

- The accuracy of cytology as a screening test in HIV-positive versus HIV-negative women;
- The natural history of preinvasive disease of the cervix in HIV-positive versus HIV-negative women;
- The impact of anti-retroviral treatment on the natural history of preinvasive lesions of the cervix.

Accuracy of cytology in HIV-positive women

Several studies have addressed the accuracy of the cytological test in HIV-positive women. Maiman *et al.* (1991) reported that the false negative rate of cytology was significantly higher in HIV-infected women than in HIV-uninfected women and recommended that routine colposcopy and histological sampling be performed in these women. A follow-up study confirmed the earlier findings (Maiman *et al.,* 1998), based on an evaluation of 285 HIV-infected and 685 HIV-negative women, among whom 255 of the HIV-infected women underwent colposcopy and biopsy. Abnormal cytology detected 62% of all biopsy-confirmed CIN and 83% of all high-grade CIN in the HIV-infected women, but 38% of all CIN diagnosed would have been missed had colposcopy not been performed. The false negative rate of cytology, however, was not significantly different from that recorded in the HIV-negative group, reflecting the limitations of cervical cytology as a screening test rather than a poorer performance of cytology in HIV-infected women.

Fink *et al.* (1994), in a cross-sectional analysis of 51 HIV-positive women examined by cytology, colposcopy and biopsy, showed that there was good reproducibility of cytological results in HIV-infected women, but a high false negative rate for cytology compared to colposcopy. For instance, of 29 women who had normal cytological findings, 24% had CIN on biopsy. The authors recommended that this high false negative rate of cytology and the high prevalence of CIN in HIV-infected women warranted the inclusion of routine colposcopy for all HIV-infected women as a part of primary screening.

Korn *et al.* (1994) tested 52 HIV-positive women by cytology, colposcopy and histology, and a group of 85 women who self-reported HIV-negative status. The prevalence of CIN was 50% in the HIV-infected group and the sensitivity of cytology was 63% with a specificity of 84%. The performance of cytology in the control group was similar, however, and it was concluded that the accuracy of cytology was not significantly lower in HIV-positive women. The authors noted a high rate of loss to follow-up in HIV-positive women, as well as a significant incidence of concurrent lower genital tract pathology, which may justify initial colposcopic evaluation in this group.

Spinillo *et al.* (1998) reported on a cross-sectional study of 241 HIV-positive women and 991 controls (404 known HIV-negative and 587 of unknown HIV status). Among HIV-positive women, the sensitivity of cytology was 73% and the specificity 97%. The corresponding figures for the control group were 84% and 99% and the differences were not statistically significant. However, the negative predictive value of cytology was significantly lower in the HIV-positive group. The authors suggested more frequent screening of HIV-positive women rather than primary screening with cytology, colposcopy and biopsy.

Goodman *et al.* (2000) undertook a prospective study of cytology and concurrent colposcopically directed biopsies in HIV-positive and -negative women to determine the accuracy of cytology in the two groups. Among 82 HIV-positive women, the prevalence of CIN was 37%, compared with 17% in the HIV-negative group; the false negative rates of cytology in the HIV-positive and -negative women were 37% and 21% respectively, if ASCUS findings on cytology were included among the negative results. These false negative rates fell to 10% and 14%, respectively, if ASCUS was counted as a positive result. The authors concluded that ASCUS diagnosis comprised the majority of false negative calls in HIV-positive women and they too recommended an initial screening colposcopy to detect cases of CIN missed by cytology. Thereafter, they recommended six-monthly cytological screening.

Boardman *et al.* (1994), who compared 41 HIV-positive women with 228 HIV-negative and 409 women with unknown HIV status, also found no difference in the performance of cytological testing between HIV-positive women and women of negative or unknown HIV status. With HIV-negative women as the reference group, the

relative risk of cytology–histology discrepancy was 1.1 for HIV-positive women compared with 1.5 for women whose HIV status was unknown.

Adachi *et al*. (1993) performed colposcopy on 48 women with a cytological diagnosis of SIL, of whom 95% had colposcopic or histological findings that were no more severe than the cytological result. They concluded that the positive predictive value was high. Del Priore *et al*. (1995) also reported a high PPV of abnormal cytology to predict disease in a series of 52 HIV-positive women. The PPV of cytology was 96% for HIV-positive women versus 78% for HIV-negative women. The sensitivity of cytology among HIV-positive women was only 57%, with a specificity of 92%. The authors concluded, however, that prediction of the presence and degree of an intraepithelial lesion by abnormal cytology was no worse in HIV-positive than in HIV-negative women.

Where colposcopy services are readily available and accessible, an initial colposcopy may be warranted, particularly in women at greatest risk of disease, such as those with significant immune compromise or borderline abnormal cytology. Where such a service is lacking, more frequent cytological surveillance, e.g., six-monthly, has been recommended, although with little supporting evidence.

Natural history of preinvasive cervical disease in HIV-positive versus HIV-negative women

It is now clear that women infected with HIV have a higher prevalence of infection with HPV and are more likely to develop persistent infection with multiple types of HPV, as well as having a higher incidence and prevalence of preinvasive lesions of cervix, possibly a more rapid progression to cervical cancer and a higher incidence of cervical cancer (Schafer *et al*., 1991; Klein *et al*., 1994; Wright *et al*., 1994; Sun *et al.,* 1997; Palefsky *et al*., 1999; Ellerbrock *et al*., 2000).

In 1992, the Centers for Disease Control and Prevention (CDC) included cervical cancer as an AIDS-defining disease, on the basis of extrapolation of data on the higher frequency of CIN in HIV-positive than in HIV-negative women (Centers for Disease Control and Prevention, 1992). A number of studies have supported this view and shown that HIV-positive women generally present with more advanced lesions and have a poorer prognosis than HIV-negative women (Maiman *et al*., 1993, 1997).

In a large population-based study by the AIDS-Cancer Match Registry Study Group in the USA, Frisch *et al*. (2000) found that, compared with the general population, HIV-infected individuals were at considerably increased risk for all types of anogenital HPV-associated cancers and their precursor lesions. This elevated risk spannned the decade from five years before the onset to five years after the diagnosis of AIDS. The relative risks of in situ cancer of the cervix (N = 722) (4.6; 95% CI 4.3–5.0) and invasive cancer (N = 44) (5.4; 95% CI 3.9–7.2) were similar in women with HIV infection or AIDS.

In the population-based Cancer and AIDS Registry Linkage Study in Italy (Dal Maso *et al*., 2003), women with HIV infection or AIDS had a relative risk of invasive cervical cancer (N = 18) of 21.8 (95% CI 12.9–34.6).

Lomalisa *et al.* (2000) presented data from South Africa on 60 HIV-seropositive and 776 HIV-seronegative women with newly diagnosed invasive cervical cancer. HIV-positive women presented with cervical cancer almost 10 years earlier than HIV-negative women (mean age 44 years versus 53 years), although the stage distribution was not different in the two groups, a finding of particular importance for screening programmes. In addition, severely immunocompromised women (e.g., CD4+ counts below 200 cells/μL) were significantly more likely to have advanced-stage disease at initial diagnosis than HIV-negative women.

A study in Senegal (Hawes *et al*., 2003) provided support for these conclusions, showing that HIV infection was associated with increased rates of cervical infection with high-risk types of HPV and that high-grade cervical cancer precursors and invasive cervical cancer were significantly more common in HIV-positive than in HIV-negative women (OR = 8.0; 95% CI 2.0–31.5). The degree of cervical abnormality was related to increased HIV viral load and increased immunosuppression, as expressed by low CD4+ cell counts.

Sitas *et al*. (2000) identified 167 cases of invasive cervical cancer among HIV-positive South African women versus 1323 among HIV-negative women (OR = 1.6; 95% CI 1.1–2.3), suggesting an increased cervical cancer risk among HIV-positive women in a country with high rates of cervical cancer and HIV seropositivity.

Numerous studies have indicated an increased prevalence of preinvasive lesions of the cervix in HIV-positive women. Mandelblatt *et al.* (1992) reviewed 21 studies from 1986 to 1990, and found five studies with sufficient data and a comparison group. All five studies showed a significant association between HIV infection and CIN, with an odds ratio for HIV-positive women of 4.9 (95% CI 3.0–8.2) compared with HIV-negative women.

Wright *et al*. (1994) conducted a cross-sectional study of 398 HIV-positive and 357 HIV-negative women; 20% of HIV-positive women had colposcopically confirmed CIN, compared with 4% of the HIV-negative women. The sensitivity and specificity of cytological testing did not differ significantly between the two groups and the authors concluded that cytology was

an effective screening test in HIV-positive women. In addition, by multiple logistic regression analysis, CIN was found to be associated with HPV infection (OR = 9.8), HIV infection (OR = 3.5), CD4+ T-lymphocyte count of less than 200 cells/μL (OR = 2.7) and age greater than 34 years (OR = 2.0).

Delmas *et al.* (2000) reported on the effect of immunodeficiency on the prevalence and incidence of SIL in 485 HIV-positive women. Compared with women with CD4+ counts of over 500 cells/μL, women with counts below 200 had a two-fold increase in both the prevalence and incidence of SIL and in non-regression from untreated low-grade SIL. In addition, these women had a lower response rate to treatment for high-grade SIL.

In the Women's Interagency HIV study (WHIS) (Massad *et al.,* 1999), baseline cytology in 1713 HIV-positive women and 482 high-risk HIV-negative women was abnormal in 38% of HIV-positive women compared with 16% of HIV-negative women. Risk factors for any abnormal cytology were CD4+ counts lower than 200 cells/μL (OR = 2.13; 95% CI 1.45–3.13), presence of HPV DNA and history of abnormal cytology.

Ellerbrock *et al.* (2000) showed that the prevalence of abnormal cytological findings was 4.3-fold higher in HIV-infected than in uninfected women, confirming the findings of other studies showing abnormal cytology rates of 23–60% (Provencher *et al.*, 1988). HIV-positive women were 4.5 times more likely to have histologically confirmed CIN at 54 months of follow-up than HIV-negative women.

Ahdieh *et al.* (2000) followed 84 HIV-negative and 184 HIV-positive injection drug users with six-monthly visits. Of the HIV-positive women, 70% were HPV DNA-positive at baseline compared with 26% of HIV-negative women. Cervical abnormalities were found in 13% of HIV-infected women versus 2% of HIV-negative women.

Following treatment, HIV-positive women have generally shown high recurrence rates ranging from 38% to 62%, compared with 15–18% in HIV-negative women (Petry *et al.*, 1994; Maiman *et al.,* 1999; Chirenje *et al.*, 2002).

Impact of anti-retroviral therapy on the natural history of preinvasive cervical lesions

The use of anti-retroviral therapy (ART) for treatment of HIV-infected individuals in developed countries has substantially reduced the associated morbidity and mortality. The increase in life expectancy may affect the burden of cervical cancer in either direction, depending on the degree to which the immune reconstitution allowed by ART is sufficient to diminish the risk of cervical cancer. The International Collaboration on HIV and Cancer (2000) found no change in the incidence of cervical cancer between 1992–96 and 1997–99 (i.e., after the use of ART had become widespread) in a reanalysis of cancer risk in 23 cohort studies in developed countries. This contrasted with marked reductions in the incidence of Kaposi sarcoma and non-Hodgkin lymphoma.

Recent studies have shown a beneficial impact of ART, with greater regression of HPV-associated lesions in treated women. In a cohort of French women, the prevalence of cervical HPV infection among 34 HIV-positive women remained unchanged at 81% five months after the initiation of ART. However the prevalence of CIN decreased from 69% to 53% after a median of five months ($p = 0.04$) and the mean CD4+ cell count was higher among women who regressed, suggesting that the loss of immune response and ART-induced reconstitution of immunity may have played a role in protecting against CIN (Heard *et al.,* 1998).

In a long-term follow-up study (Heard *et al.*, 2002) (median follow-up 17.7 months) of 168 HIV-positive women, 96 of whom were receiving ART, regression of CIN (defined as a regression to normality or to a lower grade of CIN) was seen in 40% of the 168 women. In a multivariate analysis, the grade of the lesion and the use of ART were independently associated with regression of CIN, after adjustment for CD4+ cell count. The relative hazard of regression of CIN in women receiving ART was 1.93 (95% CI 1.14–3.29; $p = 0.01$) compared with untreated HIV-positive women. In addition, a trend for a greater increase in CD4+ cell counts after six months of ART was observed in women who regressed compared with those who did not.

Minkoff *et al.* (2001) reported that women on ART were 40% more likely to show regression of cervical cytological abnormalities towards normality or lower-grade disease and less likely to show progression (OR = 0.68; 95% CI 0.52–0.88), after control for stage of HIV disease and severity of cytological abnormality. Among HIV-infected women, persistence of HPV infection and high HIV viral load were associated with cytological progression. Conversely, low HIV viral load and high CD4+ cell counts were associated with regression.

Moore *et al.* (2002) reported on 71 HIV-positive women who were examined by cytology, colposcopy and biopsy before starting ART and had at least one similar assessment six months after starting ART. The baseline prevalence of cervical disease was 55%, and at six months after starting ART 13% of the women showed regression without treatment of the cervix. No individual factor (e.g., smoking, HIV viral load, stage of HIV disease or CD4+ cell count) was significantly associated with regression, although a greater increase in CD4+ cell count in women on ART was most strongly associated with regression

(OR = 1.66 per 50 cell increase, p = 0.08).

These data suggest that the best responses of cervical disease in women on ART are seen among women with higher CD4+ cell counts or that women who respond to ART are those with the largest increase in CD4+ cell counts. It is important to note, however, that progression from CIN to cervical cancer occurs over many years. Before ART was introduced, HIV-positive women most often died of other HIV-related diseases and there was insufficient time for the development of cervical cancer. If ART leads to prolongation of life, women with cervical disease may be at greater risk of developing invasive cervical cancer if they do not enter a screening programme. Until a clear impact of ART on the regression and progression of HPV-associated lesions of the cervix is confirmed, HIV-positive women being treated with ART should undergo cervical screening and be actively treated where appropriate.