Chapter 2
Screening tests

Cervical cytology

Cytological testing involves collection of exfoliated cells from the cervix and microscopic examination of these cells after staining. The concept of utilizing exfoliative cytology to identify women with invasive cervical cancer was introduced by Papanicolaou and Babes in the 1920s (Papanicolaou, 1928; Papanicolaou & Traut, 1941). Subsequently, Papanicolaou refined the technique and demonstrated that conventional cytology could also be used to identify precancerous lesions of the cervix (Papanicolaou, 1954). The shift in emphasis from using cytology as a way to identify cases of invasive cervical cancer to using it to identify women with high-grade precursor lesions who are at risk for subsequently developing invasive cervical cancer was highly significant, as it meant that cervical cytology could be used to actually prevent the development of cervical cancer rather than simply identify cases at an early stage. In the 1960s, cervical cytology began to be widely used in many developed countries as a technique for cervical cancer prevention. Although the method was introduced over a half century ago, cytology-based screening programmes continue to be the mainstay of cervical cancer prevention.

Cytological terminology

Papanicolaou classes

The terminology developed by Papanicolaou separated cervical cytological findings into five categories or classes (Table 14) (Papanicolaou, 1954). At the time the classification was developed, there was only limited understanding of the relationship between cervical cancer precursor lesions and invasive cancers. Moreover, invasive cervical cancer was common and cervical cytology was initially viewed as a way of detecting early-stage, easily treated cancers. Therefore, the Papanicolaou classification system focused on how closely the exfoliated cells resembled those from an invasive cancer. Although the Papanicolaou classification was modified many times over the years, the problems inherent in this classification remain. For example, although is clear how Class I and Class V translate into known histological entities, Classes II, III or IV correlate less clearly with standard histopathological lesions. For example, should a carcinoma *in situ* be classified as Class IV and all grades of dysplasia as Class III, or does mild dysplasia correspond to Class II? There was also no consensus as to what other non-neoplastic conditions were combined in Class II. Such ambiguity in the Papanicolaou classification resulted in its non-uniform use by different cytologists. Modifications of the Papanicolaou classifications are still used in some countries. In the Netherlands, a modified Papanicolaou system (CISOE-A) is used for classification. This redefined and subdivided the Papanicolaou classes in order to make the terminology correlate with histopathological terminology (Hanselaar, 2002).

World Health Organization terminology

In the 1950s, some cytologists began to promote a more scientifically accurate terminology that would allow cytological diagnoses to translate directly into histological diagnoses. This terminology (Table 15) was later adopted by the World Health Organization (WHO) (Riotton *et al.,* 1973). The WHO terminology allows more precise correlation between cytological and histopathological findings, but is

Table 14. The original Papanicolaou classification

Class	Description
I	Absence of atypical or abnormal cells
II	Atypical cytology, but no evidence for malignancy
III	Cytology suggestive of, but not conclusive for, malignancy
IV	Cytology strongly suggestive of malignancy
V	Cytology conclusive for malignancy

From Papanicolaou, 1954

Table 15. Comparison of different terminologies used for cytologic reporting

Papanicolaou class system	World Health Organization	CIN	Bethesda System
Class I			Within normal limits
Class II			Benign cellular changes ASC
Class III	Mild dysplasia Moderate dysplasia Severe dysplasia	CIN1 CIN2 CIN3	Low-grade SIL High-grade SIL
Class IV	Carcinoma *in situ*	CIN3	
Class V	Microinvasive carcinoma Invasive carcinoma	Invasive carcinoma	Invasive carcinoma

Abbreviations: CIN, Cervical intraepithelial neoplasia; ASC, Atypical squamous cells; SIL, Squamous intraepithelial lesions
From Papanicolau (1954), Riotton *et al.* (1973), Richart (1968, 1973), Solomon *et al.* (2002)

difficult to use since it includes a number of different entities. These are *mild dysplasia, moderate dysplasia, severe dysplasia, epidermoid carcinoma in situ, epidermoid carcinoma in situ with minimal stromal invasion, invasive epidermoid microcarcinoma* and *invasive epidermoid carcinoma*. Studies have shown high rates of intra-observer and inter-observer variation with cervical cytology in general (Yobs *et al.,* 1987; Klinkhamer *et al.,* 1988; Selvaggi, 1999; Stoler & Schiffman, 2001). Classification systems that utilize more diagnostic categories have inherently higher rates of variability than do classification systems with fewer diagnostic categories (Yobs *et al.,* 1987; Selvaggi, 1999; Stoler & Schiffman, 2001; Kundel & Polansky, 2003). Other limitations of the WHO terminology are that it does not adequately deal with non-neoplastic conditions nor with specimen adequacy. Despite its limitations, many cytologists around the world continue to utilize the WHO terminology.

Cervical intraepithelial neoplasia (CIN) terminology

As a result of advances in understanding of the pathogenesis of cervical cancer, the cervical intraepithelial neoplasia (CIN) terminology was introduced in the late 1960s (Richart, 1968, 1973). The CIN concept emphasized that dysplasia and carcinoma *in situ* represent different stages of the same biological process, rather than separate entities. It had a major impact on how precancerous lesions were treated, since all types of cervical cancer precursor were considered to form a biological and clinical continuum. In the CIN terminology, mild dysplasia is classified as CIN 1, moderate dysplasia as CIN 2 and severe dysplasia and carcinoma *in situ* are grouped together and classified as CIN 3 (Table 15). The CIN terminology is still widely used in many countries for reporting both histological and cytological diagnoses.

The Bethesda System terminology

By the late 1980s, advances in our understanding of the role of human papillomavirus (HPV) in the pathogenesis of cervical cancer needed to be incorporated into cytological terminology. Moreover, it was recognized that clinicians were often confused by the non-standard terminologies used to report cytological results and that this had a potential adverse impact on clinical care. Therefore, in 1988, the US National Institutes of Health held a conference in Bethesda, Maryland, to develop a new terminology that would ensure better standardization and accommodate current concepts of the pathogenesis of cervical disease, so that cytological findings could be transmitted to clinicians as accurately and concisely as possible. The terminology that resulted is known as the Bethesda System. In 1991 the Bethesda System was slightly modified on the basis of experience obtained during the first three years of its use and it was further modified in 2001 to take into account the results of new research and over a decade of experience with the terminology (Luff, 1992; Solomon *et al.,* 2002).

The Bethesda system is viewed with caution in the United Kingdom, which retains its own British Society for Clinical Cytology (BSCC) 'dyskaryosis' terminology (British Society for Clinical Cytology, 1997). This can largely be mapped to the Bethesda system for comparison of data in a research setting, except for the borderline category, which may include koilocytes. Due largely to the robust nature of the 'severe dyskaryosis' category, fear of increasing the overtreatment inherent in cervical screening and the difficulty of achieving inter- and intra-observer agreement on 'low-grade' reports, the United Kingdom continues to use this terminology.

There are three distinct parts to each Bethesda System report: a state-

ment of the specimen adequacy, a general categorization and a descriptive diagnosis (Table 16). These categories assist clinicians by providing answers to three basic questions: (1) Do I need to repeat the cervical cytology? (2) Was the cervical cytology normal? (3) If the specimen was not completely normal, what specifically was wrong?

Because cervical cytology is considered a screening, rather than diagnostic, test, the 2001 Bethesda System reports cytological findings as an 'interpretation' or 'result' rather than as a 'diagnosis'. This stresses the fact that cytological findings usually need to be interpreted in the light of clinical findings, and that the test is designed to reflect the underlying disease state but does not always do so.

In this Handbook, the Bethesda System (SIL) terminology is used for cytological interpretation of screening tests unless otherwise reported.

Specimen adequacy

The 2001 Bethesda System requires that every cervical cytology specimen be assessed with respect to its adequacy (Solomon *et al.,* 2002). Specimens are classified into one of two categories: '*satisfactory for evaluation'* or '*unsatisfactory for evaluation*'. This represents a departure from the 1991 Bethesda System, which also included a third category for specimen adequacy that was called 'satisfactory for evaluation but limited by…..' or SBLB. This 'satisfactory but limited by…' category was most frequently used when a specimen lacked either endocervical cells or squamous metaplasia from the transformation zone but was in all other aspects 'satisfactory'. With the 2001 Bethesda System, cytology specimens previously classified as 'SBLB' are classified as 'satisfactory for evaluation' and a quality indicator comment is made indicating what limiting features are present.

A *'satisfactory for evaluation*' specimen must be appropriately labelled. To ensure proper identification, the woman's name or identifying number should be written on, or affixed to, the slide before it is sent to the cytology laboratory. Cytology laboratories should not accept unlabelled slides and should return them to the submitting clinician. It is also critical that the smear-taker provide pertinent clinical information to the laboratory that will evaluate the specimen, including the woman's age, date of last menstrual period, previous history of abnormal cervical cytology specimens or treatment for cervical disease, and the source of the specimen (e.g., vaginal or cervix). The minimal cellular requirements for a specimen to be considered 'satisfactory for evaluation' in the 2001 Bethesda System vary depending on whether the specimen is a conventional cytology specimen or a liquid-based cytology specimen. For classification of conventional cytology specimens as 'satisfactory for evaluation', an estimated 8000 to 12 000 well visualized squamous cells need to be present. For liquid-based cytology specimens, an estimated 5000 cells need to be present (Figure 29). Although the selection of these cut-offs is fairly arbitrary, the limit of 5000 cells for a liquid-based cytology specimen to be classified as 'satisfactory for evaluation' is based on a cell-counting study in which referent samples were diluted to produce preparations with defined numbers of squamous cells (Studeman *et al.,* 2003). A clear demarcation in sensitivity was observed using the SurePath™ procedure (see below) between specimens with less than 5000 squamous cells and those with 5000 cells or more: the sensitivity for a reference diagnosis case of low-grade squamous intraepithelial lesion (LSIL) increased from 73% for specimens with less than 5000 squamous cells to 98% for preparations with over 5000 cells.

There is much controversy over the importance of identifying a transformation zone component (e.g., squamous metaplastic cells) or endocervical cells in a cervical cytology preparation. Because the majority of high-grade precursor lesions arise within the transformation zone, it was widely believed until recently that specimens lacking a transformation zone component (TZC) or endocervical cells (EC) should be considered somewhat less than 'satisfactory for evaluation'. This view is supported by several studies that have shown the prevalence of SIL to be higher among cytology specimens that contain TZC/EC than among those that do not (Vooijs *et al.,* 1985; Mitchell & Medley, 1992; Szarewski *et al.,* 1993; Mintzer *et al.,* 1999). However, other studies have failed to confirm this association and, perhaps more importantly, several retrospective longitudinal cohort studies have found that women lacking TZC/EC are no more likely on follow-up to be diagnosed with squamous lesions than are women whose specimens contain TZC/EC (Mitchell & Medley, 1991; Mitchell, 2001). One retrospective case–control study of true

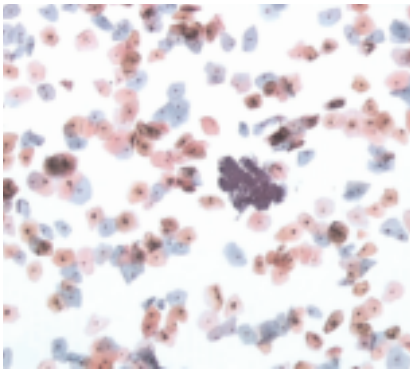

Figure 29 Liquid-based cytology: superficial and intermediate squamous cells and a cluster of columnar endocervical cells (obj. 5x)

positive and false negative cervical cytology specimens from women with CIN 3 found no difference in true positive rates between cases with or without TZC/EC (O'Sullivan *et al.,* 1998). A prospective study of women with normal cytology at entry found that although specimens containing EC at the subsequent test were at significantly higher risk of both low- and high-grade squamous intraepithelial lesions than those without EC, the presence or absence of EC at entry had no significant effect (Mitchell, 2001). In another compelling study on the lack of importance of EC, all negative cervical cytology specimens obtained in the Netherlands between 1990 and 1991 were matched with results of subsequent cytological and histological examinations (Bos *et al.,* 2001). There was no significant difference in the number of women subsequently diagnosed with CIN between women whose initial cytology specimens contained EC and those that did not. Moreover, the proportions of women diagnosed with cervical cancer were the same in both groups. It is also important to recognize that EC are less frequently found in cervical cytology specimens from women using oral contraceptives, who are pregnant or who are postmenopausal (Davey *et al.,* 2002). It has therefore been argued that specimens lacking EC or a TZC should not be considered unsatisfactory and may not need to be repeated (Davey *et al.,* 2002; Birdsong, 2001; Bos *et al.,* 2001). The 2001 Bethesda System recommends that reports should state whether or not EC or a TZC are present. Specimens lacking endocervical cells or squamous metaplastic cells should be classified as 'satisfactory for evaluation' and the quality indicator comment should indicate that these components are not present. The numeric criterion for stating that such a component is present is 10 well preserved endocervical or squamous metaplastic cells. Specimens in which inflammation, blood or poor preservation cause 50–75% of the epithelial cells to be obscured should be classified as 'satisfactory for evaluation', but a quality indicator comment made indicating that there are partially obscuring factors.

A specimen is classified as 'unsatisfactory for evaluation' when either the minimal number of epithelial cells required for interpretation is not present or blood, inflammation or poor preservation obscures more than 75% of the epithelial cell component (Figure 30). Cases which the laboratory cannot process, such as those received unlabelled, are also classified as 'unsatisfactory for evaluation' and no interpretation is rendered.

General categorization

The 'general categorization' is included as an optional component of the Bethesda System to allow clinicians to readily determine whether any degree of abnormality is present. With the 2001 Bethesda System, all cytology specimens are classified into one of three general categories. These include '*negative for intraepithelial lesion or malignancy', 'epithelial cell abnormalities' and 'other'.* These categories are mutually exclusive and specimens should be categorized according to the most significant findings.

'*Negative for intraepithelial lesion or malignancy'* includes all specimens in which no intraepithelial lesion or malignancy is identified. This includes cases with common infections such as *Trichomonas vaginalis*, fungal organisms such as Candida species, *Actinomyces* or herpes simplex virus, a shift in bacterial flora consistent with bacterial vaginosis, reparative/reactive changes, changes associated with intrauterine devices, radiation reactions or atrophic changes.

The category *'epithelial cell abnormalities'* includes both squamous and glandular cell abnormalities. This category is used whenever there are epithelial cell abnormalities, except for benign reactive or reparative changes.

The 2001 Bethesda System introduced a new general categorization, *'other'*. This category is used whenever there are no morphological abnormalities in the cells *per se*, but there are findings indicative that the woman is at some increased risk. An example is when benign-appearing endometrial cells are identified in a woman 40 years of age or older.

Squamous cell abnormalities

Atypical squamous cells (ASC): Epithelial cell abnormalities are subdivided into four categories (Table 16). 'Atypical squamous cells' (ASC) is used when cytological findings are considered suggestive but not diagnostic of a squamous intraepithelial lesion (SIL) (Figure 31). The term ASC was retained in the 2001 Bethesda System because of the wide recognition that these cells imply a significant risk for an underlying high-grade cervical intraepithelial lesion (SIL). In various studies, the prevalence of CIN 2 or 3 in women with ASC has varied between 10% and 20% (Wright *et al.,* 2002a). The ASC category roughly correlates with the 'borderline dyskaryosis' category used in the

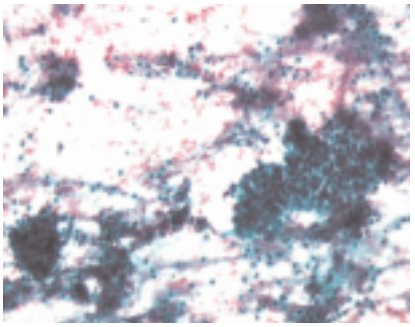

Figure 30 Unsatisfactory smear because of inflammation. Cell cluster difficult to analyse. Repeat after local treatment (obj. 10x)

Table 16. The 2001 Bethesda system

Specimen adequacy

Satisfactory for evaluation (*note presence/absence of endocervical transformation zone component*)
Unsatisfactory for evaluation (*specify reason*)
- Specimen rejected/not processed (*specify reason*)
- Specimen processed and examined, but unsatisfactory for evaluation of epithelial abnormality because of (*specify reason*)

General categorization (optional)

Negative for intraepithelial lesion or malignancy
Epithelial cell abnormality
Other

Interpretation/result

Negative for intraepithelial lesion or malignancy

Organisms
Trichomonas vaginalis
Fungal organisms morphologically consistent with *Candida* species
Shift in flora suggestive of bacterial vaginosis
Bacteria morphologically consistent with *Actinomyces* species
Cellular changes consistent with herpes simplex virus
Other non-neoplastic findings (Optional to report; list not comprehensive)
Reactive cellular changes associated with inflammation (includes typical repair), radiation, intrauterine contraceptive device
Glandular cells status posthysterectomy
Atrophy

Epithelial cell abnormalities

Squamous cell
Atypical squamous cell (ASC)
of undetermined significance (ASCUS)
cannot exclude HSIL (ASC-H)
Low-grade squamous intraepithelial lesion (LSIL)
High-grade squamous intraepithelial lesion (HSIL) (can use modifiers to separate into CIN 2 and CIN 3)
Squamous-cell carcinoma
Glandular cell
Atypical glandular cells (AGC) (specify endocervical, endometrial or not otherwise specified)
Atypical glandular cells, favour neoplastic (specify endocervical or not otherwise specified)
Endocervical adenocarcinoma *in situ* (AIS)
Adenocarcinoma

Other (List not comprehensive)
Endometrial cells in a woman ≥ 40 years of age

From Solomon *et al.* (2002)

United Kingdom. However, neither the WHO terminology nor the CIN terminology incorporates a category similar to ASC. The 2001 Bethesda System also clearly separates ASC from reactive/reparative changes and an interpretation of ASC should not be made whenever a cytopathologist identifies minor cytological abnormalities. The term ASC should be used only when the cytological findings are suggestive, but not diagnostic, of SIL. Currently, approximately 4–5% of all cervical cytology specimens are classified as ASC in the USA (Jones & Davey, 2000).

The '*atypical squamous cell' category is formally subdivided into two subcategories: 'atypical squamous cells – of undetermined significance'* (ASCUS or ASC-US) and *'atypical squamous cells – cannot exclude a high-grade SIL'* (ASC-H). This subdivision was felt to be important because

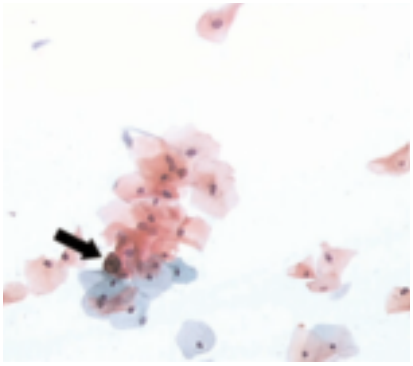

Figure 31 Parakeratotic cell (arrow), with an eosinophilic cytoplasm denser than normal superficial cells and a relatively regular but enlarged nucleus: ASCUS (rule out LSIL) (obj. 10x)

women with ASC-H (Figure 32) are at considerably higher risk for having CIN 2 or 3 and of being high-risk HPV DNA-positive than are women with ASCUS (Genest *et al.,* 1998; Sherman *et al.,* 1999, 2001; Selvaggi, 2003). Information from the US National Cancer Institute ASCUS–LSIL Triage Study (ALTS) clinical trial indicates that the risk that a woman with ASC-H has CIN 2 or 3 is over twice that of a woman with ASCUS (Table 17) (Sherman *et al.,* 2001). Moreover, the prevalence of high-risk HPV DNA-

Table 17. Prevalence of high-risk HPV DNA and CIN 2 and CIN 3 in women with ASCUS and ASC-H in the ASCUS-LSIL triage study (ALTS)*

Cytology result	No.	% high-risk HPV DNA-positive	% biopsy-confirmed CIN 2+	% biopsy-confirmed CIN 3+
ASCUS	764	63.2%	11.6%	4.7%
ASC-H	116	85.6%	40.5%	24.1%
HSIL	213	98.7%	59.2%	37.6%

* Study provides the results for liquid-based cytology specimens that were tested for high-risk types of HPV using Hybrid Capture 2
From Sherman *et al.* (2001)

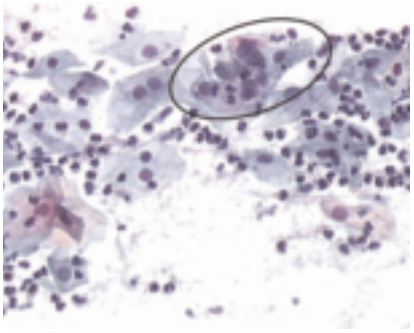

Figure 32 Inflammatory smear with parabasal squamous cells with enlarged nuclei: ASC-H (ellipse) (obj. 20x)

positivity among women with ASC-H is almost as high as that of women with a high-grade squamous intraepithelial lesion (HSIL) cytological result. Therefore the recommended management of women with ASCUS and ASC-H differs (Wright *et al.,* 2002a).

Low-grade squamous intraepithelial lesion: The LSIL category in the Bethesda System includes both HPV effects and CIN 1 (i.e. mild dysplasia). Most cytologists consider the cytopathic effects of HPV, including multinucleation, perinuclear halos and nuclear atypia with irregular nuclear outlines and hyperchromasia, to overlap the cytological features of CIN 1. These features are referred to as 'koilocytosis', a term derived from the Greek *koilos*, meaning hollow.

The classical studies of Reagan and others identified the key cytological features of CIN 1 (Table 18) (Reagan & Hamomic, 1956). The cells are of the superficial or intermediate-cell type. They are classically described as having nuclei 4–6 times the size of a normal intermediate-cell nucleus (Figure 33). However, nuclei may vary in size and, in many cases of LSIL that are characterized by marked HPV cytopathic effects, are only twice the size of a normal intermediate-cell nucleus. The nuclei are usually hyperchromatic, and multinucleation is common. The chromatin is finely granular and uniformly distributed. The cells typically occur as individual cells or as sheets of cells with well defined cell borders.

High-grade squamous intraepithelial lesion: Because the Bethesda System combines moderate and severe dysplasia together with carcinoma *in situ* in the HSIL category, there is wide variation in the cytological appearance of HSIL. When applying the 2001 Bethesda System, many cytopathologists utilize the option of subdividing HSIL into CIN 2 and CIN 3 lesions. As the severity of the lesion increases, the degree of differentiation and the amount of cytoplasm decreases, the nuclear:cytoplasmic ratio increases, and the degree of

Table 18. Cytological features of squamous intraepithelial lesions

Bethesda system	LSIL	HSIL		
CIN terminology	CIN 1	CIN 2	CIN 3	
WHO terminology	Mild dysplasia	Moderate dysplasia	Severe dysplasia	Carcinoma *in situ*
Cell type	Superficial or intermediate	Parabasal	Basal	Basal, spindle, pleomorphic
Cell arrangement	Singly or sheets	Singly or sheets	Singly or sheets	Singly or sheets or syncytia
Number abnormal	+	++	+++	++++
Koilocytosis	+++	+	+/–	+/–
Nuclear size	+++	++	+	+
Hyperchromasia	+	++	+++	++++
Nuclear:cytoplasmic ratio	+	++	+++	++++

From Reagan & Hamomic, 1956

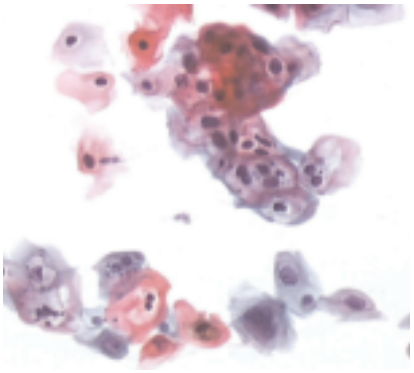

Figure 33 LSIL: typical eosinophilic and basophilic koilocytes associated with some parakeratosis and binucleated cells (obj. 20x)

nuclear atypia also increases. HSIL of the moderate dysplasia type typically contains cells similar to those seen in LSIL, as well as atypical immature cells of the parabasal type (Figure 34). The nuclei of these cells are more hyperchromatic and irregular than typically seen in LSIL. In severe dysplasia, the overall size of the cells is reduced compared to mild and moderate dysplasia, but because the cells demonstrate minimal differentiation, the nuclear:cytoplasmic ratio is greatly increased. In severe dysplasia, there are usually considerably greater numbers of neoplastic cells that are typically found individually. Carcinoma *in situ* can be of the small-cell type, of the large-cell non-keratinizing type or of the large-cell keratinizing (pleomorphic) type. Although separation of carcinoma *in situ* into these three different cytological types has little clinical significance, all three have quite different cytological appearances. Small-cell lesions consist of small basal-type cells similar to those seen in severe dysplasia but which demonstrate even less cytoplasm and higher nuclear:cytoplasmic ratios (Figure 35). Because of their small size, these cells can easily be overlooked during routine screening and such cases account for a disproportionate percentage of false negative cytological results. The cells of large-cell non-keratinizing lesions typically form syncytial-like cell sheets in which individual cell membranes are difficult to identify. These cells have enlarged, hyperchromatic nuclei and minimal amounts of cytoplasm. The keratinizing large-cell type of carcinoma *in situ* is composed of pleomorphic, highly atypical cells, many of which have thick keratinized cytoplasm. These cells are often spindled or tadpole-shaped and have extremely dense nuclear chromatin (Figure 36).

Invasive squamous-cell carcinoma: Cytologically, squamous-cell carcinomas of the cervix are subdivided into keratinizing and non-keratinizing types. Non-keratinizing carcinomas (Figure 37) typically have large numbers of malignant cells that form loose cell sheets and syncytial arrangements. The cells have enlarged nuclei with coarsely clumped chromatin, prominent macronucleoli and focal chromatin clearing. A key cytological feature is the presence of a 'dirty' background containing blood and necrotic material. This is often referred to as a tumour diathesis. This characteristic background is usually less prominent in liquid-based cytology specimens.

Cervical smears from women with *keratinizing carcinomas* contain malignant cells of a variety of shapes and sizes (Figure 38). Some of the cells are pleomorphic or tadpole-shaped with nuclei that are irregular in shape and

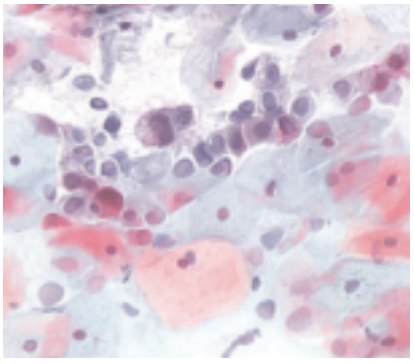

Figure 34 Parabasal cells arranged in a pile with nuclear enlargement, irregular nuclear outlines and coarse chromatin. HSIL (moderate dysplasia) (obj. 20x)

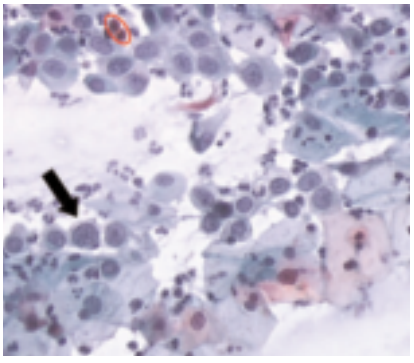

Figure 35 HSIL (severe dysplasia): inflammatory smear containing many parabasal cells with enlarged nuclei with irregular chromatin (black arrow). Some cells with a mildly eosinophilic cytoplasm (ellipse) (obj. 20x)

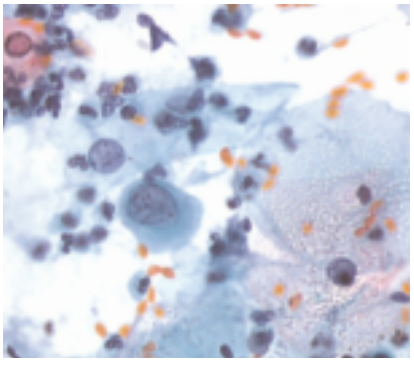

Figure 36 HSIL (severe dysplasia): basal cells with enlarged nuclei and irregular or very dense and opaque chromatin (arrow), accompanied by an atypical mature cell (obj. 40x)

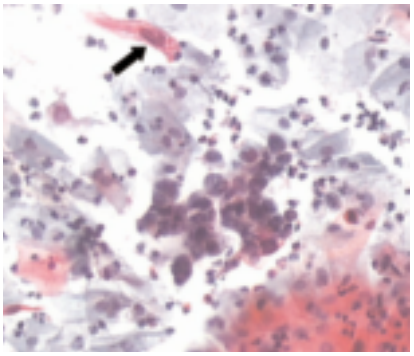

Figure 37 Invasive squamous cell carcinoma
One cluster of pleomorphic and poorly differentiated malignant cells and one isolated cell of abnormal shape (arrow). Inflammation, blood and necrosis in the background (obj. 20x)

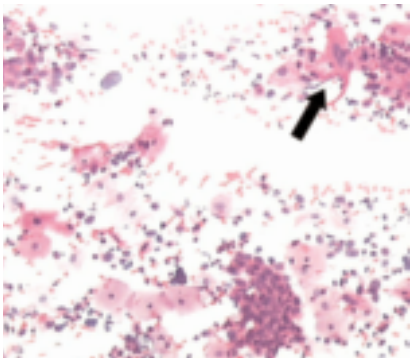

Figure 38 Invasive squamous cell carcinoma
Pleomorphic malignant cells, isolated or in clusters, sometimes keratinized or necrotic with bizarre cell shapes (arrow). Inflammation, blood and necrosis in the background (obj. 10x)

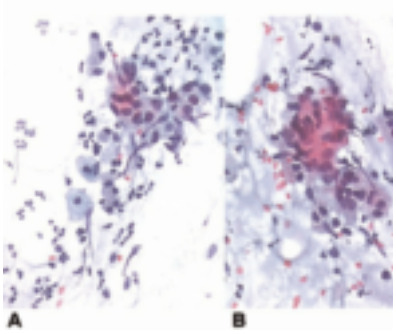

Figure 39 Smear from the transformation zone and endocervix
Sheets of atypical glandular cells (AGC) with enlarged nuclei with similar chromatin pattern in all cells (A and B: obj. 20x)

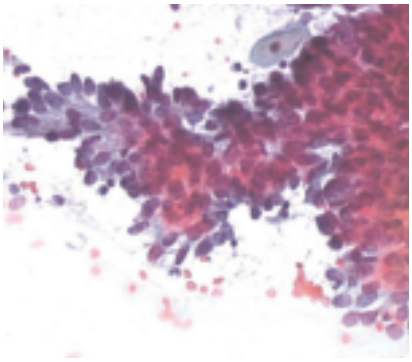

Figure 40 Endocervical adenocarcinoma *in situ* (AIS)
Atypical columnar endocervical cells, with enlarged, elongated and hyperchromatic nuclei. Typical feathering and palisading. (obj. 20x)

quite hyperchromatic. Unlike non-keratinizing squamous-cell carcinoma, keratinizing squamous-cell carcinomas usually do not have a 'dirty' background or evidence of tumour diathesis.

Glandular cell abnormalities

Glandular cell abnormalities are categorized into four categories: *atypical glandular cells* (AGC), *atypical glandular cells – favour neoplasia, adenocarcinoma* in situ and *adenocarcinoma*. Whenever possible, atypical glandular cells are categorized as to whether they are endocervical or endometrial in origin.

Atypical glandular cells (AGC): Glandular cytological abnormalities are considerably less common than squamous abnormalities and most cytologists tend to be less comfortable recognizing and diagnosing them. In addition, the criteria used to differentiate reactive endocervical changes from neoplasia are less well established than those used for squamous lesions. Cytologists even have difficulty in differentiating atypical endocervical cells from cases of CIN 2 or CIN 3 that have extended into endocervical crypts. This accounts for the high prevalence of squamous abnormalities (approximately 30%) detected in women referred to colposcopy for AGC (Eddy *et al.,* 1997; Veljovich *et al.,* 1998; Ronnett *et al.,* 1999; Jones & Davey, 2000; Krane *et al.,* 2004).

The cytological features of atypical glandular cells vary depending on the degree of the underlying histopathological abnormality and whether or not the cells are endocervical or endometrial in origin. Atypical glandular cells of endocervical origin frequently form dense two- or three-dimensional aggregates that have minor degrees of nuclear overlapping. In some cases, the chromatin is somewhat granular and nuclear feathering can be seen at the periphery of the cellular aggregates (Figure 39). In cases interpreted as *atypical glandular cells – favour neoplasia,* there is more marked cytological abnormality and typically a greater number of abnormal cells.

Adenocarcinoma in situ: In cases of adenocarcinoma *in situ*, there are usually a larger number of atypical glandular cells that form crowded cellular clusters (Figure 40). The sheets are usually three-dimensional. The cells within these sheets occasionally form rosettes and have extensive feathering of the cells at the periphery. Individual endocervical cells are highly atypical with enlarged round, oval or elongated nuclei that vary in size from cell to cell. In most cases, the chromatin is coarsely clumped and multiple mitoses are seen.

Adenocarcinoma: Invasive adenocarcinomas should be subclassified into the endocervical or endometrial type whenever possible. The cytological diagnosis of invasive adenocarcinoma is relatively straightforward. Adenocarcinoma cells from either an endocervical or an endometrial primary type have enlarged nuclei, high nuclear:cytoplasmic ratios, coarsely clumped chromatin and prominent nucleoli (Figure 41). They can occur singly or in clusters.

Other terminologies

Although the 2001 Bethesda System classification is applied in many countries, other classification systems are also widely used. As mentioned previously, many countries prefer to subclassify high-grade intraepithelial lesions into at least two categories. This is the approach used in the United Kingdom, where squamous intraepithelial abnormalities are divided into five categories (borderline changes; mild, moderate, severe dyskaryosis and severe dyskaryosis or possibly invasive cancer) (British Society for Clinical Cytology, 1997).

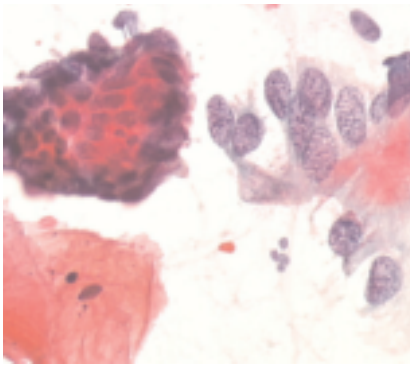

Figure 41 Histologically proven invasive adenocarcinoma
More or less cohesive mallignant columnar cells next to a less atypical cell group (obj. 40x)

Conventional cervical cytology

The importance of proper specimen collection cannot be overemphasized. Although no formal studies have demonstrated that educating clinicians on the optimal technique of obtaining cervical cytology samples improves specimen quality, there is considerable anecdotal evidence that this is important (Krieger *et al.,* 1998). One half to two thirds of false negative cervical cytology results are attributable to either poor patient conditions at the time the cervical specimen is collected or the manner in which it is collected (Morell *et al.,* 1982; Gay *et al.,* 1985; Vooijs *et al.,* 1985; Agency for Health Care Policy and Research, 1999). Therefore it is important that clinicians and nurses obtaining specimens be adequately trained in specimen collection and that they avoid situations that may reduce the performance of the test (McGoogan *et al.,* 1998). This is especially important in low-resource settings, where women may undergo screening only once or twice in their lifetime.

Preparing the woman

Whenever possible, appointments for a cervical cytology examination should be scheduled approximately two weeks after the first day of the last menstrual period. Patients should be instructed to avoid sexual intercourse and douching for 24 to 48 hours before having the cytology specimen collected. In addition, women should not use any intravaginal products or medicine for several days before the smear is taken. Women using an intravaginal estrogen product should discontinue its use several days before the examination.

Circumstances that may interfere with the interpretation of a cervical cytology test include active menstruation, significant cervical or vulvovaginal infections and a timing less than eight weeks post-partum. When a woman is actively menstruating, blood and cellular debris from the endometrium tend to obscure the cells on the smear, particularly during the first few days. Similarly, a cytology specimen should not be obtained when an abnormal vaginal or cervical discharge is observed. Women with a discharge should be evaluated for cervicitis and vaginitis using appropriate tests and be treated before the cytology specimen is taken, otherwise the specimen may be compromised by the inflammatory exudates or mildly reactive cells may be misinterpreted as a significant cytological abnormality.

There is controversy as to the ideal timing of post-partum smears. Smears obtained less than eight weeks post-partum are often difficult to interpret because of marked inflammation and reparative changes, so a high rate of mild cytological abnormalities may be diagnosed. Another factor that can adversely affect the interpretation of cervical cytology specimens is severe atrophy.

Although one should strive to collect specimens under ideal conditions, failure to comply with suggested screening intervals presents a greater risk to women. For previously non-compliant women, particularly those at risk for cervical neoplasia, a smear obtained under less than ideal conditions is preferable to no smear at all.

Equipment

To collect a conventional cervical cytological specimen, the equipment required is a speculum, a light source, a collection device, a glass slide and fixative. Since most cervical cancer precursors and invasive cancers occur in the transformation zone, the use of specially designed devices that sample this area is recommended. The most common is a wooden or plastic spatula that conforms to the curvature of the portio. It is critical that the endocervical canal be sampled in order to obtain

reasonable sensitivity (Martin-Hirsch *et al.,* 1999) and many spatula-type devices have extended tips designed to collect cells from this area. Either a moistened cotton swab or a brush-type endocervical sampler device (e.g., cytobrush) can be used to collect a second sample directly from the endocervical canal after the portio has been sampled (Koonings *et al.,* 1992; Kohlberger *et al.,* 1999). Recently developed collection devices that sample the endocervix and exocervix simultaneously do not provide a significantly lower false negative rate than the combination of spatula and a conical cervical brush (Szarewski *et al.,* 1993).

There is no consensus as to whether a single-slide technique, with both samples of the ectocervix and endocervix placed on the same slide, or a technique in which the two samples are put on two separate slides is preferable. Comparative studies of the two techniques have reported similar results (Saitas *et al.,* 1995; Quackenbush, 1999). The single-slide approach has the advantage of reducing screening time and laboratory workload and it decreases the storage space required for archiving slides. When a single-slide technique is utilized, there also is no consensus on whether the specimens from the ectocervix and endocervix should be mixed together on the slide or kept separate as in the V (vagina) C (ectocervix) E (endocervix) technique.

Collecting the sample

A conventional cytology specimen is typically obtained using a spatula and conical cervical brush. The slide must first be labelled with the woman's name or number. Laboratories should have a written protocol specifying what is considered adequate labelling and should not accept inadequately labelled specimens. The person collecting the specimen should ensure that a test requisition is accurately and legibly filled out before collecting the specimen. The information most commonly requested by laboratories includes:

- Woman's name and indication if there has been a name change in the last five years. Some laboratories also use unique patient identifier numbers
- Date of birth or age
- Menstrual status (date of last menstrual period, whether the woman is pregnant, post-partum, on hormone replacement therapy, or has had a hysterectomy)
- Previous history of abnormal cervical cytology, or treatment for CIN or cancer
- Whether the clinician considers the woman to be at high risk for developing CIN or cancer. Possible risk factors include smoking, infection with HIV, lack of previous screening and multiple partners.
- Specimen source – vaginal or cervical

Good visualization of the cervix is important for obtaining an adequate specimen. Cervical cytology specimens are generally collected with the woman in the dorsolithotomy position. A sterilized or single-use bivalve speculum of appropriate size is inserted into the vagina in such a manner as to allow complete visualization of the cervical os and as much of the transformation zone as possible. The cervix should not be contaminated with lubricant or water-soluble gel that may obscure the smear. Therefore the smear must be obtained *before any bimanual examination.* Gentle removal of excess mucus and discharge from the cervix with a large cotton-tipped applicator can produce a better-quality smear (Kotaska & Matisic, 2003), but vigorous cleansing may remove many of the most easily exfoliated cells. Saline should not be used to help clear debris from the surface of the cervix. It is also preferable not to apply 3–5% acetic acid to the cervix before taking the cytology specimen, as this can reduce the cellularity of the smear and produce poor staining (Griffiths *et al.,* 1989; Cronje *et al.,* 1997).

Before the specimen is collected, the cervix should be carefully inspected with the naked eye for grossly visible masses or ulcerations that may indicate an invasive cervical cancer. If a grossly visible lesion is identified, the woman should be referred for further confirmation. In many cases, the lesion can be directly sampled and the cellular sample obtained can be submitted separately for cytological assessment. The procedure for collecting cells from the cervix varies depending on the type of device used and the number of slides to be prepared. If a spatula and conical cervical brush are utilized, the first step is to place the spatula firmly against the ectocervix with the long projection extending into the endocervical canal. The spatula is then rotated several times 360° around the portio and removed. It is important to ensure that the entire squamocolumnar junction is sampled, since this is the site where most CIN lesions develop. In most women, the spatula will come into contact with the squamocolumnar junction if the pointed end is placed in the os, but in young women with a large ectopy, the spatula may need to be moved laterally to sample a peripherally positioned squamocolumnar junction. When rotating the spatula, it is easy to miss part of the cervix; this can be alleviated by directly visualizing the cervix while sampling. Transfer is best performed by using the spatula to thinly

spread the cells onto the glass slide. It is important to ensure that as much cellular material as possible is transferred from *both* sides of the spatula.

The endocervical canal is then sampled, using a conical cervical brush, which is placed in the endocervical canal so that the last few bristles remain visible and then gently rotated 90° to 180° once. One such rotation will adequately sample the endocervical canal and generally does not produce bleeding. Material from both sides of the spatula should be spread onto the slide.

If collection devices that simultaneously sample both the endocervix and the ectocervix are used, the manufacturers' directions should be followed for each type of device.

Cell fixation must be performed within a few seconds of specimen collection in order to prevent air-drying, which obscures cellular detail and hinders interpretation (Somrak *et al.,* 1990). Immersing the slide in alcohol or spraying it with a specially formulated spray fixative can prevent air-drying. With immersion fixation, the slide is either immersed in alcohol and transferred to the laboratory in the container of alcohol or allowed to fix for 20 to 30 minutes in the alcohol, removed and allowed to air-dry. Various different spray fixatives are available. Only spray fixatives specifically designed for cytological specimens should be used and the manufacturer's instructions for a given product must be followed. The fixative should be liberally applied such that the slide appears moist over its entire surface. In order to prevent disruption of the cellular layer on the slide, the container of spray fixative should generally be held 15–25 cm from the slide during application.

Performance of conventional cytology

Despite the proven effectiveness of cervical cytological screening in reducing the incidence of cervical cancer, over the last decade the accuracy of cervical cytology has been questioned. Two factors need to be considered when assessing the accuracy of any screening or diagnostic test. One is whether the test is specific in detecting a given condition; the other is the sensitivity of the test for detecting the condition. Several large meta-analyses have indicated that both the sensitivity and specificity of cervical cytology are lower than previously thought (Fahey *et al.,* 1995; McCrory *et al.,* 1999; Nanda *et al.,* 2000). [The Working Group considered the estimates of cytology test performance obtained through these meta-analyses to be of concern, given current cytology practices. In particular, it felt that it is very unlikely that specificities as low as 60–70% would be observed in a modern cytological screening practice.] Table 19 presents the sensitivities and specificities of conventional cervical cytology observed in a number of recent large cervical cancer screening studies. Even within the confines of research studies, a wide range of performance has been reported.

Liquid-based cervical cytology

Liquid-based cytology (LBC) was introduced in the mid-1990s as a way to improve the performance of the test. Rather than having the clinician prepare the cytological specimen at the bedside by spreading the exfoliated cells onto a glass slide, the cells are transferred to a liquid preservative solution that is transported to the laboratory, where the slide is prepared.

Table 19. Performance of conventional cytology in various large research studies

Author	Country	Ages	Study size	Sensitivity (%)	Specificity (%)	Histological cut-off
Cuzick *et al.* (1999a)	United Kingdom	34+	2988	86	98	CIN 2+
Hutchinson *et al.* (1999)	Costa Rica	18+	8636	55	98	CIN 2+
Ratnam *et al.* (2000)	Canada	18–69	2098	56	62	CIN 2+
Denny *et al.* (2000a)*	South Africa	35–65	2944	70	85	CIN 2+
Denny *et al.* (2002)*	South Africa	35–65	2754	40	96	CIN 1+
Cuzick *et al.* (2003)	United Kingdom	30–60	11 085	77	96	CIN 1+
Petry *et al.* (2003)	Germany	30+	8466	44	98	CIN 2+
Salmerón *et al.* (2003)	Mexico	15–85	7868	59	98	CIN 1+
Sankaranarayan *et al.* (2004b)	India	25–65	10 591	65	92	CIN 2+

Cytological cut-off for referral for all studies is ASC or greater except for those studies marked by asterisk, where a cut-off of LSIL or greater was used.
Sensitivity and specificity are estimated cross-sectionally (see Chapter 4)

A number of different LBC techniques are in use worldwide. These include ThinPrep®, SurePath™, Cytoscreen™, Cyteasy®, Labonord Easy Prep, Cytoslide, SpinThin and PapSpin. The first two of these are approved for use in the USA by the Food and Drug Administration (FDA) and are the most widely used methods worldwide. They are therefore the best characterized in terms of performance. With the ThinPrep method, clumps of cells and mucus are broken up by mechanical agitation and then the liquid preservative solution is filtered through a membrane filter with a pore size specifically designed to trap epithelial cells while allowing contaminating red blood cells and inflammatory cells to pass through. The epithelial cells collected on the membrane filter are then transferred onto a glass slide and stained. This produces a relatively thin, monolayer-type preparation. The ThinPrep-2000 processor allows one specimen to be processed at a time, whereas the newer ThinPrep-3000 processor is more fully automated and allows up to 80 samples to be processed at a time. In contrast, with the SurePath method, clumps of cells and mucus are broken up by aspiration through a syringe. The cell suspension is then layered on top of a density gradient and the red blood cells and inflammatory cells are separated from the epithelial cells by density gradient centrifugation. The resulting cell pellet containing predominantly epithelial cells is then inserted into a robotic workstation, where it is resuspended and transferred to a glass microscope slide. The SurePath method allows up to 48 samples to be processed at a time.

LBC is purported to have a number of advantages over conventional cervical cytology. These include a more representative transfer of cells from the collection device to the glass slide, a reduction in the number of unsatisfactory cytology specimens, the availability of residual cellular material for subsequent molecular testing or for making additional glass slides, and possibly increased detection of HSIL.

Performance of liquid-based cytology methods

Numerous studies have evaluated the comparative performance of the two most commonly used LBC methods (ThinPrep and SurePath) and conventional cytology with respect to test positivity, their sensitivity and specificity for identification of CIN, the time required for evaluation of the specimens, and specimen adequacy. Although there is reasonable agreement that LBC improves specimen adequacy and reduces screening time compared to conventional cytology, there is considerable controversy surrounding the relative sensitivity and specificity of the two approaches, largely due to a lack of well designed comparative studies.

Most comparative studies have utilized one of two types of study design: split-sample studies and historical control studies. Split-sample studies collect cells from the cervix using a single collection device and a conventional cervical cytology specimen is prepared first. Residual cells remaining on the device are then transferred to a liquid-based cytology preservative. Therefore each woman acts as her own control and detection rates in conventional and LBC specimens are compared. The other widely used study design, known as 'direct to vial', compares the performance of LBC collected in the routine manner (direct transfer to the preservative solution) during a given time period with historic control data obtained using conventional cytology. Both study designs have significant limitations. With split-sample studies, it is difficult to ensure that the two cytology specimens are comparable. Since the conventional cytology slide is prepared first and the LBC specimen is prepared second, this design would seem to lead inherently to bias against LBC. Therefore it has been argued that split-sample studies do not demonstrate the full benefit that could be obtained when LBC is utilized in routine clinical practice. Studies utilizing historical controls avoid the need to prepare several cytology specimens from a single woman, but introduce other potential biases, including the comparability of the populations being compared.

Other significant limitations found in many of the studies evaluating LBC include failure to compare test performance with a reference standard of 'blinded' colposcopy/biopsy and a study population of women followed up for a prior cytological abnormality rather than women undergoing routine screening. A review of new cervical cytology methods conducted in 2001 for the US Preventive Services Task Force and the Agency for Healthcare Research and Quality found that out of 962 potentially relevant studies, not one met their predefined inclusion criteria (Hartmann *et al.,* 2001). This was commonly due to lack of an adequate reference standard, but most studies were excluded for more than one reason. At the time the review was conducted, only one study, from Costa Rica, had applied a definitive clinical reference standard to a random sample of women with normal screening test results and allowed the relative sensitivity and specificity for LBC and conventional cervical cytology to be calculated (Hutchinson *et al.,* 1999). The Costa Rica study was a split-sample study rather than a direct-to-vial study. The other studies that were reviewed used various types of clinical reference standard, including a combination of histological follow-up and conventional cytology follow-up with incomplete data, a consensus expert panel diagnosis of the index specimens, histological follow-up or

consensus expert panel diagnosis in cases of missing follow-up, and histological follow-up of HSIL combined with a balanced follow-up diagnosis of all other available follow-up data. The most common reference standard used in studies of LBC performance has been the expert panel review of selected cytology specimens. Unfortunately, with expert panel review, the screening test findings are not related to the true disease status of the cervix, making determination of false negatives and false positives, and hence sensitivity and specificity, impossible. Cervical biopsy diagnoses obtained as part of routine follow-up of women with abnormal cervical cytology results are another commonly used reference standard in studies of LBC. However, unless the pathologist is blinded to the original cytological findings, it is quite possible that the interpretation of cervical biopsy specimens will be biased. Large, randomized controlled clinical trials comparing the performance of LBC and conventional cytology need to be conducted by laboratories in which the techniques are well established. Although the results of no such studies are yet available, one large randomized trial is currently under way in the Netherlands (M.A. Arbyn, personal communication).

Several systematic, evidence-based reviews of the published literature on LBC have been published (Nanda *et al.,* 2000; Payne *et al.,* 2000; Bernstein *et al.,* 2001; Hartmann *et al.,* 2001; Sulik *et al.,* 2001; Abulafia *et al.,* 2003; Klinkhamer *et al.,* 2003; Arbyn *et al.,* 2004a). These reviews are based on test positivity ratios or detection rates, i.e., relative sensitivities and specificities of histologically confirmed lesions. They have come to somewhat conflicting conclusions (Table 22). It is important to note that the comparative utility of LBC relative to conventional cervical cytology will vary from one setting to another. The National Health Service of the United Kingdom recently agreed to introduce LBC throughout the country, in view of the reduction of inadequate specimens from 9% with conventional cervical cytology to 1–2% with LBC (National Institute for Clinical Excellence, 2003).

Table 20 presents data from a number of 'direct-to-vial' studies. Although there is considerable variation between the studies in the prevalence of HSIL identified using either conventional cytology or LBC, on average the use of LBC increased the rate of detection of HSIL in these stud-

Table 20. Comparison of identification of SIL using conventional cytology with LBC in representative "direct-to-vial" studies

Reference	LBC test	Population	Conventional			Liquid-based cytology			Increase in HSIL
			No.	LSIL	HSIL	No.	LSIL	HSIL	
Bolick & Hellman (1998)	TP	Screening	39 408	0.8%	0.3%	10 694	2.3%	0.8%	173%
Dupree *et al.* (1998)	TP	Screening	22 323	0.9%	0.2%	19 351	1.4%	0.3%	50%
Papillo *et al.* (1998)	TP	Screening	18 569	0.9%	0.5%	8541	1.6%	0.7%	55%
Carpenter & Davey (1999)	TP	High-risk	5000	4.4%	1.9%	2727	6.9%	2.4%	26%
Diaz-Rosario & Kabawat (1999)	TP	Screening	74 756	1.6%	0.26%	56 339	2.7%	0.52%	102%
Guidos & Selvaggi (1999)	TP	Screening	5423	1.0%	0.3%	9583	3.6%	1.0%	233%
Vassilakos *et al.* (1999)	SP	Screening	88 569	1.6%	0.4%	111 358	2.5%	0.7%	79%
Hatch (2000)	TP	High-risk	16 260	2.9%	1.5%	7934	6.1%	3.2%	116%
Tench (2000)	SP	Screening	10 367	0.6%	0.5%	2231	1.0%	0.7%	46%
Weintraub & Morabia (2000)	TP	Screening	126 619	0.5%	0.1%	39 455	1.8%	0.5%	400%
Obwegeser & Brack (2001)	TP	Screening	1002	3.7%	1.8%	997	4.7%	1.6%	– 11%
Baker (2002)	TP	Screening	4872	2.8%	0.7%	3286	4.1%	1.0%	43%
Cheung *et al.* (2003)	TP	Screening	191 581	1.0%	0.25%	190 667	1.7%	0.24%	– 4%
Moss *et al.* (2003)	TP	Screening	67 856	2.3%	1.4%	34 128	2.6%	1.7%	21%
	SP	Screening	43 280	2.3%	1.4%	47 642	2.3%	1.2%	–14%
Colgan *et al.* (2004)	SP	Screening	445 225	1.4%	0.40%	445 011	1.8%	0.35%	–

Abbreviations: TP, ThinPrep; SP, SurePath

Table 21. Comparison of specimen adequacy in conventional cytology with LBC in "direct-to-vial" studies

Reference	LBC test	Population	Conventional			Liquid-based cytology		
			No.	Limited (%)	Unsatisf. (%)	No.	"Limited' (%)	Unsatisf. (%)
Bolick and Hellman (1998)	TP	Screening	39 408	17.8	1.0	10 694	11.6	0.3
Dupree *et al.* (1998)	TP	Screening	22 323		2.0	19 351		3.8
Diaz-Rosario and Kabawat (1999)	TP	Screening	74 756	22.0	0.2	56 339	18.7	0.7
Carpenter and Davey (1999)	TP	High-risk	5000	19.4	0.6	2727	10.5	0.3
Guidos and Selvaggi (1999)	TP	Screening	5423	21.4	1.2	9583	0.7	0.5
Vassilakos *et al.* (1999)	SP	Screening	88 569	4.7	1.5	111 358	1.2	0.2
Tench (2000)	SP	Screening	10 367	31.0	2.9	2231	15.8	0.4
Weintraub and Morabia (2000)	TP	Screening	130 050	27.8	0.3	39 790	8.1	0.2
Obwegeser and Brack (2001)	TP	Screening	1002	2.5	0	997	5.5	1.4
Baker (2002)	TP	Screening	4872	18.2	0.7	3286	9.1	0.8
Cheung *et al.* (2003)	TP	Screening	191 581	2.6	0.48	190 667	0.5	0.32
Moss *et al.* (2003)	TP	Screening	74 584		9.7	34 813		2.0
	SP	Screening	47 632		9.1	21 456		0.9

ies. The wide variety of study populations makes comparisons difficult. This is because estimates of performance are influenced by outlying results of a few studies. In a comprehensive formal meta-analysis of all published 'direct-to-vial' studies that adjusted for outlying results, Arbyn *et al.* (2004a) found a pooled ratio for detection rate of HSIL in ThinPrep specimens versus conventional cytology of 1.72 (95% CI 1.42–2.08) and for SurePath specimens versus conventional cytology of 1.47 (95% CI 1.14–1.89). It is important to bear in mind the limitations to interpretation of these studies, as described above, and that the actual number of additional cases classified as HSIL using LBC is quite small—only about three cases per 1000 women screened.

Specimen adequacy

The effect of LBC on specimen adequacy rates has been evaluated in a number of the 'direct-to-vial' studies (Table 21). Both the ThinPrep and SurePath methods appear to produce fewer specimens classified as either 'limited by obscuring factors' such as blood, inflammation or poor preservation than does conventional cytology. In addition, in many studies both methods have reduced the number of specimens classified as 'unsatisfactory for evaluation'. In a recent pilot study in the United Kingdom (Moss *et al.,* 2003), the use of ThinPrep reduced the 'inadequate' rate from 9.7% to 2.0%. The use of SurePath reduced the 'inadequate' rate from 9.1% to 0.9%. For all study sites combined, there was an 82.7% reduction (rate ratio 0.173, 95% CI 0.17–0.19). The reduction was significant in each of three age groups: 20–34, 35–49 and 50–64 years. In a meta-analysis of the comparative performance of LBC, Arbyn *et al.* (2004a) estimated the ratio of the inadequacy rate versus conventional cytology of ThinPrep in 'direct-to-vial' studies to be 0.70 (95% CI 0.39–1.27) and of SurePath to be 0.13 (95% CI 0.07–0.26).

Specimen interpretation time

A few studies have evaluated the impact of LBC on specimen interpretation time. LBC seems to be associated with shorter interpretation times than required for conventional cytology specimens. Payne *et al.* (2000), in their systematic review for the United Kingdom National Health Service, provided estimates of three minutes for LBC compared with 4–6 minutes for conventional cytology. This is not surprising given that the total surface area that needs to be screened is considerably less for both ThinPrep and SurePath than for conventional cytology specimens. The need for continuous adjustment to focus is also reduced using LBC, since the cells tend to be in the same plane of focus. With conventional cytology specimens, the screener needs to continually adjust the focus to evaluate clusters of cells. Payne *et al.* (2000) reported, however, that cytologists in Edinburgh found screening monolayers to require more intense concentration than screening conventional cytology specimens, making it more tiring. In part, this reflects the fact that occasionally only one or two HSIL cells are present

Table 22. Systematic reviews of comparative performance of LBC and conventional cytology

Author	Subgroup	Indicator	Key conclusions
Nanda *et al.* (2000)	ThinPrep	Histologically confirmed lesion	Higher sensitivity of LBC, but only three studies were evaluated
Payne *et al.* (2000)		Test positivity or histologically confirmed lesion	Some evidence that LBC offers an improvement in sensitivity
Bernstein *et al.* (2001)	ThinPrep	Test positivity	ThinPrep is as good as, or superior to, conventional cytology for diagnosing CIN
Hartmann *et al.* (2001)	All studies	Histologically confirmed lesion	Current evidence is inadequate to gauge whether LBC is "better" than conventional cytology
Sulik *et al.* (2001)		Histologically confirmed lesion	LBC demonstrated higher sensitivity (90%; 95% CI 77–96%) than conventional cytology (79%; 95% CI: 59–91%) for CIN 2 or more severe
Abulafia *et al.* (2003)	ThinPrep only	Test positivity	ThinPrep tends to be more sensitive than conventional smears in detecting CIN
Klinkhamer *et al.* (2003)	All studies	Histologically confirmed lesion	Indications that SurePath has lower sensitivity than conventional for ASC or greater No definitive statement can be made for detection of LSIL or higher or HSIL or higher for SurePath because of conflicting results Indications that ThinPrep has higher sensitivity than conventional for ASC or greater Likely that ThinPrep has higher sensitivity than conventional for LSIL or higher Likely that ThinPrep has higher positivity rate and greater absolute sensitivity than conventional for HSIL
Arbyn *et al.* (2004a)	Split-sample studies	Test positivity	More LSIL in LBC than in conventional cytology Positivity rates for ASC and HSIL not statistically different
	Direct-to-vial studies		More LSIL detected by LBC 80% (95% CI 52–112%) ThinPrep; 54% (95% CI 25–90%) SurePath More HSIL detected by LBC 72% (95% CI 42–108%) ThinPrep; 47% (95% CI 14–89%) SurePath Positivity rates for ASC were the same There was no reduction in positive predictive value for CIN 2 and CIN 3 of LBC versus conventional cytology

in an LBC specimen, necessitating careful scrutiny of every individual cell.

Availability of residual cellular material for molecular testing

One of the major benefits of LBC in many settings is the availability of residual cellular material for molecular testing for agents such as *Chlamydia trachomatis* or HPV. In the USA, the 2001 Consensus Guidelines for the Management of Women with Cytologic Abnormalities considered HPV DNA testing of residual LBC fluid to be the preferred approach to managing women with ASCUS cytological results (Wright *et al.,* 2002a), on the grounds that such reflex HPV DNA testing offers the advantage that women do not need to return to the office or clinic for an additional clinical examination. In addition, the 40–60% of women who are high-risk HPV DNA-negative will be spared a colposcopic examination and can be rapidly assured that they do not have a significant cervical lesion. A comprehensive study of triage methods for women with ASCUS indicated that reflex HPV DNA testing provides the same or greater life expectancy benefits and is more cost-effective than either a programme of repeat cytology or immediate colposcopy (Kim *et al.,* 2002).

Quality assurance and quality control issues

An advantage of cervical cytology over screening methods such as visual screening is that even though quality assurance and quality control programmes can be developed for both, the availability of archival glass slides facilities such programmes. Various definitions for quality control and quality assurance are used by laboratories. In general, *quality control* can be thought of in terms of the actual assessments that are done to ensure high quality and *quality assurance* can be thought of in terms of the entire process of maintaining minimum standards and continually striving for excellence. Quality assurance should be a coordinated effort that is designed to control, detect and prevent the occurrence of errors and hopefully to improve patient care. In general, there are three stages to the process of quality control (Bozzo, 1991):

- Setting standards for what one wishes to control and defining the benchmarks;
- Developing a mechanism for assessing what one wishes to control;
- Defining the response to be taken when deficiencies are identified.

For cervical cytology screening, quality assurance programmes can include a number of types of activity and should take into account country- and location-specific needs. What may be considered acceptable or even mandatory in one setting may serve simply to limit the availability of screening in other settings. It is critical, however, that any cervical cytology laboratory or programme have an established quality assurance programme. In general, it is preferable for cytology services to be centralized as much as possible, to facilitate quality assurance. The use of computerized data collection systems that can integrate cytological findings, histological findings and follow-up information is highly desirable (Miller *et al.,* 2000).

Preanalytic quality control

Preanalytic quality control measures include the records that laboratories should maintain relating to specimen receipt, preparation of specimens, staining of specimens and upkeep of equipment and microscopes, as well as records of personnel and their training and education.

Training

Training of both the cytotechnicians who perform the initial screening in the laboratory and the pathologists who provide the final interpretation is critical to obtaining optimal performance of a cervical cytology programme. Cytopathologists should either receive formal training in an established academic programme or be trained in an established national centre for cervical cytology for at least six months (Miller *et al.,* 2000). This training should typically include not only the interpretation of cervical cytology specimens, but also cervical histopathology. A cytopathologist who will run a laboratory is generally selected for leadership potential and ability to organize, run and manage a successful cytopathology laboratory, and will require training in laboratory management skills.

In most developed countries, cytotechnicians undergo 1–2 years of formal didactic training in order to develop a high level of competence in evaluating all types of cytological specimens, including gynaecological cytology. However, in some countries consideration is now being given to intensive six-month training programmes focusing only on gynaecological cytology. In addition, cytotechnicians should periodically participate in competence-based education programmes. Unfortunately, cytotechnology training programmes are not available in many developing countries and extended formal training programmes are not an option. In these settings, cytotechnicians are often 'bench-trained', being tutored by a person with some level of training in interpreting gynaecological cytology specimens. Training in this manner should be avoided unless the laboratory where it occurs processes at least 15 000 specimens annually and training should last for at least six months (Miller *et al.,* 2000). Although there is little evidence that cytotechnicians who are trained in such a 'hands-

on' fashion perform less well than those who receive formal training, the variability in training inherent in this approach is a cause for concern. Whenever possible, cytotechnicians should receive formal, structured, competence-based training in interpreting cervical cytology specimens. The International Federation of Cytology has an international qualification for cytotechnicians, which can be used to ensure that competence has been obtained.

Workload limits – maximum and minimum

It is now widely accepted that, because of the repetitive nature of screening cytology specimens, there should be workload limits on the number of specimens that a cytotechnician can screen in any given period. In the USA, federal regulations require that anyone performing primary screening of cervical cytology specimens should evaluate no more than 100 cytology specimens per 24-hour period and in not less than eight hours (Federal Register, 1992). In addition, every laboratory must establish individual workload limits for each cytotechnician, based on their experience and skill. This must be reassessed every six months using laboratory-defined performance standards. In many European countries, this workload limit is considered too high and other limits are used. In the United Kingdom, for example, time limits rather than slide limits are used. Cytotechnicians are restricted to screening for only four hours per day, regardless of whether they are screening conventional or LBC specimens. Since LBC specimens can be screened more rapidly than conventional cytology specimens, this means that greater numbers of LBC specimens can be screened by each cytotechnician. A recent consensus panel recommended that a daily workload limit of 60 cases was preferable (Miller *et al.,* 2000).

It is also important that a laboratory process a minimum number of specimens per year in order to maintain an adequate level of competence (Krieger *et al.,* 1998). In reviews of US laboratories by the College of American Pathologists, screening error rates were found to be greatest in laboratories processing less than 5000 specimens per year and having no dedicated screening cytotechnicians (College of American Pathologists, 1997). In the United Kingdom, laboratories are now required to process at least 15 000 specimens per year. Evaluation of a minimum annual number of specimens is also to be considered desirable in low-resource settings. The Peruvian Society of Cytopathology does not certify laboratories that process under 5000 specimens annually (Salvetto & Sandiford, 2004). A recent World Health Organization consensus panel recommended that each laboratory should process at least 20 000 specimens yearly in order to maintain acceptable skills (Miller *et al.,* 2000).

Review of abnormal cases

It is generally accepted that a pathologist should review all specimens deemed by the screening cytotechnician to have any degree of cytological abnormality (American Society of Cytopathology, 2001). Identification of discordant cases provides an element of quality control for the screening process and allows identification of specific cytotechnicians and specific areas of cytology requiring additional education. It is important for quality monitoring that all reviews be documented.

Rescreening of negative cases

Some form of rescreening of specimens initially considered negative is important for quality control. In the USA, federal regulations stipulate that at least 10% of all samples interpreted as negative by each cytotechnician must be reassessed by either a pathologist or a supervising cytotechnologist before the result is reported (Federal Register, 1992). This regulation is controversial for a variety of reasons. One is the level of discrepancy that is considered significant. It has been argued that negative specimens classified as atypical (e.g., ASC) upon review should not be considered errors, because of the inherent subjectivity of this diagnosis (Krieger *et al.,* 1998). Another problem with performing 10% rescreening of negative cases is that significant lesions are quite uncommon in the reviews. Given an underlying rate of SIL of only 2–3% in the screened population and assuming that even a poorly performing cytotechnician will be able to identify 75% of specimens containing SIL, large numbers of specimens must be rescreened to determine which cytotechnicians or laboratories are performing poorly. This lack of statistical power greatly hampers its use as a quality control measure (Hutchinson, 1996).

The technique of rapid rescreening of all negative specimens has been the subject of a number of studies and appears to present an attractive alternative to the 10% rescreening approach. Using this technique all, or most, of the specimens classified as negative by a laboratory undergo a second, more rapid evaluation by a different screener. This is the approach to rescreening adopted in the United Kingdom by the National Health Service (NHSCSP, 2000). Another approach is referred to as 'prescreening', in which all specimens undergo a rapid review before the intensive screening. In a recent meta-analysis of published data, Arbyn *et al.* (2003) found that the pooled estimated sensitivity of rapid prescreening for HSIL or more severe lesions was 86% and that the technique showed diagnostic properties that support its use as a quality

control measure. The same group previously demonstrated that rapid prescreening was superior to 10% random rescreening in identifying cases that were missed (Arbyn & Schenck, 2000).

Cytology–histology correlations and clinical follow-up

If a laboratory has access to histological specimens obtained at the time of colposcopy for an abnormal cytological finding, it should compare all premalignant and malignant cytological results with the histopathological observations. This allows the laboratory to refine its cytological criteria. If histological specimens are not available, the laboratory may attempt to obtain referral and follow-up information. It is important that the laboratory obtain follow-up information on women with HSIL to ensure that they have not been lost to follow-up.

Measuring the performance of the laboratory

Laboratories need to carefully and continuously monitor their performance as a whole, as well as that of individual cytotechnicians. Information that can be useful for a given laboratory includes the percentages of specimens classified as having a given result (e.g., ASC, LSIL, HSIL, etc.), the rate of unsatisfactory specimens, the ASC:LSIL ratio, the laboratory turnaround time, etc. One of the most important measures is screening sensitivity (Krieger *et al.,* 1998; NHSCSP, 2000; American Society of Cytopathology, 2001). However, it is very difficult to determine the sensitivity of screening in a real-world laboratory setting. One approach that has been proposed to estimate screening sensitivity in a laboratory is to calculate the 'false negative proportion', which is essentially the number of false negative LSIL or greater specimens identified through a 100% rapid rescreen programme divided by the total number of LSIL or greater specimens identified through regular screening and the rapid rescreen process combined (Krieger *et al.,* 1998).

Proficiency testing

Proficiency testing programmes provide laboratories, cytopathologists and individual cytotechnicians with sets of stained cytology specimens on which the interpretation has been agreed to according to a set procedure. The slide sets are then evaluated by the person being tested and their interpretation is compared with that of the panel or with their peers (Coleman & Evans, 1999; NHSCSP, 2000). This allows the performance of both whole laboratories and individual cytotechnicians or cytopathologists to be compared against others in an unbiased manner. Periodic retesting should be conducted every 6–12 months (Miller *et al.,* 2000). Individuals who perform poorly on proficiency testing should receive additional training to improve their skills and any who continue to perform poorly after retraining should be reassigned to non-screening tasks.

Recent evidence suggests that performance on proficiency testing provides some evidence of the real-world performance of cytotechnicians (Keenlyside *et al.,* 1999). A recent report from Peru and Nicaragua has shown that proficiency testing can be implemented successfully in developing countries (Salvetto & Sandiford, 2004).

Visual inspection

The use of visual inspection methods to screen for cervical neoplasia began with the use of Schiller's test in the 1930s (Schiller, 1933). In the 1980s, the idea of looking at the cervix with the naked eye for early detection of disease (known as 'down-staging') in low-resource settings was promoted (Stjernswärd, 1987). Over the last ten years, the use of dilute (3–5%) acetic acid applied to the cervix before inspection (visual inspection with acetic acid, VIA) has been investigated. More recently, the application of Lugol's solution has been used and is referred to as visual inspection with Lugol's iodine (VILI).

Visual tests are inherently subjective. Published studies of the test performance characteristics vary with regard to important methodological aspects that result in biases and other difficulties in generalizing the findings to other populations. For example, studies may use different definitions of test positivity. Differences in training of personnel and in the light sources used also generate variability in test performance characteristics across different study settings. Varying abilities of colposcopists to detect lesions and of pathologists to interpret histology accurately also affect the assessment of test performance.

Colposcopy with directed biopsy is the usual reference standard by which the performance of visual tests is assessed, but biases may impair the validity of the assessment. Verification bias arises if colposcopy is not applied equally to all women because of the study design. Blinding between those performing the visual test under evaluation and those performing the reference colposcopy is crucial to avoid information or expectation bias. Test performance for detection of CIN 2 or worse lesions and potential biases of all studies reviewed are summarized below.

Unaided visual inspection (VI)

Visual inspection (VI) (also called 'down-staging' or 'unaided visual inspection') consists of a clinical examination of the cervix using only a speculum and a light source. Test positivity is defined by the presence or

absence of specific characteristics, usually with low and high thresholds of positivity (Table 23). Only one of the six published studies reporting test characteristics of VI (Table 24) did not suffer from obvious verification bias (see Glossary and Chapter 4) (Basu *et al.,* 2002); it found sensitivity to be low (< 50%) irrespective of the threshold used to define test positivity. The other five studies used cytology as the reference standard. In all studies, the high-threshold definition of test positivity (corresponding to a 5–10% positivity rate) was associated with rather low sensitivity (30–60%). Gains in sensitivity using the low-threshold definition of test positivity led to concomitant decreases in specificity. Thus it is clear that VI lacks sufficient sensitivity for use as a primary screening test.

Visual inspection using acetic acid (VIA)

VIA involves naked-eye inspection of the cervix one minute after application of a 3–5% solution of acetic acid using a cotton swab or a spray. Test positivity is based on the appearance of acetowhite areas in the transformation zone, close to the squamocolumnar junction or the os. The cervix is examined using a bright light source such as a torch or halogen focus lamp. VIA is also known as direct visual inspection (DVI), acetic acid test (AAT) and cervicoscopy.

Dilute acetic acid causes what is thought to be a reversible coagulation of intracellular proteins, resulting in noticeable opacity and a decrease in the usual reddish hue imparted by the subepithelial vasculature. This effect, called acetowhitening, is not specific to cervical neoplasia and may also occur in immature squamous metaplasia and in inflamed, regenerating cervical epithelium. The degree of opacity due to the acetowhite reaction varies according to the thickness of the neoplastic change present in the epithelium and thus according to the grade of intraepithelial neoplasia.

The most common features observed using VIA are summarized in Table 25. VIA results are reported using negative and positive categories. VIA-positive cervices are illustrated in Figure 42.

In 17 published studies, test positivity rates ranged from 3% to 53% (Table 26). Seven studies were designed to minimize verification bias. In two other studies (Denny *et al.,* 2000a, 2002), only women negative by cytology, VIA, HPV DNA testing and cervicography were not subjected to colposcopy, reducing susceptibility to bias. Seven of these nine studies (Londhe *et al.,* 1997; University of Zimbabwe/JHPIEGO Cervical Cancer Project, 1999; Denny *et al.,* 2000a; Belinson *et al.,* 2001; Denny *et al.,* 2002; Cronjé *et al.,* 2003; Sankaranarayanan *et al.,* 2004a), accounting for more than 95% of the total sample size, reported sensitivities of approximately 75%.

One study used a gold standard for enhanced disease ascertainment that was based on directed biopsy of any abnormal area(s), four-quadrant biopsies and endocervical curettage (ECC) in all women (Belinson *et al.,* 2001). The other study that had enhanced design features was based on 55 000 women enrolled at 11 sites in six West African countries and India (Sankaranarayanan *et al.,* 2004a). Each site followed a common testing protocol that included VIA, VILI and colposcopy with directed biopsy, as required, performed by separate individuals. Although similar training methods and test result definitions were used, there was substantial variation in the reported positivity rates (7–27%), sensitivity (56–94%) and specificity (74–94%).

Numerous studies have shown VIA to have sensitivity similar to that of cervical cytology for identifying women with HSIL, but much lower specificity (Table 27). Only two studies compared the accuracy of VIA and HPV DNA testing (Table 28); these showed the two tests to have similar accuracy. The reproducibility of VIA has been documented to be equivalent to that of histology, cytology and colposcopy (Sellors *et al.,* 2002). In the multicentre study in Africa and India, the agreement between master trainers and local providers using 36 cervical photographs was fair (raw agreement, 64.5%; kappa, 0.38) (Sankaranarayanan *et al.,* 2004a).

Table 23. Test definition for visual inspection

Test definition	Characteristics
Normal	Normal-looking cervix, nabothian cysts
Positive (low threshold)	Cervicitis, erosion, polyp, wart, unhealthy cervix, reddish-looking cervix
Positive (high threshold)	Low-threshold features plus bleeding on touch, bleeding erosion, hypertrophied elongated cervix, growth, ulcer

Visual inspection using acetic acid with low-level magnification (VIAM)

VIAM is VIA with low-level magnification (2–4 x), using a hand-held device to inspect the cervix one minute after application of acetic acid. Table 29 presents test results from four studies comparing VIA and VIAM. None of these studies documented any significant difference in test performance characteristics between VIAM and VIA.

Table 24. Studies of visual inspection[a]

Study	Sample size	Population (age, recruitment, location)	Provider	Reference diagnosis	Positivity rate (%)	Sensitivity (%)	Specificity (%)	Comments
Singh *et al.*[b] (1992)	44 970	Opportunistic India		Cytology/ histology	(H) 11 (L) 69	63	89	Verification bias
Bhargava *et al.*[b] (1993)	3608	Opportunistic India	Midwife	Cytology/ histology	(H) 5 (L) 65	25 92	96 37	Verification bias
Sujathon *et al.* (1995)	3602	30+ Opportunistic and referred India	Cyto-pathologist	Cytology	(L) 63	89	50	Not designed for accuracy estimation
Nene *et al.*[c] (1996)	2135	35–60 Community-based	Health worker	Cytology/ histology	(H) 6 (L) 57	60 90	94 43	Verification bias
Wesley *et al.* (1997)	2843	30+ Opportunistic India	Health worker	Cytology/ histology	(H) 6 (L) 45	29 66	94 55	Verification bias
Basu *et al.* (2002)	6399	Community-based India	Health worker	Colposcopy/ histology	(H) 7 (L) 25	32 49	93 76	

[a] Some test characteristics of the table are not exactly those reported in corresponding publications. Estimates have been computed when they were not provided or have been corrected to achieve comparability between studies. This correction was performed to take into account differences in study design or analysis, due to various factors: different threshold of test positivity, different disease definition, only a subset of the population used for estimation of characteristics, improper computation method, etc.
[b] Detection of any lesions
[c] Detection of cancer
Test positivity was defined at high-threshold (H) and at low-threshold (L), the sensitivity was estimated with the threshold CIN2–3 unless otherwise specified
The study with no verification bias is highlighted

A correlation study, with a sample size of 2080 previously screened women and a positivity rate of approximately 5% for VIAM, reported poor associations between VIAM and HPV test positivity and between VIAM and cytology test results (Rodriguez *et al.,* 2004).

Visual inspection using Lugol's iodine (VILI)

The use of Lugol's solution to aid inspection of the cervix with the naked eye was described in 1933 by Schiller, but fell into disuse as cytological testing became available (Schiller, 1933; Wright, 2003). Several decades later, research on visual inspection methods led to the observation that nurses and midwives recognized non-staining areas on the cervix after application of Lugol's solution more readily than acetowhite areas (Sankaranarayanan & Wesley, 2003), which led to renewed interest in this technique (referred to in the past as 'Lugol's iodine test' and 'Schiller's iodine test').

Lugol's iodine stains glycogen stored in cervical epithelial cells. Mature squamous epithelium stores more glycogen than either columnar epithelium or immature squamous metaplastic epithelium. The application of iodine solution to the cervix thus results in black or dark brown staining of mature squamous epithelium. Columnar epithelium does not stain and retains its reddish hue. Areas of immature metaplasia stain a very light brownish hue, if at all. Neoplastic squamous epithelium

Table 25. Test definition of visual inspection with acetic acid

Denomination	Possible thresholds			Characteristics
	A	B	C	
Normal	negative	negative	negative	Normal looking cervix: no white lesion, smooth, uniform, featureless Atypical cervix: ectopion, polyp, cervicitis, inflammation, Nabothian cysts
Indeterminate				Severe inflammation or cervicitis so that cervix cannot be adequately assessed for acetowhite lesion
Ill-definite lesion	positive	positive		Pale white lesion (acetowhite lesion), poorly circumscribed and faintly acetowhite Focal, small punctuated areas of acetowhitening usually involving the transformation zone
Definite lesion			positive	Dense white lesion with sharp border; one border abutting the squamo-columnar junction
Suspicious cancer				Cervical ulcer or growth cauliflower-like growth or ulcer Fungation mass

contains little or no glycogen and does not stain with Lugol's iodine, taking a bright mustard or saffron yellow colour. Atrophic epithelium stains partially with Lugol's iodine, which makes interpretation difficult in postmenopausal women. A condylomatous lesion may not stain or only partially stain with Lugol's iodine. Areas of leukoplakia (hyperkeratosis) and areas partially denuded of squamous epithelium do not stain with iodine and remain colourless in a surrounding black or dark brown background. Results of VILI are categorized in Table 30. Images of VILI-positive cervices are shown in Figure 43.

The single published report of VILI test characteristics (Sankaranarayanan *et al.,* 2004a) involved 54 981 women aged 25–65 years. The reference standard was colposcopy-directed biopsy. VIA and VILI were performed independently by blinded individuals in order to minimize information (expectation) bias. In this setting, VILI was more sensitive than VIA and equally specific (Figure 44). The reproducibility of

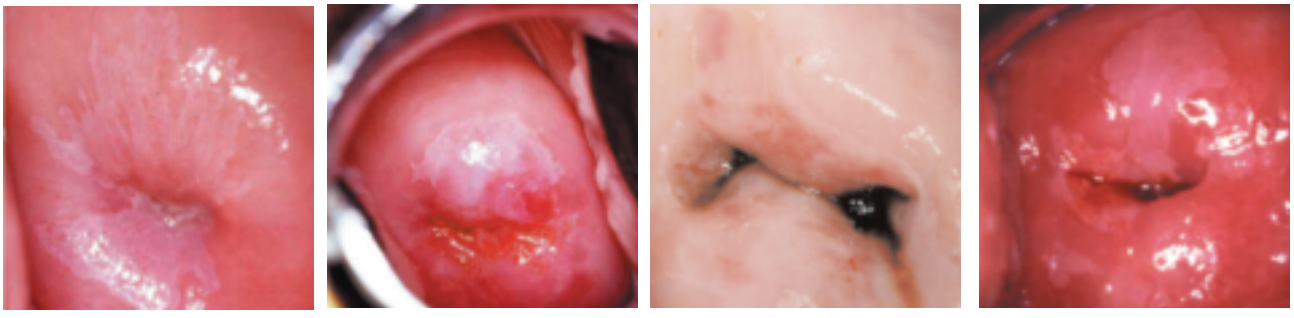

Figure 42 Example of VIA-positive lesions

Table 26. Studies of visual inspection with acetic acid [a]

Study	Sample size	Population (age, recruitment, location)	Provider	Reference diagnosis	Positivity rate (%)	Sensitivity (%)	Specificity (%)	Comments
Slawson *et al.* (1992)	2690	15–45 Family practice USA	Clinician	Colposcopy/histology for Pap+ or VIA+	3	29	97	Not designed for test accuracy estimation
Cecchini *et al.* (1993)	2036	17–83 Opportunistic Italy	Smear-taker	Colposcopy/histology for Pap+, VIA+ or or cervicography+	25	88	75	Not designed for test accuracy estimation
Megevand *et al.* (1996)	2426	20–83 Opportunistic South Africa	Nurse	Colposcopy/histology for Pap+ or VIA+	3	65	98	Not designed for test accurary estimation
Londhe *et al.* (1997)	372	Opportunistic India	Clinician	Colposcopy	53	78	49	Only 74% of the enrolled patient underwent colposcopy
Sankaranarayanan *et al.* (1998)	2935	20+ Opportunistic India	Cytotech.	Colposcopy/histology for Pap+ or VIA+	10	87	91	Not designed for test accuracy estimation
University of Zimbabwe/ JHPIEGO (1999)	2148	25–55 Opportunistic Zimbabwe	Midwife	Colposcopy/histology	40	77	64	
Sankaranarayanan *et al.* (1999)	1268	22–70 Opportunistic India	Nurse	Colposcopy/histology for Pap+ or VIA+	36	95	68	Not designed for test accuracy estimation
Denny *et al.* (2000)	2944	Opportunistic South Africa	Nurse	Colposcopy/histology for Pap+, VIA+, HPV+ or cervicography+	18	67	83	Minimal verification bias
Cronjé *et al.* (2001)	6298	Mean age: 34 Opportunistic South Africa	Nurse	Histology for VIA+ and 20% of VIA-	18	50	84	Unbiased estimation not possible from the published data

Table 26 (contd)

Study	Sample size	Population (age, recruitment, location)	Provider	Reference diagnosis	Positivity rate (%)	Sensitivity (%)	Specificity (%)	Comments
Belinson *et al.* (2001)	1997	35–45 Opportunistic China	Gynaecol.	4 quadrant histology + ECC	28	71	74	
Singh *et al.* (2001)	402	Mean age: 37 Referred Symptomatic India	Gynaecol.	Colposcopy/histology	42	87	82	
Denny *et al.* (2002)	2754	35–65 Opportunistic South Africa	Nurse	Colposcopy/histology for Pap+, HPV-, VIA+ or cervicography+	25	70	79	Minimal verification bias
Rodriguez-Reyes *et al.* (2002)	376	19–45 Opportunistic Mexico		Histology	48	92	59	
Cronjé *et al.* (2003)	1093	21–65 Opportunistic South Africa	Nurse	Histology	53	79	49	
Ngelangel *et al.* (2003)	3316	25–65 Opportunistic Philippine	Nurse	Colposcopy/histology	10	37	91	
Tayyeb *et al.* (2003)	501	30–60 Opportunistic Pakistan		Colposcopy/histology for Pap+ or VIA+	31	94	78	Verification bias
Sankaranarayanan *et al.* (2004a)	54 981	25–65 Opportunistic India, Africa (11 studies)	Midwife	Colposcopy/histology	16 [R[b]:7–27]	77 [R[b]:56–94]	86 [R[b]:74–94]	

[a] Some test characteristics of the table are not those reported in the corresponding publications. Estimates have been computed when they were not provided or have been corrected to achieve comparability between studies. This correction was performed to take into account differences in study design or analysis, due to various factors: different threshold of test positivity, different disease definition, only a subset of the population used for estimation of characteristics, improper computation method, etc.
[b] R stands for the range within the studies reported

Outcome threshold CIN 2–3
The studies with no verification bias are highlighted

Table 27. Comparison of VIA and cytology accuracy in published studies[a]

Study	Sample size	Positivity rate (%)		Sensitivity (%)		Specificity (%)	
		VIA	Cytology	VIA	Cytology	VIA	Cytology
Slawson *et al.* (1992)	2690	3	6	29	87	97	95
Cecchini *et al.* (1993)	2036	25	4	88	63	75	96
Megevand *et al.* (1996)	2426	3	13	65	100	98	88
Londhe *et al.* (1997)	372	53	6	78	22	49	95
Sankaranarayanan *et al.* (1998)	2935	10	10	87	86	91	91
University of Zimbabwe/ JHPIEGO (1999)#	2148	40	13	77	44	64	91
Sankaranarayanan *et al.* (1999)	1268	36	16	95	62	68	87
Denny *et al.* (2000)	2944	18	15	67	80	83	87
Cronjé *et al.* (2001) #	6298	18	2	50	19	84	99
Singh *et al.* (2001)	402	42	42	87	81	82	79
Denny *et al.* (2002) #	2754	25	70	57	79	96	
Ngelangel *et al.* (2003)	3316 (VIA)						
	3195 (Cytology)	10	2	37	14	91	98
Tayyeb *et al.* (2003)	501	31	16	94	47	78	89
Cronje *et al.* (2003) #	1093	53	9	79	53	49	95
Sankaranarayanan *et al.*[b] (2004b)	22 663	17	9	72	65 [R[c]: 38–81]	84	92 [R[c]: 86–99]

[a] Some test characteristics of the table are not the ones reported in corresponding publications. Estimates have been computed when they were not provided or have been corrected to achieve comparability between studies. This correction was performed to take into account differences in study design or analysis, due to various factors: different threshold of test positivity, different disease definition, only a subset of the population used for estimation of characteristics, improper computation method, etc.

[b] Subset of five Indian studies from Sankaranarayanan *et al.*(2004a)

[c] R stands for the range within the studies reported

Cytology threshold: ASCUS+, unless otherwise indicated (# , LSIL+)

Outcome threshold: CIN 2–3

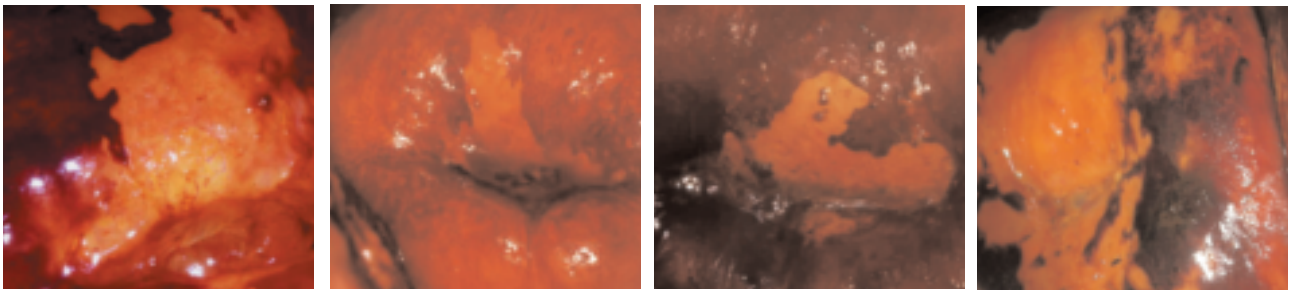

Figure 43 Example of VILI-positive lesions

Table 28. Comparison of VIA and HPV testing accuracy in published studies

Study	Sample size	VIA			HPV testing		
		Positivity rate (%)	Sensitivity (%)	Specificity (%)	Positivity rate (%)	Sensitivity (%)	Specificity (%)
Denny *et al.* (2000)	2944	18	67	83	16	73	86
Sankaranarayanan *et al.* (2004c)[a]	18 085	11	65 [R[b]: 54–79]	89 [R[b]: 89–90]	7 [R[b]: 6–9]	65 [R[b]: 45–81]	94 [R[b]: 92–95]

Outcome threshold: CIN 2–3

[a]Subset of five Indian studies from Sankaranarayanan *et al.* (2004a)

[b]R stands for the range within the studies reported

Table 29. Comparison of accuracy of visual inspection with acetic acid, with or without magnification (VIA and VIAM), in published studies

Study	Sample size	VIA			VIAM			
		Positivity rate (%)	Sensitivity (%)	Specificity (%)	Device	Positivity rate (%)	Sensitivity (%)	Specificity (%)
Denny *et al.* (2000)	2944	18	67	83	x 2.5 (hand-held device)	18	67	83
Denny *et al.* (2002)	2754	25	70	79	x 4.5 Aviscope	27	74	77
Ngelangel *et al.* (2003)	3316 (VIA) 3447 (VIAM)	10	37	91	Speculo-scope (6 x 16 magnification)	11	34	90
Sankaranarayanan *et al.* (2004d)	16 900 (3 studies)	14 [R[a]: 11–19]	[R[a]: 56–71]	[R[a]: 82–90]	x 4 magnifi-cation Hand-lens (2 studies) and Aviscope (1 study)	14 [R[a]: 11–18]	64 [R[a]: 61–71]	87 [R[a]: 83–90]

Outcome threshold: CIN 2–3

[a] R stands for the range within the studies reported

VILI appears to be greater than that of VIA.

Quality control for visual inspection tests

The substantial variability in test performance characteristics of visual inspection tests reflects, at least in part, the subjective nature of visual inspection. Definitions of test result categories should be standardized to improve reproducibility. Due to the subjective nature of visual inspection methods, it is difficult to maintain the quality of assessment among trained staff. Adequate training, routine process measurements (e.g., test positivity rates, histological confirmation rates) and on-site supervision are critical to support high-quality visual inspection-based screening services. Although reliable methods of correlating daily competence with proficiency testing results have not been

Table 30. Categorization of visual inspection with Lugol's iodine (VILI) test results

VILI-negative	Patterns include the following: • A normal pattern of dark brown or black staining of the squamous epithelium and no change in colour of the columnar epithelium, or • patchy, indistinct, ill-defined, colourless or partially brown areas, or • pale areas of no or partial iodine uptake present on polyps, or • a leopard-skin appearance (associated with *T. vaginalis* infection), or • pepper-like, non-iodine uptake areas seen in the squamous epithelium, far away from the squamocolumnar junction, or satellite, thin, yellow, non-iodine uptake areas with angular, or digitating margins, resembling geographical areas seen far away from the squamocolumnar junction
VILI-positive	Presence of a dense, thick, bright, mustard-yellow or saffron-yellow iodine non-uptake area seen in the transformation zone close to or abutting the squamocolumnar junction or, in the absence of a visible squamocolumnar junction, close to the os, or when the entire cervix turns bright yellow
VILI-positive, invasive cancer	Frank, nodular, irregular, ulceroproliferative growth visible on the cervix, which turns densely yellow on application of iodine

Adapted from Sankaranarayanan & Wesley (2003)

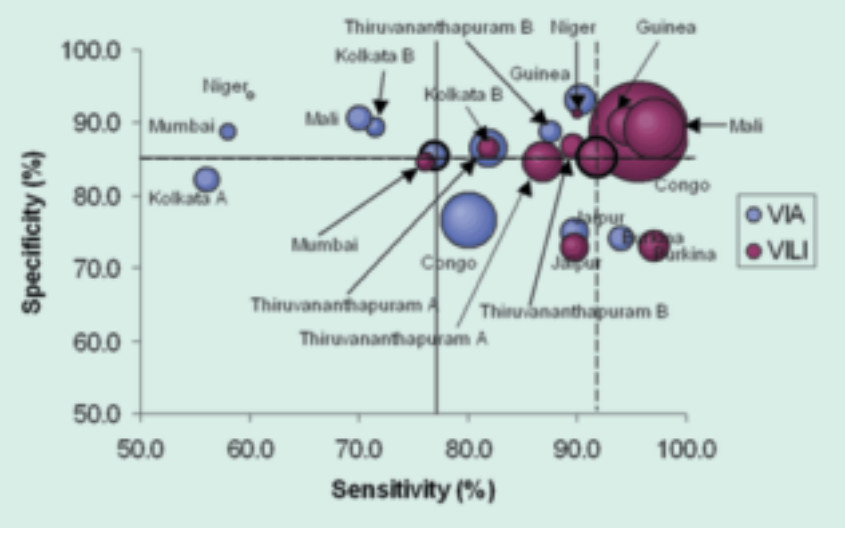

Figure 44 Plot of sensitivity and specificity for VIA and VILI for each of the studies. The size of the bubble reflects the precision of the estimates. The bigger the bubble, the lower the variance of the sensitivity and specificity and the higher the precision due to a larger study sample size. The bubbles with thick borders represent the pooled estimates. (Adapted from Sankaranarayanan *et al.*, 2004a)
Note: bubble size = $k/(d_{se}^2 + d_{sp}^2)$, where k is a constant, d is the difference between the lower and upper confidence limits, se is sensitivity and sp is specificity.

established for cytological screening programmes (Vooijs *et al.,* 1998), visual inspection-based screening programmes may utilize periodic assessments of practitioners' skills using collections of VIA and VILI cervical photographs.

There is no consensus on the number of visual inspections that should be performed correctly by an individual before he/she is deemed competent, nor on the minimum daily rate that is required to maintain skills. The periodic computation of test positivity rates for visual inspections performed by each individual may be useful for monitoring visual inspection. However, because there is no permanent record when visual assessments are made by VIA or VILI, unless a photographic image is taken for subsequent review by a supervisor (Sellors *et al.,* 2002; Wright, 2003), these positivity rates are not verifiable by audit. Visual inspection

methods coupled with immediate cryosurgical treatment for test-positive cases ('see and treat') have also been suggested for cervical screening in low-resource settings (Denny *et al.,* 2002, Gaffikin *et al.,* 2003). Because biopsies to exclude invasive carcinoma are not performed before ablative treatment, the 'see and treat' approaches carry the risk of undertreatment of invasive carcinoma and reduced opportunities to diagnose potentially curable invasive disease. It has been argued that, outside of research settings, it may not be feasible to monitor the safety or confirm the effectiveness of 'see and treat' programmes (Suba *et al.,* 2004).

Advantages and potential hazards of visual inspection methods

Concerns about personal modesty and discomfort caused by the vaginal speculum are common to all screening techniques. Local irritation of tissues and allergic reactions to iodine or vinegar are rare. In both VIA and VILI, staining of the epithelium is temporary, although iodine staining lasts longer (up to 45 minutes) than acetowhitening (Sankaranarayanan *et al.,* 2004a).

Application of visual inspection methods should probably be restricted to women under the age of 50 years. With increasing age, the squamocolumnar junction migrates inward from the readily visible portion of the ectocervix towards the endocervical canal, so lesions probably become more difficult to identify with visual methods in older women. In addition, the accuracy of visual inspection may be highly dependent on the underlying prevalence of sexually transmitted diseases, which may increase the level of inflammation and render visual inspection difficult to assess.

Visual inspection-based tests are simple, safe and well accepted. They require a very low level of infrastructure and can be performed by a wide range of personnel, such as doctors, nurses, midwives, paramedical workers and trained non-medical personnel, after a short period of training (1–3 weeks). In addition, results are available without delay, allowing immediate referral for confirmatory testing. However, if immediate treatment is performed, high rates of overtreatment may result, given the relatively low test specificity of VIA and VILI.

Colposcopy

Colposcopy is a procedure that allows illuminated stereoscopic and magnified (typically 6–40 x) viewing of the cervix and the vagina. For colposcopy, the woman is placed in the lithotomy position, with the cervix exposed with a bivalve speculum in place, and various solutions (normal saline, 3–5% dilute acetic acid and Lugol's iodine) are applied to the cervical epithelium in sequence. A green filter is rarely used except when the subepithelial vascular pattern is examined (Jordan, 1985; Sellors & Sankaranarayanan, 2003). The aim is to examine the transformation zone, an area bounded laterally by the original squamocolumnar junction, in which metaplastic squamous epithelium develops, the medial or internal border being defined by the new squamocolumnar junction. This latter junction defines the upper limit of the squamous metaplastic process, which in certain conditions may become abnormal. When such abnormal areas develop within this zone, they are graded according to morphological features, namely, acetowhiteness, margins, blood vessels and iodine uptake.

Hinselmann (1925) first described colposcopy. The modern colposcope is a binocular microscope with a variable-intensity light source providing a stereoscopic view of the cervix, with a field of view and depth of focus that vary inversely with the magnification selected.

Provision of a high-quality colposcopic service requires the availability of a trained colposcopist and access to a competent histopathologist able to perform assessment of removed biopsy material. The findings from video colposcopy seem to agree with those obtained with traditional optical colposcopy (Ferris *et al.,* 2000b).

Colposcopic findings

Terminology to describe the morphological findings in a standard fashion has evolved over the years (Dexeus *et al.,* 1977; Jordan, 1985; Sellors & Sankaranarayanan, 2003; Walker *et al.,* 2003) (Table 31). Many of the qualitative descriptions have been quantified as to the degree of abnormality and have been combined into a scoring system (Table 32) that is used by many colposcopists to grade abnormal squamous epithelial areas (Reid & Scalzi, 1985; Reid, 1993). The uncommon glandular epithelial lesions tend to be more difficult to diagnose and appear as strikingly dense acetowhite or milky-white areas compared to the surrounding villi of columnar epithelium. Microinvasive and frankly invasive squamous cancers are densely acetowhite with markedly atypical blood vessels (bizarre, irregular branching and gross fluctuations in calibre and course). The surface configuration gradually changes from small protuberances, excrescences or microconvolutions in microinvasive cancer to frankly invasive cancer with strikingly raised edges, irregular surface contour, and strikingly bizarre blood vessels that bleed spontaneously or on touch.

Histological confirmation

Biopsies are obtained under colposcopic visualization from the locations with the most severe changes, in order to histologically confirm the degree of severity of the neoplastic process. Since it is essential to rule out the presence of cancer, it is standard practice in some settings to obtain a

Table 31. International terminology for colposcopy from the International Federation for Cervical Pathology and Colposcopy

I	Normal colposcopic findings Original squamous epithelium Columnar epithelium Transformation zone
II	Abnormal colposcopic findings Flat acetowhite epithelium Dense acetowhite epithelium[a] Fine mosaic[a] Fine punctuation Coarse punctuation[a] Iodine partial positivity Iodine negativity[a] Atypical vessels[a]
III	Colposcopic features suggestive of invasive cancer
IV	Unsatisfactory colposcopy Squamocolumnar junction not visible Severe inflammation, severe atrophy, trauma Cervix not visible
V	Miscellaneous findings Condylomata Keratosis Erosion Inflammation Atrophy Deciduosis Polyps

From Walker *et al.* (2003)
[a] Major changes

histological sample from the endocervical canal if the new squamocolumnar junction (and thus the entire transformation zone) cannot be examined. Debate continues as to whether histological sampling of the endocervical canal should be performed routinely in all women undergoing colposcopic examination or only in circumstances such as when the new squamocolumnar junction cannot be seen or the colposcopic examination is deemed to be unsatisfactory. It is also suggested to be used when the colposcopic examination is satisfactory but a cytological test indicates a higher grade of lesion (Spirtos *et al.,* 1987; Fine *et al.,* 1998; Pretorius *et al.,* 2004).

The colposcope can also be used to assess the remainder of the lower genital tract (vagina, vulva and perianal skin), especially if no cervical lesion is found in a woman with abnormal cytology. Women who are HIV-positive tend to have multifocal disease involving the vagina, vulva and perianal areas, and therefore these regions need to be examined (Abercrombie & Korn, 1998).

Primary screening and diagnosis

Colposcopy continues to be used routinely as part of a standard gynaecological examination by many clinicians in some European and Latin American countries, probably as a result of the long-standing tradition rooted in German medical teaching from the time of Hinselmann in Hamburg (Jordan, 1985; Dexeus *et al.,* 2002). When colposcopy has been evaluated for primary screening, it has been usually accompanied by simultaneous cytology (Dexeus *et al.,* 2002). The rationale behind this combined testing approach is that it decreases false negative and false positive rates associated with cytology alone and also reduces the need for call-back for repeat cytology, the colposcope being used as a guide to collection of the cytology specimens (Van Niekerk *et al.,* 1998). Within Germany at least, there is some reluctance to support the continued use of colposcopy as a screening tool to assist in the taking of cytological specimens, since there is no evidence that the quality of smears is improved (Hilgarth & Menton, 1996). Furthermore, constraints limiting the application of colposcopy to universal screening include its high cost relative to cytology, the availability and accessibility of adequately trained colposcopists, and the lower ability of colposcopy to detect endocervical lesions (Van Niekerk *et al.,* 1998; Belinson *et al.,* 2001).

Since colposcopy was introduced in the 1960s to the English-speaking world, it has been selectively applied for diagnosis in women who are referred because of an abnormal cytological test. Current indications for colposcopy are listed in Table 33.

Biases and caveats in the assessment of colposcopy

Most assessments of the sensitivity and specificity of colposcopy and directed biopsy are susceptible to bias. The colposcopic impression confounds the reference standard of diagnosis (histology) since it dictates where the histological specimen is obtained from, leading to an inflated estimate of the accuracy of colposcopy. In contrast to

Table 32. Combined colposcopic index, commonly used to score and document abnormal areas seen on colposcopic examination

Colposcopic sign	Zero points	One point	Two points
Colour	Less intense acetowhitening (not completely opaque); indistinct, semi-transparent acetowhitening. Acetowhitening beyond the margin of the transformation zone Snow-white colour with with intense surface shine	Intermediate, shiny, grey-white shade	Dull, oyster-white
Lesion margin and surface configuration	Feathery, indistinct, or finely scalloped edges Angular, irregularly shaped, geographic margins Satellite lesions with margins well removed from the new squamocolumnar junction Lesion with a condylomatous or micropapillary contour	Regularly-shaped lesion with sharp, straight edges	Rolled, peeled edges Internal margins separating lesions with differing scores, the more central one with the higher score tending to be nearest the new squamocolumnar junction
Blood vessels	Fine punctuation or mosaic pattern	Absent vessels (after application of acetic acid	Coarse punctuation or mosaic pattern
Iodine staining	Positive iodine staining (mahogany-brown colour) Negative iodine staining in an area that scores 3 or less on the first 3 criteria	Partial iodine uptake giving a veriegated pattern	Negative for uptake giving a mustard yellow appearance in area that is significant (4 or more points) by the other 3 criteria

From Reid & Scalzi (1985); Reid (1993)
A score of 0–2 is compatible with CIN 1; 3–5 with CIN 1 or 2; and 6–8 with CIN 2 or 3.

studies in which colposcopy is used for primary screening (with or without cytology), studies assessing colposcopy as a diagnostic procedure are conducted on women referred with abnormal screening cytology and having, therefore, a higher probability and possibly a more severe spectrum of cervical pathology. Since women with more pronounced findings and disease may be selected by screening, the performance of colposcopy in a diagnostic capacity may exceed its accuracy and reproducibility when it is used as a screening tool. If possible, all women evaluated with a test under assessment should have the reference standard applied to avoid verification bias and where this is not possible, statistical correction should be made. When colposcopic findings are compared with the pathological diagnosis, the colposcopist and the pathologist should be blind to corresponding information from the other test.

In relation to a large multidisciplinary study of precancerous lesions in China, Belinson *et al.* (2001) observed that increased use of technology alone does not guarantee that detection improves. Important factors are whether the quality of light used optimizes perception, the adequacy of training of the personnel, and the attributes of the population studied, such as prevalence of cervical inflammation. The definition of abnormality and certainty thresholds used by colposcopists in a study is important, since these determine the replicability

Table 33. Indications for colposcopy

- Positive screening test result suggesting an increased probability of cervical neoplasia, e.g., cytology[a], visual inspection with acetic acid (VIA) and/or Lugol's iodine (VILI)
- Suspicious-looking cervix (where cancer cannot be excluded); regardless of the screening test result
- Presence of clinically apparent leukoplakia since a hyperkeratotic area may obscure a lesion and preclude adequate cytological sampling of the underlying area
- Presence of external genital warts; regardless of the screening test results (in some systems) (Howard *et al.*, 2002; Li *et al.*, 2003)
- Women at increased risk of cervical neoplasia[b]

[a] Abnormal cytology including ASCUS (with positive oncogenic HPV test), LSIL, HSIL
[b] Those who are HIV-positive; those with external genital warts

of findings and the test cut-off for what are the minimal criteria for abnormality.

Studies of diagnostic colposcopy

Two meta-analyses have been performed on the accuracy of diagnostic colposcopy applied to women referred with abnormal cytology. Mitchell *et al.* (1998a) performed a systematic review of 86 articles published between 1960 and 1996, nine of which met the inclusion criteria and eight were eligible for meta-analysis. At the cut-off level of normal versus abnormal on colposcopy, the average weighted sensitivity, specificity and area under the receiver operating characteristic (ROC) curve of histological CIN 2 or more were 96%, 48% and 80%, respectively. At the cut-off level of normal and LSIL versus HSIL and cancer on colposcopy, the corresponding results were 85%, 69%, and 82%. This suggests that, independent of prevalence and compared with low-grade lesions, high-grade lesions and cancer are diagnosed with higher sensitivity. Olaniyan (2002) reviewed publications from 1966 to 2000 and the results of his meta-analysis, based on eight studies, seven of which were included in the previous meta-analysis, were similar.

A recent study of the diagnostic accuracy of colposcopy in China (Belinson *et al.,* 2001) included methodological features intended to reduce selection bias and to assess the degree to which colposcopically directed biopsy is confounded with the colposcopic impression and the reference standard. In this study, vaginal and cervical specimens from 8497 women (aged 27 to 56 years) were screened for 13 oncogenic types of HPV (Hybrid Capture 2 assay) and by liquid-based cytology (AutoCyte, TriPath, Burlington, NC) (Pretorius *et al.,* 2004). Colposcopy was performed on 3063 women who had an abnormality on screening and a directed biopsy was obtained from any abnormality. If colposcopy showed no lesion in a quadrant of the transformation zone, a biopsy was obtained in the original squamocolumnar junction in that quadrant. An ECC was then performed after biopsies had been obtained. Based on all of the women who had colposcopy (including 11 with unsatisfactory colposcopy), the sensitivity and specificity of colposcopy for detection of CIN 2 or worse lesions were 62.4% (234/375; 95% CI 57.3–67.3%) and 93.7% (7612/8122; 95% CI 93.2–94.2%), respectively (Pretorius *et al.,* 2004).

Among the women with satisfactory colposcopy in the same study, directed biopsy detected 57.1% of high-grade lesions and cancers, while four-quadrant biopsy and ECC detected 37.4% and 5.5%, respectively. Among women referred for a cytological abnormality, directed biopsies were 4.8 times more likely to show a high-grade lesion or cancer than four-quadrant biopsies (26.5% versus 5.5%). The yields of CIN 2 or higher from four-quadrant biopsies for women referred because of HSIL, LSIL or ASCUS with a positive HPV test were 17.6%, 3.6% and 1.7%, respectively. One of 20 women in whom CIN 2 or worse was detected only by ECC had cancer despite satisfactory colposcopy.

A cohort study of 255 colposcopically negative women with abnormal cytology and 726 controls with normal cytology were followed for five years to assess the probability of false-negative colposcopy (Milne *et al.,* 1999). Subsequent neoplasia was found in 19% versus 3% of controls ($p < 0.0001$).

Studies of screening colposcopy

In a cross-sectional study, 1997 unscreened Chinese women (aged 35–45 years) first were assessed by VIA performed by a gynaecologist and then a second gynaecologist (blinded to the VIA results) performed colposcopy with directed biopsies being taken from abnormal areas (Belinson *et al.,* 2001). All women also had a biopsy taken from each of the four quadrants (and all had an ECC) in order to estimate the performance of colposcopy in a screening setting. Sensitivity and specificity of colposcopy and directed biopsy for high-

grade CIN or cancer were 81% (95% CI 72–89%) and 77% (95% CI 75–78%) compared with the combined histological findings from the directed, four-quadrant and ECC specimens.

A similar study in Germany enrolled 4761 women 18–70 years of age who were screened by conventional cytology (obtained under colposcopic vision), HPV testing of cervicovaginal samples by PCR and probing for 13 high-risk types and colposcopy when they visited one of ten gynaecologists for standard care (Schneider *et al.,* 2000). Biopsy and EEC were performed where appropriate and if colposcopy was normal, biopsies at 6 and 12 o'clock and ECC were obtained. The sensitivity and specificity of screening colposcopy for detecting at least CIN 2, by histological confirmation, were 13.3% (95% CI 7.0–20.5) and 99.3% (95% CI 99.0–99.6), respectively.

Five studies of the simultaneous use of colposcopy and cytology to detect cervical cancer, performed more than 30 years ago, showed that the combined sensitivity of the two methods for cervical cancer varied from 95.0% to 99.4% (Dexeus *et al.,* 1977). A recent case series from a German university using colposcopy and cytology for primary screening showed that the sensitivity of colposcopy for detecting at least CIN 2 was 90.8% (148/163) based on directed biopsy (Hilgarth & Menton, 1996). A similar study in the USA, based on 196 women who were screened opportunistically in a gynaecologist's practice, gave estimated sensitivities of screening cytology, colposcopy and their combination of 48%, 76% and 91% (Davison & Marty, 1994). The estimated specificities were 100%, 96% and 96%, respectively.

Validity of visual signs

Reid and Scalzi (1985) published a scoring system which quantified the degrees of difference within certain morphological parameters (Table 32). These included reference to the colour of the cervical epithelial, blood vessel structure and the surface configuration of the epithelium of the transformation zone, as well as the degree of iodine staining. However, few major studies have studied the incorporation of this scoring system within a colposcopic management regime. One retrospective study of 134 women with biopsy-proven lesions using the modified Reid index score showed that it gave more accurate prediction of low-grade versus high-grade disease than when the 1976 International Nomenclature for Colposcopic Classification was employed (Carriero *et al.,* 1991).

Prospective research on the predictive validity of visual signs in 425 women with abnormal cytology referred to a Canadian colposcopy clinic has shown that among three morphological characteristics routinely evaluated within the abnormal transformation zone (borders, degree of acetowhitening, abnormal blood vessels), performance based on acetowhitening was as good as all three signs combined (Shaw *et al.,* 2003). A prospective study of 2112 women referred to the Cook County Hospital Dysplasia Clinic in Chicago did not use standardized grading criteria, but did show an association between histology and colposcopic impression ($p < 0.001$), although agreement was poor (kappa, 0.20) (Massad & Collins, 2003).

The size of a lesion (categorized as the number of quadrants with positive histology) affects the sensitivity of colposcopy for detecting at least CIN 2 when the lesion grade on referral cytology or histology is controlled (Pretorius *et al.,* 2001). Colposcopy had a sensitivity of 65% (95% CI 47–79%) if the lesion involved only one quadrant of the cervical surface and 100% if more surface was involved (Belinson *et al.,* 2001). Shafi *et al.* (1991) excised the entire transformation zone by loop electrosurgical excision procedure (LEEP) and confirmed an association between lesion area and histological grade. A study that estimated lesion size from cervigrams concluded that lesion size affects the sensitivity of cytology (Barton *et al.,* 1989). Colposcopically inapparent high-grade lesions, remaining after a directed biopsy was taken, were evenly distributed among the four quadrants at 2, 4, 8, and 10 o'clock (Pretorius *et al.,* 2004).

While most studies of colposcopically directed biopsy have shown less than perfect sensitivity for detecting the presence of a higher-grade lesion found on a subsequent LEEP specimen (Howe & Vincenti, 1991; Barker *et al.,* 2001), the rate of underestimation among HIV-positive women may be substantially higher (Del Priore *et al.,* 1996).

Reproducibility of colposcopy

Observer agreement studies of visual methods have been conducted using cervical photographs taken after the application of dilute acetic acid. Between three expert colposcopists, intra-observer and inter-observer agreement was poor to good when assessing border characteristics (range of inter-observer kappa, 0.13–0.41; of intra-observer kappa, 0.26–0.58) and the colour of acetowhitening (range of inter-observer kappa, 0.21–0.47; of intra-observer kappa, 0.34–0.75). There was excellent agreement as to the site of the lesion from which a biopsy should have been obtained (raw agreement, 95.3%, 143/150) (Sellors *et al.,* 1990). Ferris *et al.* (2000b) studied the inter-observer agreement within pairs of colposcopists using optical and video colposcopes and found that colposcopic impression agreement with histopathology (kappa, 0.60; 95% CI 0.53–0.68), biopsy intent agreement (79.9%) and biopsy site selection

agreement by quadrant (A, 78.3%; B, 81.3%; C, 85.3%; D, 82.7%) were not significantly different ($p \geq 0.3$), despite the use of different colposcopes. Similar agreement was obtained when telecolposcopy (using a video colposcope) was viewed by an expert colposcopist at a remote location and compared with the video colposcopy performed by an expert on-site. The kappa values for colposcopic impression and histopathology agreement varied between 0.16 and 0.31 (p values not given) and for biopsy intent, kappa was 0.32 ($p = 0.002$) (Ferris *et al*. 2002).

Assuming that colposcopists use the same definitions, reproducibility of colposcopic assessment depends in part on colposcopists using similar 'thresholds of certainty' for categorizing findings as to normal versus abnormality and grade.

Quality control

Like other medical services, colposcopy services can be audited and compared with national standards, such as those established for the English National Health Service, for process and outcome (Ferris *et al.,* 2002). Indicators recommended for periodic audit include waiting time for colposcopy by grade of referral smear; adequacy of communication between primary and secondary level; frequency of procedures; agreement between colposcopic diagnosis, referral cytology and histology; treatment method by histological diagnosis; efficacy of treatment (e.g., whether histological evidence of CIN is present in over 90% of women undergoing 'see and treat'); and follow-up rates at one year (Luesley, 1996).

Cervical imaging using colpophotography, video colpography, and telecolposcopy has been studied. All methods give a true representation of what is seen at colposcopy and have been recommended for teaching and audit, as well as for patient care (Sellors *et al.,* 1990; Etherington *et al.,* 1997; Milne *et al.,* 1999; Harper *et al.,* 2000; Li *et al.,* 2003). Harper *et al.* (2000) showed that a telescopy network that allows transmission and sharing of static colposcopic images for consultation and teaching purposes on a regular basis was technically feasible, acceptable to women and health care providers living in remote areas, and gave good inter-observer agreement between the on-site colposcopists and the off-site review colposcopists as to degree of abnormality (kappa = 0.68; 95% CI 0.54–0.82). Ferris *et al.* (2003) showed that network telecolposcopy using high-speed telecommunications lines and computer telecolposcopy using modems and telephone lines to transmit static images was superior to cervicography as measured by the number of confirmed CIN lesions detected and timeliness of results. On-site colposcopy had the highest sensitivity to detect CIN because of the stereoscopic vision, the ability to manipulate the cervix and view the acetowhite reaction as it occurs, and the ability to resolve vascular and epithelial features. Compared with telecolposcopy, the ability to assess whether a colposcopic examination is satisfactory appears to be better with on-site colposcopy (Sellors *et al.,* 1990).

Documentation of colposcopic images and data using the latest digital photographic and information systems allows not only recording and comparison of colposcopic findings with subsequent examinations, but also the retention of data for audit, post-treatment follow-up and comparison of data between units.

Potential side-effects of colposcopy

A routine colposcopic examination involves some discomfort due to the insertion of the vaginal speculum and more when a tissue specimen is obtained by punch biopsy. Psychological morbidity should be appreciated and counselling considered (Howard *et al.,* 2002a). Studies using measures of anxiety such as the State-Trait Anxiety Inventory have consistently shown that anxiety scores before colposcopy are markedly elevated to levels seen in patients awaiting surgery, and fall immediately after colposcopy is completed. The fears that women have before colposcopy relate to cancer, fertility, danger to partner, social stigma and pain or embarrassment during the procedure. Other women may have a significant level of anxiety about the examination because of a possible history of sexual abuse. Educational booklets and counselling are effective in reducing anxiety (Ferris *et al.,* 2003). Colposcopy service providers need to be sensitive and responsive to women's needs in order to provide an acceptable service and to optimize adherence to appointments.

Cervicography

Cervigrams are replicate photographs of the cervix taken after application of 5% acetic acid, using a camera with a fixed focal length and internal light source. The images are projected onto a screen at a fixed distance to simulate magnification and are interpreted by a trained evaluator.

It is now possible to achieve equal visual resolution with digital cameras, producing images that can be immediately downloaded and transmitted for expert review, the images being evaluated with computer-generated magnification as needed. Future efforts related to cervicography will depend on digital techniques capable of generating images as good as those using high-quality film, with the advantages of 'telemedicine'-based screening and centralized image analysis (Wright, 2003).

Cervigrams are interpreted using the categories presented in Table 34. With these criteria, initial studies showed poor reproducibility because of differing distinction of the very subtle acetowhite lesions that represent either immature squamous metaplasia or HPV changes.

In a large inter-observer study among 3637 women, a comparison of dichotomous results (positive versus not) assigned by the initial versus the second evaluator yielded a kappa statistic of 0.5, indicating only moderate agreement beyond that expected by chance (Schneider *et al.,* 2002).

The deficiencies and inconclusive results of several small-scale studies of cervicography were summarized by Nuovo *et al.* (1997). In later large-scale evaluations, summarized in Table 35, cervicography proved insufficiently accurate to serve as a stand-alone screening test. In summary, cervicography appeared somewhat less accurate than cytology, primarily because of inferior specificity. This was a result of the overcalling of acetowhite epithelial changes.

The percentage of technically inadequate cervigrams varies widely by study; satisfactory results depend on the experience of the evaluator (De Sutter *et al.,* 1998). Adjudicated cervigram reviews and histological re-confirmation of CIN 2, CIN 3 or cancer did improve performance over a single interpretation, but suggested the upper limit of sensitivity (Schneider *et al.,* 2002). The sensitivity and specificity depended on the cut-point of positivity and targeted disease end-point, but no choice of cut-point generated excellent overall accuracy of detection of CIN 3 and cancer.

A major limitation of cervicography (and possibly, by extension, other static visual techniques) is the poor sensitivity among older women, whose transformation zones are often beyond the field of vision (Schneider *et al.,* 1999). Women aged 40–60 years would be expected to represent a sizeable proportion of women being screened in the low-resource settings where a non-cytological technique such as cervicography might be particularly helpful (Wright, 2003). However, in this group, the targeted precancerous lesions can be small enough to be easily missed. In one evaluation, the sensitivity of cervicography for detection of CIN 2, CIN 3 and cancer was only 30.0% among 2196 post-menopausal women, compared with 54.7% among 6264 pre-menopausal women, using a positive cut-point, and findings were similar using an atypical cut-point (see Table 34 for definitions) (Schneider *et al.,* 1999).

The relative accuracy of direct visual inspection compared with distant, expert review of a static visual image is not clear. There appears to be a trade-off between colposcopic

Table 34. Cervigram classification[a]

Not referred for colposcopy:	
Negative:	No definite lesion is visible
Atypical 1 (A1):	A lesion inside the transformation zone is visible; based on the lesion's site and morphology, the lesion is presently considered to be of doubtful significance
Atypical 2 (A2):	A lesion outside the transformation zone is visible; based on the lesion's site and morphology, the lesion is presently considered to be of doubtful significance
Technically defective:	The cervigram slide is not adequate for evaluation
Referred for colposcopy:	
Positive (all categories below): A lesion is visible and colposcopy is recommended because of the lesion's site and morphology, or no definite lesion is visible, but the appearance warrants colposcopy to exclude significant disease.	
Positive 0 (P0):	Probably normal variant; appearance warrants colposcopy to exclude significant disease
Positive 1A (P1A):	A lesion extending into the canal, the visible portion of which is presently considered to be of doubtful significance
Positive 1B (P1B):	Compatible with a low-grade lesion
Positive 2 (P2):	Compatible with a high grade lesion
Positive 3 (P3):	Compatible with cancer

[a] As of 1 January 1995, National Testing Laboratories worldwide revised the atypical category. Previously, atypical 1 referred to trivial lesions outside the transformation zone and atypical 2 referred to trivial lesions inside the transformation zone.

Modified from Schneider *et al.* (1999)

Table 35. Selected screening studies of cervicography

Study	Population	Disease target threshold	Cervicography	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Coibion *et al.* (1994)	Belgium, *n* = 4015	> CIN 1, n = 123	Old Atyp.[a]	86	99	76	99.0
Schneider *et al.* (1996)	Germany, *n* = 967	> CIN 2, n = 38	Atyp. or Pos.[a]	45	91	17	97.6
Baldauf *et al.* (1997)	France, *n* = 1351, mixed screening/ referral population	> CIN 1, n = 168	Positive[a]	51	96	44	97.1
De Sutter *et al.* (1998)	Belgium, *n* = 5192	> CIN 2, *n* = 33	Positive[a]	55	97	11	99.7
Schneider *et al.* (1999)	Costa Rica, *n* = 8460[b]	> CIN 2, *n* =136	Atyp. or Pos.	63	85	6	99.3
			Positive	49	95	14	99.1
Denny *et al.* (2000)	South Africa, *n* = 2611	> CIN 2, *n* = 79	Positive	58	91	58	93.4
Costa *et al.* (2000)	Italy, n = 992	> CIN 2, *n* = 90c	Atyp. or Pos.	76	91	51	97.4
Cronjé *et al.* (2001)	South Africa, *n* = 1747[d]	> CIN 1, *n* = 342	Positive	42	79	32	84.8
Cronjé *et al.* (2003)	South Africa, *n* = 1093	> CIN 2, *n* = 90	Positive (not P0)	49	88	26	95.0
Ferreccio *et al.* (2003)	Costa Rica, *n* = 8457	> CIN 3, *n* = 110 including follow-up	Atyp. or Pos.	62	85	5	99.4

[a] Evaluation not performed according to National Testing Laboratory criteria.
[b] Population-based screening of a high-risk province, where attempts were made to vary cervicography cut-point and disease end-point to explore performance.
[c] Women with negative colposcopy presumed to be disease-negative.
[d] Analysis of subgroup of large group of screened women. Subgroup included those with biopsied acetowhite lesions, as well as 1/5 of women with seemingly normal cervix. Predictive values not adjusted for sampling.

Sensitivity and specificity are estimated cross-sectionally (see Chapter 4)

expertise and the loss in visual discrimination inherent in examining an image compared to real-time examination. In one cross-sectional screening study with limited statistical power due to small numbers of precancerous outcomes, cervicography was apparently more accurate than direct visual inspection by nurses, due to increased specificity (Rodriguez *et al.*, 2004). However, in a statistically more powerful study, distant review by experts of digitized, static colposcopy images was significantly less sensitive (but more specific) than colposcopy performed by local gynaecologists and nurses with varied training (Ferris & Litaker, 2004).

Since it has been concluded, on the basis of accumulated data, that cervicography is inadequate as a stand-alone screening technique, research has shifted to evaluation of combining cervicography with

cytology or HPV for screening, and to a possible role for cervicography in the triage of women with equivocal cytology. These topics are considered below in the section on combined techniques.

HPV DNA testing

Research on the use of HPV DNA assays as a potential cervical cancer screening tool began in the late 1980s, as a reflection of the emerging evidence that these viruses played a causal role in the genesis of cervical neoplasia (zur Hausen, 1976; Deligeorgi-Politi *et al.,* 1986). Although much of that research began with a focus on viral detection as an end in itself (reviewed by Schiffman, 1992), attention soon turned to the potential clinical utility of HPV testing for identifying cervical cancer precursors (Lörincz *et al.,* 1990; Wilbur & Stoler, 1991; Lörincz, 1992). The basic assumption was that standardized molecular testing of exfoliated cervical cells for the putative causal agent of cervical cancer could have acceptable diagnostic performance, while being more reproducible and more easily adapted for automated, high-volume testing in clinical practice than conventional cytological testing. Concerns in the USA about the quality of smears processed in cytopathology laboratories added pressure to study the potential use of HPV testing as an adjunct to cytology (Reid *et al.,* 1991; Reid & Lörincz, 1991), despite some opposing views (Nuovo & Nuovo, 1991; Beral & Day, 1992). More recently, cytology has been characterized not only as a sufficient screening test, but also as a likely necessary component of future screening programmes based on HPV or visual testing, due to the low relative specificity of non-cytological methods (Suba & Raab, 2004).

Techniques to detect the presence of HPV in cervical cell specimens have evolved considerably in the last 25 years, from (i) simple scoring of koilocytes (a type of cytopathic effect taken to indicate the presence of HPV in the host epithelial cells) in cervical smears (Komorowski & Clowry, 1976) to (ii) immunocytochemical staining (Syrjänen & Pyrhonen, 1982); non-amplified nucleic acid hybridization methods, such as (iii) dot blot (Parkkinen *et al.,* 1986), (iv) Southern blot (Okagaki *et al.,* 1983) and (v) filter in-situ hybridization (Schneider *et al.,* 1985); signal-amplified, immunoassay-based nucleic acid hybridization techniques such as (vi) the Hybrid Capture™ (HC) assay (Farthing *et al.,* 1994); and (vii) a variety of type-specific (Dallas *et al.,* 1989) and general or consensus-primer (Gregoire *et al.,* 1989; Manos *et al.,* 1989; Snijders *et al.,* 1990; Roda Husman *et al.,* 1995; Kleter *et al.,* 1998; Gravitt *et al.,* 2000) polymerase chain reaction (PCR) techniques. In addition, adaptations of the solution-based, non-amplified hybridization and PCR protocols have been used to detect HPV DNA in histological sections or smears, to allow confirmation of the presence of the virus in particular target cells. Such *in situ* techniques (Gupta *et al.,* 1985; Nuovo *et al.,* 1991) have been useful in molecular pathology studies, but have found little interest as potential screening tools for cervical cancer and its precursors. Serological assays to detect antibodies to HPV capsid or functional protein antigens have also received attention as investigational tools in epidemiological and clinical studies (Jochmus-Kudielka *et al.,* 1989; Galloway, 1992). However, as with *in situ* assays, they have not been considered as candidate methods for screening cervical cancer precursors. Serology detects humoral immune response to HPV antigens, which may reflect cumulative exposure to HPV infection acquired in mucosal sites other than genital, and thus is not suitable, in principle, as a screening tool.

Early clinical studies used non-amplified DNA hybridization methods (without signal amplification) to gauge the screening utility of HPV testing to identify and manage cervical lesions. Such methods are no longer used, however, because of their insufficient sensitivity and specificity for epidemiological and clinical studies (Franco, 1992; Schiffman & Schatzkin, 1994). Only the commercially available HC assay and a few PCR protocols have been the focus of investigations conducted in the last 10 years. Although a number of biotechnology companies are currently developing HPV DNA diagnostic systems for clinical use, few are yet available commercially or have reached the stage of large-scale clinical studies. For this reason, this overview focuses primarily on the HC assay and on the more popular PCR protocols that have been used in screening studies.

Hybrid Capture™ assay

Most clinical investigations of HPV testing have used first- or second-generation Hybrid Capture™ (HC) systems (Digene, Inc., Gaithersburg, MD), the only HPV test currently approved by the US FDA. The HC system is a nucleic acid hybridization assay with signal amplification for the qualitative detection of DNA of high-risk, cancer-associated HPV types in cervical specimens. It cannot determine the specific HPV type present, since detection is performed with a combined probe mix. The first HC assay (HC1) was a tube-based detection system and probed for only nine of the high-risk HPV types: 16, 18, 31, 33, 35, 45, 51, 52 and 56. The second-generation HC system (HC2) has improved reagents and is based on a microplate assay lay-out that targets 13 high-risk HPV types: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56,

58, 59 and 68. A probe set for a few non-oncogenic HPV types (6, 11, 42, 43, 44) has been available for both the HC1 and HC2 assays but its utility has not been sufficiently investigated in clinical or epidemiological studies. It is often designated as probe A, whereas probes for high-risk HPV types are referred to as probe B.

HC2 is an entire system that can be used with a dedicated cervical sampler kit containing a special cervical conical brush and a vial with specimen transport medium (STM). The brush is designed for optimal collection of cells from both the ectocervix and endocervix. The brush is shaped as a cone (Christmas-tree-like) that fits the cervical canal and samples the endo- and ecto-cervix. This brush is inserted gently into the cervical canal and fully rotated three times. It is then retrieved without touching the vaginal wall and inserted into the collection tube containing STM. The tip is broken and the tube is closed. According to the manufacturer, specimens in STM can be held at room temperature for up to two weeks and can be stored for an additional week at 4°C. If not tested in the first three weeks after collection, they can be stored at –20°C for up to three months.

HC2 is a solution hybridization assay that uses long synthetic RNA probes that are complementary to the DNA sequence of the 13 high-risk HPV types (or to the probe A types) listed above. The initial reaction step denatures the exfoliated cells in STM, thus releasing host and any existing HPV DNA molecules to the solution. HPV DNA molecules then bind (i.e., hybridize) with the respective RNA probe, resulting in the formation of DNA–RNA hybrids reflecting the composition of HPV types present in the mixture. This hybridization step occurs in solution inside the wells of a specially treated 96-well plastic microtitration plate previously coated with polyclonal IgG antibodies that are specific for RNA–DNA hybrids, regardless of sequence homology. Any such hybrids will then be captured by the solid-phase-bound antibodies, hence the name 'hybrid capture'. After washing steps to remove unbound molecules, a solution of a conjugate reagent consisting of the same anti-RNA–DNA hybrid antibody covalently linked with the enzyme alkaline phosphatase is added to the wells. Conjugate antibody molecules will then bind to any solid-phase-bound hybrids. After further washing to remove unbound molecules, a solution containing a chemiluminescent dioxetane substrate is added to the wells. Cleavage of the substrate by the alkaline phosphatase releases a luminescent reaction product into the solution. The intensity of the light emitted is proportional to the amount of HPV DNA originally present in the specimen and is measured in a luminometer provided with the system. The reaction signal of each specimen is expressed on a scale (relative light units or RLU) relative to the average reactivity measured in triplicate wells with a positive control containing 1.0 pg of HPV16 DNA per ml. Specimens yielding RLUs greater than or equal to 1.0 are considered positive; some studies have assessed the validity of this cut-point using ROC curve analysis (Schiffman *et al.,* 2000). In most clinical settings, the manufacturer (Digene) certifies the laboratory that intends to perform HC2 testing, thus ensuring quality control.

Because the RLU signal is proportional to the amount of HPV DNA present in the specimen, the HC2 assay has occasionally been used to infer viral load, on a semi-quantitative basis (Clavel *et al.,* 1998; Sun *et al.,* 2001; Cuzick *et al.,* 2003). The assay is easy to perform in clinical practice and amenable to automation, which makes it attractive for high-volume screening use. To this end, a robotic assay workstation named Rapid Capture System™ (Digene) is available, which performs specimen transfer, all pipetting operations, incubations, shakings and washings. However, the denaturation of specimens in the sample device tubes still has to be performed by hand. This automatic station increases the accuracy of the test and allows a single user to test 352 specimens within four hours.

Since it is based on signal, rather than target amplification (as in the case of PCR protocols), HC2 is less prone to cross-specimen contamination, thus obviating the need for special laboratory facilities to avoid cross-contamination (Coutlee *et al.,* 1997). In practice, only the high-risk probe mix (probe B) is used for cervical lesion screening, which reduces the time and cost to perform the test. At the standard FDA-approved cut-off of 1 pg/ml (RLU $\geq$ 1.0) and even at higher discriminant levels, there is cross-reactivity between certain HPV types not present as targets in the probe B set (e.g., 53, 66, 67, 73) and the RNA probes used in that set (Peyton *et al.,* 1998; Vernon *et al.,* 2000; Howard *et al.,* 2002b, 2004). Cross-reactivity with non-cancer-causing types would have an adverse impact on test specificity in settings with high prevalence of the low-risk types. On the other hand, cross-reactivity with other high-risk types not represented in the probe B set may be beneficial for test sensitivity (Castle *et al.,* 2003).

Polymerase chain reaction

PCR is based on the repetitive replication of a target sequence of DNA flanked at each end by a pair of specific oligonucleotide primers, which initiate the polymerase-catalysed chain reaction. Because of the exponential increase in the amount of target DNA sequence after a few reaction cycles of denaturation, annealing and extension, PCR has very high levels of molecular

sensitivity and permits the detection of less than 10 copies of HPV DNA in a mixture. Therefore, PCR has a lower threshold of molecular detection for HPV DNA than the HC assay. PCR is based on target amplification with type-specific or consensus or general primers. The latter are able to amplify sequences from several different HPV types because they target conserved DNA regions in the HPV genome. The amplified DNA products can be revealed by ethidium bromide staining following agarose or acrylamide gel electrophoresis, which permits presumptive verification of the expected molecular weight of the amplified target, thus confirming positivity. Verification can also be done by methods that further probe the post-amplification products for their sequence homology with the target. Dot blot, Southern blot or line strip hybridization are used to this end and generally result in improved molecular sensitivity and specificity as compared with electrophoresis and staining (Gravitt & Manos, 1992; Gravitt *et al.,* 1998). Finally, use of restriction enzymes to analyse the fragment length signatures in combination with probe hybridization (Bernard *et al.,* 1994) and direct DNA sequencing provide the highest possible resolution to distinguish the HPV types present in a biological specimen.

The very high sensitivity of PCR is its very limiting factor in terms of clinical applicability. Molecular threshold does not correlate directly with clinical sensitivity and specificity (Snijders *et al.,* 2003). Because millions of copies of the DNA target can be produced from a single molecule, there is a high probability of contamination of other specimens and control samples with HPV sequences in airborne droplets and aerosolized reaction mixtures. In fact, cross-contamination was a major problem in some early applications of PCR in HPV testing. Extreme care is needed in PCR testing laboratories. Several procedures are well established to minimize the potential for contamination, the most important of which is the separation of pre-amplification and post-amplification areas.

Judicious analysis of sequence homology among different genes of distinct HPV types using software that aligns DNA sequences will reveal countless segments that could serve as candidates for PCR primer design. In fact, many type-specific and consensus HPV testing PCR protocols have been published in the last 15 years. However, because of the requirements for validation, reproducibility, and general acceptability, relatively few have become established to the point of being widely used in clinical and epidemiological studies. Primer systems targeting sequences in the L1, E1, E6 and E7 genes have been most commonly used. Because of their well conserved sequences, L1 and E1 have been targeted by the consensus primer protocols. E6 and E7, on the other hand, have more sequence variation among HPV genotypes, making them less suitable as targets for amplification of a broad spectrum of HPV types (Gravitt & Manos, 1992).

The most widely used PCR protocols are of the consensus or general primer (degenerate or non-degenerate) type, i.e., they can potentially amplify sequences of multiple HPV types with one primer set in one reaction pass. The size of the amplified product is the same irrespective of the HPV type present in the starting mixture, and thus electrophoresis cannot reveal the actual type present in the sample. The post-amplification hybridization or sequencing techniques described above must be used to identify the HPV type or types originally present. Three consensus primer systems (and their technical variations) based on L1 sequence detection have become well established. They can detect essentially all types of HPV that infect the mucosal areas of the lower genital and upper aerodigestive tracts. Two of these, the MY09/11 (Manos *et al.,* 1989) and the GP5/6 (Snijders *et al.,* 1990; Van den Brule *et al.,* 1990) systems have evolved into variants with better primer composition and internal oligonucleotide probing, such as the PGMY09/11 (Gravitt *et al.,* 1998, 2000) and the GP5+/6+ (Roda Husman *et al.,* 1995; Jacobs *et al.,* 1995, 1997; Van den Brule *et al.,* 2002) protocols. Over the years, the original radioactively labelled hybridization probes have gradually been abandoned in favour of biotinylated probes and enzyme immunoassay formats. The third protocol is designated SPF10 LiPA, for line probe assay based on the SPF10 primer set (Kleter *et al.,* 1998; Quint *et al.,* 2001). Although these three consensus protocols amplify targets within the L1 gene of HPV, they do so for segments of considerably different sizes: 450 base pairs (bp) for MY09/11, 140 bp for GP5/6, and 65 bp for SPF10 LiPA. The size of the amplified product is not a trivial matter. Although discrimination of sequence homology is better for longer gene segments and thus would in theory permit improved HPV type resolution, shorter fragments tend to yield better sensitivity with severely degraded specimens, such as paraffin-embedded, archival tumour tissue. Damage is often pronounced in DNA extracted from such archival specimens, resulting in DNA fragments of less than 200 bp. In these circumstances, a protocol targeting a short fragment, such as GP5+/6+ or SPF10 LiPA, tends to yield fewer false negative results (Gravitt & Manos, 1992).

The newly developed Roche prototype Microwell plate assay (Roche MWP) employs an oligonucleotide set which amplifies a short fragment of the L1 gene of high-risk HPV types (170 bp, compared with 450 bp with

PGMY09/11). This amplicon is immobilized using a pool of capture molecules bound to the wells of a microtitre well plate and visualized by colorimetric detection. The new test was developed to employ the TaqGold DNA polymerase, which minimizes the amount of non-specific amplification and increases the sensitivity of the test. Since it amplifies a shorter fragment, it is considered to be more sensitive than PGMY09/11 PCR and also suitable for less well preserved specimens; it has been reported that these primers detect about 13% more HPV in cervical smears than the PGMY primers (Iftner & Villa, 2003). However, because these primers were designed for high-risk types only (HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 68), this test is not truly generic, but rather comparable to the HC2 test. In addition, the use of PCR assays aiming at maximum sensitivity for detection of HPV in a screening setting irrespective of concomitant disease may be inappropriate with regard to clinical usefulness.

The reproducibility and agreement of HPV testing results among the three most popular PCR protocols, as well as between them and the FDA-approved HC2 assay, for overall HPV detection have been extensively studied. While agreement at the overall positivity level may be considered adequate in clinical settings, concordance at the level of type detection leaves much to be desired (Qu *et al.,* 1997; Kleter *et al.,* 1998, 1999; Peyton *et al.,* 1998; Swan *et al.,* 1999; Gravitt *et al.,* 2000; Castle *et al.,* 2002c; Van Doorn *et al.,* 2002; Castle *et al.,* 2003).

Biomedlab Co. (Republic of Korea) has developed an HPV oligonucleotide microarray-based system for detection of HPV types that currently allows detection of 22 HPV types, by immobilizing HPV type-specific oligonucleotide probes and a control (beta-globin probe) on an aldehyde-derivatized glass slide. Target DNA is subjected to a standard PCR in the presence of fluoresceinated nucleotides (labelled with Cy5 or Cy3) employing primers for both the beta-globin (PC03/04) and for the L1 region (modified GP5/6 primers) of several HPV types. Randomly labelled PCR products are then hybridized to specific oligonucleotides on the chip, which is afterwards scanned by laser fluorescence. In the case of multiple infections, multiple hybridization signals can be seen. Because signal detection in microarrays is subject to variation, additional levels of control would be desirable. These should include quality control of the efficiency of the PCR reaction and the hybridization conditions, include a measurement of the homogeneity of the probes on the chips and allow some sort of quantification. In addition, the read-out requires expensive equipment for signal detection and would need to be performed with the help of special software that allows threshold settings.

Sensitivity and specificity of HPV assays

Dozens of studies have provided data on the diagnostic performance of HPV DNA testing methods. However, only some of them provided direct comparisons with cytological testing in detecting high-grade precancerous lesions and clearly specified the type of population, i.e., whether it was a primary screening or secondary triage study. The vast majority of studies either did not provide data on cytology or presented data on mixed series of subjects and could not be unequivocally designated as screening or triage settings. Table 36 summarizes the main features of selected published studies that provided data on the comparison of HPV testing with cytology in primary screening for cervical cancer and its precursors; it also gives estimates of sensitivity and specificity for HPV testing and for cytology in the same studies. Data on the performance of HPV testing in triage studies are presented later in this chapter. For all studies, specificity estimates are based on women free of histologically demonstrable squamous lesions.

The studies vary considerably in terms of investigational design, choice of population and methods, which, as expected, leads to enormous variability in the results observed. Most studies assessed HPV test performance on the basis of prevalent lesions using simple cross-sectional designs or retrospective case series, whereas some assessed both prevalent and seemingly short-term incident lesions based on cross-sectional investigations with extended follow-up (ASCUS-LSIL Triage Study (ALTS) Group, 2003a, b). Lesion definition varied across studies and included either CIN 1 or CIN 2/3 or worse lesions, diagnosed by histology on specimens obtained by colposcopy-guided biopsy. Sometimes the SIL terminology was used for these histological diagnoses. In some studies, the colposcopic result was used if no biopsy was taken. Some studies used direct community recruitment, but usually the study population was clinic-based. None of the studies was based on long-term follow-up for more relevant end-points, such as incidence of CIN 2 or 3 or cancer or mortality from invasive cervical cancer (see Chapter 4).

For many of the studies, the purpose was to compare HPV testing with other screening technologies for cervical cancer (primarily cytology). None of the investigations was a randomized controlled trial; all were based on concomitant testing for HPV and cytology alone or with additional tests. Such investigations are known as split-sample studies because the cervical specimen, collected in single or multiple exfoliative procedures using a swab, a cytobrush or other collection device, is split into several sub-

Table 36. Selected studies that provided data on the comparison of HPV testing with cytological testing[a] in primary screening for cervical cancer and its precursor lesions: characteristics and estimates of screening performance

Study, country	Study size	Age (years)	HPV test	Study features	Sensitivity (%)		Specificity (%)		Comments
					HPV	Pap	HPV	Pap	
Cuzick *et al.* (1995), UK	2009	20–45	Type-specific PCR (16, 18, 31, 33)	Women free of cytological abnormalities at enrolment	75	46	96	98	
Cuzick *et al.* (1999), UK	2988	34+	HC1, HC2, MYO9/11 PCR	Women free of cytological abnormalities at enrolment	95	79	95	99	HPV indices based on HC2 (*N* = 1703)
Kuhn *et al.* (2000), South Africa	2944	35–65	HC1, HC2	Unscreened population, community recruitment	88	78	82	97	LSIL in histology excluded, HPV indices based on HC2 (*N*=424)
					73		88		LSIL in histology excluded, HPV indices based on HC1 (*N*=2861)
Ratnam *et al.* (2000), Canada	2098	18–69	HC1, HC2	Multiple screening practices, 10% random sample of Pap-/HPV- women referred for colposcopy	68	27	91	96	All ages, bias-adjusted[b] specificity includes CIN 1
Schiffman *et al.* (2000), Costa Rica	8554	18+	HC1, HC2	Population-based, HPV positivity not a criterion for colposcopy referral	88	78	89	94	Conventional cytology and HC2, all ages
					93		80		HC2, ages 18–30
					81		90		HC2, ages 31–40
					93		94		HC2, age > 40
Wright *et al.* (2000), South Africa	1356	35–65	HC2	Subset of sample in Kuhn *et al.* (2000)	84	61	83	96	Clinician-collected cervical samples tested for HPV
Schneider *et al.* (2000), Germany	4761	18–70	GP5+/6+ PCR	Multiple screening practices, cross-sectional plus 8 months follow-up testing	89	20	94	99	Bias-adjusted[b]
Belinson *et al.* (2001), China	1997	35–45	HC2	Unscreened population, community recruitment, all women underwent colposcopy	95	87	85	94	ThinPrep cytology
Blumenthal *et al.* (2001), Zimbabwe	2073	25–55	HC2	Completed recruitment (irrespective of HIV sero-status) of study in Womack *et al.* (2000), all women underwent colposcopy	80	44	61	91	Bias-controlled[b]
Clavel *et al.* (2001), France	7932	15–76	HC2	Women free of cytological abnormalities at enrolment, cross-sectional plus 15 months follow-up testing, HPV positivity alone not a criterion for immediate colposcopy referral	100	68	87	95	All ages, paired set with conventional cytology (*N*=2281)
					100	88	86	93	All ages, paired set with ThinPrep cytology (*N*=5651)

Table 36 (contd)

Study, country	Study size	Age (years)	HPV test	Study features	Sensitivity (%)		Specificity (%)		Comments
					HPV	Pap	HPV	Pap	
					100	58	90	96	Age > 30, paired set with conventional cytology (*N*=1550)
					100	84	88	95	Age > 30, paired set with ThinPrep cytology (*N*=4121)
Kulasingam *et al.* (2002), UK	4075	18–50	HC2, MYO9/11 PCR	Family planning clinic recruitment, ThinPrep cytology, 41% random sample of Pap-/HPV- women referred for colposcopy	91/88	57	73/79	90	All ages, detection of > CIN 3, HPV tests: HC2/PCR, bias adjusted[b]
					74/70	46	71/78	89	Age < 30, detection of ≥CIN 2, HPV tests: HC2/PCR, bias-adjusted[b]
					63/57	36	83/87	96	Age ≥ 30, detection of ≥CIN 2, HPV tests: HC2/PCR, bias-adjusted**
Petry *et al.* (2003), Germany	8466	30+	HC2 and PCR	Primary screening network of clinics and gynaecological practices: patients attending routine cervical screening stratified to be representative of all of Germany. Bias controlled	98	37	95	99	HC2, CIN 2/3+, bias-adjusted[b]
					97	40	95	99	HC2, CIN 3+, bias-adjusted[b]
Salmeron *et al.* (2003), Mexico	7868	15–85	HC2	Women attending opportunistic screening. Self-collected vaginal compared with clinician-collected cervical samples. Those positive for HPV or by cytology were referred for colposcopy and biopsy	93	59	92	98	CIN 2/3+, clinician-collected samples
Sankaranarayanan *et al.* (2004), India	18 085	25–65	HC2	Women attending primary screening in 3 different sites in India. All subjects investigated by colposcopy and, when necessary, underwent a biopsy. Averted verification bias in the design.	46	37	92	87	Kolkata
					69	70	94	99	Mumbai
					81	72	95	98	Trivandrum

[a] LSIL threshold (majority of studies) or LSIL or persistent ASCUS (Kulasingam *et al.*, 2002). Unless otherwise indicated, results are based on conventional cytology. Unless otherwise stated, estimates shown are for CIN 2/3 lesions or worse as disease outcome.
[b] Verification bias: bias-controlled denotes verification of disease status among all participants and bias-adjusted denotes correction of estimates based on the verification of disease status in a random sample of test-negative women.

samples for testing. Studies varied in terms of timing of collection, collection method, or whether or not visual methods for cervical inspection were used as adjunct screening techniques.

One advantage of HPV DNA testing is that it is suitable for self-sampling, as in many of the studies shown in Table 36. Self-sampling is likely to improve compliance, and is particularly appealing in populations with social or religious limitations on the acceptability of vaginal examinations.

A few large randomized controlled trials of HPV testing in primary cervical cancer screening are currently in progress. Of note are the UK "HPV in Addition to Routine Testing" (HART) investigation (Cuzick *et al.,* 2003), the Dutch POBASCAM trial (Bulkmans *et al.,* 2004), the UK "A Randomized Trial in Screening to Improve Cytology" (ARTISTIC) (H. Kitchener, personal communication), the Osmanabad trial in India (R. Sankaranarayanan, personal communication), the Italian trial (G. Ronco, personal communication), the Canadian Cervical Cancer Screening Trial (CCCaST) (E. Franco, personal communication) and a trial in Finland (Nieminen *et al.,* 2003).

The majority of the estimates in Table 36 must be interpreted with caution because of selection biases and other issues that affect computation of screening performance indices. For instance, the sensitivity and specificity estimates of most studies shown in Table 36 are relative, not absolute, because they are not based on interval cancer incidence and are subject to verification bias (see Chapter 4). The latter occurs whenever the probability of disease verification via the gold standard is dependent on the screening test result. In general, such studies used a design in which only women with one or more positive screening tests were referred for colposcopy and biopsy, which prevented the unbiased estimation of absolute sensitivity and specificity (their estimates should be considered relative). These studies relied on the fact that with two or more tests, there were always combinations of either cytology-negative or HPV-negative women with verified disease status available for analysis. However, the biasing effects of the unequal verification of disease status can be strong and may lead to estimates of screening efficacy that cannot be generalized for cost considerations and other public health uses (Franco, 2000). Such verification bias was averted (by applying the gold standard of disease verification to all women) or corrected (by extrapolating the screening results from a random fraction of women with negative screening tests to those without colposcopic verification) in a few studies, as indicated in Table 36.

An important assumption in dealing with the issue of verification bias is the expectation that the gold standard of colposcopy-guided biopsy provides perfect ascertainment of disease. Studies that either avoided or corrected for the putative bias assumed that a colposcopy-guided biopsy accurately reveals the existence of cervical lesional tissue, which was then used to ascertain the distribution of diseased and non-diseased women, allowing the computation of adjusted estimates of screening validity. While the approach is correct for its intended purpose, i.e., to obtain an improved estimate of the distribution of disease conditional on test results, it should be recognized that a simple colposcopy or even a colposcopy-guided biopsy cannot guarantee that a lesion will be detected. In many test-negative women, the colposcopist cannot visualize lesional tissue and may decide that the colposcopic impression of no disease alone serves as definitive diagnosis. However, a lesion could be hidden in the endocervical canal and not visible. Although this pitfall could be minimized by adopting a colposcopy protocol in which blind biopsies are collected, it is still possible that a fraction of the existing lesions will remain undetected, because of either their location or size. Therefore, in any cross-sectional survey of screening efficacy, the ethically acceptable gold standard for cervical lesions (colposcopy-guided biopsies) is an imperfect one because of inadequate sampling of the entire cervical tissue that is at risk for squamous-cell malignancy. Only a more aggressive diagnostic approach such as a detailed histological examination of serial sections from cone biopsies or from specimens collected LEEP or by large loop excision of the transformation zone (LLETZ) would approach the definition of an acceptable gold standard of disease, but such an approach even in a sample of test-negative women would be unethical as well as impractical.

Even if tissue sampling could be done optimally with respect to lesion site and time of development, one needs to consider also the misclassification of lesion outcome status that exists even with histopathological ascertainment. Studies that involve multiple expert pathologists indicate that the reproducibility in grading histopathology specimens is not high, even with large specimens, such as LEEP-obtained tissue samples. Therefore, a study that is simply based on lesion ascertainment by a single expert pathologist will be more prone to lesion misclassification than one employing a panel of readers that reaches a consensus diagnosis in every case.

Furthermore, as the design of screening efficacy studies evolves from the traditional single-opportunity sampling, cross-sectional layout to long-term, repeated sampling investigations over many years, disease case definition becomes a more dynamic process, requiring the juxtapositioning of screening and diagnostic test results

obtained from multiple samples collected over time. This process involves combining the results from different diagnostic approaches, which may be differentially triggered by the severity of the lesion grade presumed by the test (HPV or cytology), e.g., colposcopy with simple biopsy for equivocal or low-grade lesions, LEEP for high-grade lesions, etc. Natural history investigations of HPV and cervical neoplasia are examples of studies that have to grapple with this added complexity by having to differentiate between prevalent and incident lesions, progression and regression, and relating them to screening test performance. Calculation of sensitivity and specificity in such studies involves the combination of diagnostic information over multiple samples, which greatly reduces the chance that any lesions are missed through the pitfalls described above for a cross-sectional study relying on colposcopy-guided biopsies alone. On the other hand, the repeated sampling layout of these investigations obviates the need for invasive diagnostic procedures among women testing consistently negative for both HPV and cytology over many visits. The longitudinal nature of the investigation ends up providing the test and diagnostic data that approaches the true distribution of disease dynamics, conditional on study duration. Therefore, correction for verification bias is not a critical issue in these longitudinal studies with intensive follow-up of test-negative women and repeated histological sampling of test-positive cases. However, such studies do have to contend with the issue of distinguishing between prevalent and incident lesions to properly assign the distribution of disease for the purposes of gauging screening test efficacy.

As described in Chapter 4, the ideal estimation of sensitivity assumes follow-up and clinical surveillance, via a cancer registry or otherwise, for invasive cancer that is diagnosed between the screening tests.

Another issue that affects screening performance of HPV tests is the type of specimen. In theory, clinician-collected cervical specimens are ideal in terms of sampling exfoliated cells from the target tissue. Therefore, clinical correlations between lesion severity and presence of HPV should be optimal with such specimens. However, in public-health practice, convenience for the patient and cost-saving considerations have led some to propose self-sampling as a viable alternative to collection by a clinician; the assumption is that the loss in screening accuracy would not be substantial to the point of offsetting the benefits of simplifying specimen collection (Sellors *et al.,* 2000). Self-sampling of genital specimens remains an attractive option in developing countries and in remote regions where health-care providers cannot be available at point-of-care settings. However, issues of validity, acceptability and training present obstacles to wider application of self-sampling.

Costs and potential hazards

The most important obstacles to more widespread acceptance of HPV testing in cervical cancer screening are its high unit cost and the fact that the technology is not in the public domain, as it is for cervical cytology. The cost-effectiveness of HPV testing is heavily dependent on assumptions related to the intrinsic cost of the test, the infrastructure available in the setting where the screening will be implemented, the length of interval between screening visits, and the existing expenditures incurred by quality assurance imposed by local legislation.

There are no additional physical hazards associated with the application of HPV testing technology for the purpose of cervical cancer screening, as the specimen used in the test is the same as that collected for a traditional cytological test. Only minor discomfort and very minimal risk are associated with obtaining exfoliated cervical cell samples. On the other hand, little is known about the psychological and emotional impact of communicating positive HPV test results to women. As knowledge about HPV has become more widespread, there has been a gradual shift in how the medical and public-health communities consider cervical cancer prevention; the perspective has moved from an oncological one to a model in which a sexually transmitted infection is the target (Franco, 2003). Implementation of testing for HPV in primary screening for cervical cancer would lead to a large proportion of women having to be told that they harbour a sexually transmitted viral infection that can ultimately cause cancer. There is a dearth of research on the merits and consequences of conveying this information. The vast majority of such women will not be required to change their lifestyle or to be referred for a more aggressive diagnostic procedure on the basis of this information, since their infection will be found to be transient. Therefore, it is debatable whether conveying this information would bring any real benefit to a screening participant. Practically nothing is known on the potential negative impact, including social and legal implications, of imparting this information. Also the dynamics of between-gender transmission of HPV infection are poorly understood. Such information is important in screening contexts, e.g. for health providers to convey meaningful information on risk to couples.

Another concern with the use of HPV testing in cancer screening is the potential for a breakdown in quality-control safeguards if too many commercial test suppliers enter the market without a certain level of regulatory control of performance standards by health-care or government agencies. At present there are only a

few commercial suppliers of HPV diagnostic systems. The two major ones can afford to keep up strict quality-control standards in reagent batch production and performance characteristics by passing on the costs to the consumers or their health insurers, private or public. However, increased competition resulting in diminishing market shares and reductions in the cost of testing might lead test manufacturers to relax their standards of quality. Such a scenario could prove disastrous in many respects, since there are theoretically many more variables that can affect the performance of HPV testing than there are for cytology-based screening. It is imperative, therefore, that early performance and proficiency standards be agreed upon by public health agencies involved with quality assurance of cervical cancer screening.

Other emerging techniques

This section describes three new developments in screening methods: (1) computer-assisted cytological interpretation of cervical smears, (2) use of physical real-time devices and (3) detection of molecular surrogate markers of cancer progression.

Computer-assisted reading of cervical smears

The aim of automation-assisted screening is to increase the sensitivity of cytological testing by finding, for instance, small abnormal squamous and glandular cells, known to be very difficult to detect in conventional screening; it should also increase specificity by selecting only lesions corresponding to objective reproducible criteria. Automated screening should also increase productivity by excluding normal slides or part of the slides from manual screening by selecting most atypical images from a slide to be checked by the cytologist, so as to allow more slides to be screened without increasing the number of staff.

The PAPNET system, that is no longer commercially available, included neural network software and traditional imaging technology. It selected 128 of the most suspect fields in conventional cervical cytological specimens and presented these images on a video review screen. The cytotechnician then interpreted the images on the screen and decided whether to carry out manual screening. Another system that was introduced in the 1990s was AutoPap. This computerized scanning device was originally designed for algorithmic classification of conventional cervical cytology specimens, but was later approved for liquid-based cytology specimens. The device was initially approved in the USA by the FDA as a method to be used for quality control. In the quality control mode, only those specimens classified as normal ('within normal limits') were reviewed through the device. Subsequently AutoPap was approved by the FDA for primary screening (Dunton, 2000). In the primary screening mode, all slides are processed through the device and then, on the basis of an 'abnormality index' assigned to the slide by the algorithmic processing feature, each slide is either filed without manual review by a cytotechnician (up to 25% of all specimens) or reviewed manually in the normal manner.

Both of these devices have served as prototypes for newer devices that are being developed to help automate the evaluation of cervical cytology specimens. Recently, the ThinPrep Imaging System (Cytyc, Boxborough, MA, USA) has received FDA approval for use in primary screening of liquid-based cytology specimens in the USA. This system uses image analysis and algorithmic processing to identify a fixed number of the worst microscopic fields on a given slide; a motorized computer-controlled microscope stage then takes the cytotechnician directly to these specific microscope fields. TriPath Imaging has published results using a similar type of device, referred to as the Focal Point location-guided screening device, that is not yet FDA-approved (Wilbur *et al.,* 2002). This device is based on the earlier AutoPap device.

Several studies have reported the test accuracy of automation-assisted screening (Kok & Boon, 1996; Wilbur *et al.,* 1996; Koss *et al.,* 1997; Michelow *et al.,* 1997; Bartels *et al.,* 1998; Doorneward *et al.,* 1999; Halford *et al.,* 1999; PRISMATIC Project Management Team, 1999; Bergeron *et al.,* 2000a; Duggan, 2000; Kok *et al.,* 2000). They show generally a better test sensitivity with at least the same specificity as conventional screening. Most studies were retrospective (quality control) and/or involved rather small numbers of smears. One larger prospective study, conducted by the PRISMATIC Project Management Team (1999), including 21 700 smears, also showed equal sensitivity but better specificity for automated screening, as well as higher productivity. Results of only two randomized prospective public health trials in a primary screening setting have been reported. One of these studies found clearly higher detection rates of *in situ* and invasive carcinoma (Kok & Boon, 1996). However, the second study, integrated in the Finnish organized screening programme and involving several cytological laboratories, did not clearly confirm this result (Nieminen *et al.,* 2003) (Table 37), showing sensitivity and specificity nearly equal to those of traditional cytological screening (Table 38).

The few randomized prospective studies and other performance studies have shown that automation-assisted screening may be feasible as a part of

Table 37. Comparison of histologically verified cervical lesions between the PAPNET® arm and the conventional screening arm: number (N) and proportion (per 1000) of screenees, odds ratios (OR), with 95% confidence intervals (CI) (logistic regression)

Histological diagnosis	PAPNET arm Total 65 527		Conventional arm Total 25 767		OR	Significance
	N	per 1000	N	per 1000		
Invasive cancer	44	0.67	8	0.31	2.16	$p < 0.05$
In situ carcinoma	79	1.20	18	0.68	1.76	$p < 0.05$
CIN 3	124	1.89	44	1.70	1.11	NS

NS, not significant
From Kok & Boon (1996)

Histological diagnosis	PAPNET arm Total 36 225		Conventional arm Total 72 461		OR	95% CI
	N	per 1000	N	per 1000		
Invasive cancer	3	0.08	4	0.06	1.50	0.30–6.80
CIN 3	51	1.4	100	1.4	1.02	0.72–1.42
CIN 2	51	1.4	104	1.4	0.98	0.70–1.36
CIN 1	40	1.1	96	1.3	0.83	0.57–1.20
Normal and other	36 080	996	72 157	996	1.00	Reference

From Nieminen *et al.* (2003)

Table 38. Specificity of the PAPNET and conventional Pap-smear test with cut-off levels of ASCUS+ and LSIL+ for invasive cancer and for an outcome of CIN 2+ or invasive cancer in primary screening setting

	Negative histology	Negative Pap smear	Specificity %
Cytological threshold: ASCUS+			
Outcome: invasive cancer			
PAPNET	36 222	33 447	92.3
Conventional	72 453	67 241	92.8
Outcome: CIN2+			
PAPNET	36 171	33 447	92.5
Conventional	72 353	67 240	92.9
Cytological threshold: LSIL+			
Outcome: invasive cancer			
PAPNET	36 222	35 972	99.3
Conventional	72 453	71 890	99.2
Outcome: CIN2+			
PAPNET	36 171	35 970	99.4
Conventional	72 353	71 887	99.4

From Nieminen *et al.* (2003)

routine primary screening and with the devices tested it seems to perform at least as well as conventional screening in an organized well functioning programme. Automation-assisted screening may improve the results of a suboptimal screening organization, but may have no advantage over a well organized, high-quality screening programme other than possibly handling more samples with same quality.

A new generation of automated devices for use with liquid-based cytology is now being launched, the performance of which has not yet been evaluated in randomized trials

Physical real-time devices

Advantages of physical real-time devices would be to allow non-invasive on-spot diagnosis, an ability to acquire an objective machine-generated result, applicability in primary health-care settings, the non-requirement for highly trained colposcopists, and high acceptability by women (Soler & Blumenthal, 2000; Basen-Engquist *et al.,* 2003; Wright, 2003).

Normal and neoplastic cervical epithelia have different physical and biochemical properties, yielding distinct patterns in conductance of electrical pulses and reflectance of light waves (Mahadevan *et al.,* 1993; Ramanujam *et al.,* 1994; Richards-Kortum *et al.,* 1994; Wright *et al.,* 2002c). These differences have been applied in fluorescent spectroscopic devices which capture electro-physical signals from the stimulated cervix and analyse the patterns using algorithms to discriminate between normal and neoplastic tissue (Burke *et al.,* 1999; Follen Mitchell *et al.,* 1999). The TruScreen (formerly Polarprobe, Polartechnics Limited, Sydney, Australia) is a portable device that measures the response of the cervical surface to low-voltage electric stimuli and light waves of four different wavelengths. The sensor captures the emitted signals and computer software integrates the information and provides a diagnosis of CIN or the absence of CIN (Mould & Singer, 1997; Singer *et al.,* 2003). In a study of 651 women in ten international centres, the relative sensitivity for histologically confirmed CIN 2 or worse lesions as diagnosed by TruScreen was 70%; the corresponding sensitivity for cytology was 69% and for a combination of TruScreen and cytology was 93% (Singer *et al.,* 2003).

Fluorescence spectroscopy is based on the measurement of autofluorescence from tissue molecules such as FAD, NADPH and collagen that emit light after excitation with low-power laser light of certain wavelengths (Burke *et al.,* 1999; Ferris *et al.,* 2001a; Follen Mitchell *et al.,* 1999). Multi-modal spectroscopy integrates several types of spectroscopy such as intrinsic fluorescence, diffuse reflectance and light scattering. These integrated techniques allow the examination of both biochemical characteristics and morphological features (nuclear size, blood perfusion, cellular changes), so as to optimize the distinction between normal and abnormal tissue (Nordstrom *et al.,* 2001; Georgakoudi *et al.,* 2002). Spectroscopic instruments are continually being improved (Drzek *et al.,* 2003).

The performance of fluorescence spectroscopic devices has been found promising in several small trials, usually conducted by the manufacturer and most often on selected groups of women. In a series of 111 women, accuracy to detect CIN 2 or worse lesions was higher for multimodal spectroscopy than for cytology (area under the ROC curve (AUC): 95% for spectroscopy and 78% for cytology) (Ferris *et al.,* 2001a). Larger multicentre trials are needed to confirm these preliminary results.

Molecular surrogate markers

Because both HPV DNA testing and cytological screening yield considerable numbers of women to be referred for colposcopy and biopsy, who are subsequently found not to harbour high-grade disease, it seems worthwhile to look for markers which, at one test occasion, might identify women susceptible to progression with a high predictive value.

Biomolecular pathways leading from HPV infection to the development of cervical dysplasia and cancer are becoming well understood (zur Hausen, 2000, 2002). Continued expression of the viral early oncogenes E6 and E7 appears to be an essential factor in the neoplastic transformation and maintenance of immortalized cell lines, apparently by inactivation of the tumour-suppressor proteins p53 and pRb, respectively (zur Hausen, 1994, 2000) (see Chapter 1). Certain key DNA, RNA or protein markers arising during the neoplastic transformation process might be measured to predict the progressive character of disease in screening, diagnosis and prognosis. However, it is not clear that one common molecular carcinogenetic pathway is involved and it is therefore possible that no single molecular marker will ever on its own allow distinction between progressive and non-progressive disease.

Potential markers of progression include messenger RNA for the E6 or E7 proteins, HPV DNA sequences integrated into the human genome, over-expression of cell cycle regulator proteins or proliferation protein markers, and determination of certain genetic or immunological profiles (Sotlar *et al.,* 1998; Arias-Pulido *et al.,* 2002; Bibbo *et al.,* 2002; Kadish *et al.,* 2002; von Knebel Doeberitz, 2002; Altiok, 2003; Sherman, 2003; Solomon, 2003; Wang & Hildesheim, 2003) (see Table 39).

Table 39. Potential markers for HPV-induced cervical intraepithelial neoplasia or cancer

Marker	Change in expression	Family	Rationale
Protein markers			
$p16^{INK4A}$	Increased	Cyclin-dependent kinase (CDK) inhibitor	E7-mediated degradation of the Rb gene gene yields enhanced transcription of the gene coding for p16.
p53	Decreased	Anti-tumour regulating protein, involved in apoptosis	Viral E6 protein from oncogenic HPV types binds p53, facilitating its degradation through the ubiquitation pathway.
Ki-67	Increased	Cell proliferation marker	Abnormal cell proliferation beyond basal cell layers.
PCNA (proliferating cell nuclear antigen)	Increased	Cell proliferation marker	Abnormal cell proliferation beyond basal cell layers.
Cyclin E	Increased	Protein associating with CDK2	Cyclin E associated with CDK2 drives cells from G1 to S phase through phosphorylation of pRb and other targets.
Mcm5, Cdc6 c-myc	Increased	Proliferation markers	Abnormal cell proliferation beyond basal cell layers.
Telomerase	Increased	Nucleoprotein consisting of hTR (RNA) and hTERT (enzyme)	Controls length of telomeres and plays a role in cell immortalization
RNA markers			
E7 or E6 mRNA	Presence	Viral mRNA, trancripts of E6 or E7 gene	Presence of mRNA for E6 or E7 indicates active expression of of oncogenes. Presence of E6 or E7 mRNA in the absence of L1 HPV DNA might indicate integration of viral DNA.
DNA markers			
	Decrease in the ratio E2/E6 and E2/E7 viral DNA	Viral DNA sequences integrated in host genome	Viral integration often occurs at the E2 gene of the HPV genome. Disruption of the E2 gene yields a more intensive transcription of the oncogenes E6 and E7. In the episomal state E2 and E6 DNA are present in equal amounts, while in the integrated form, less or no intact E2 is present.
		Markers of genetic host – polymorphism of p53 gene	p53 with arginine at position 72 of p53 should have a higher affinity for the E6 oncogene. Arg/Arg homozygotes should therefore have a higher risk of cervical neoplasia than Arg/Pro heterozygotes or Pro/Pro homozygotes[a]

[a] Increased risk for CIN or cervical cancer among Arg/Arg homozygotes is an inconsistent finding in the literature (see Chapter 1).

p16

Cyclin-dependent kinases (CDKs), cyclins and CDK inhibitors are key molecules that control the cell cycle and coordinate DNA synthesis, chromosome separation and cell division (Morgan, 1997). Viral oncoproteins interfere directly or indirectly with several of these CDKs (Cho *et al.,* 2002). p16INK4A inhibits the CDK4/6 interaction with cyclin D1, preventing progression through the G1/S checkpoint of the cell cycle (Keating *et al.,* 2001). Accumulation of p16INK4a mRNA and protein has been reported in response to inactivation of the retinoblastoma gene product (pRb) through binding with viral E7 (Xiong *et al.,* 1993; Serrano, 1997). However, this overexpressed p16 is inert since pRb-function is neutralized by E7 (Medema *et al.,* 1995; Khleif *et al.,* 1996). Overexpression of p16 protein is considered to be a marker for progression from HPV infection to cervical cancer (von Knebel-Doeberitz, 2002).

Immuno-detection of p16 using monoclonal antibodies in histological material was described by Klaes *et al.* (2001). Other studies of p16 immunoreactivity in histological material with different levels of abnormality and HPV infection status (Table 40) have used various types of primary and secondary antibody and chromogen. The sensitivity of p16-immunohistochemical detection of CIN 2 or worse lesions in histological preparations varied between 70% and 100%, while the specificity ranged from 34% to 100% (Keating *et al.,* 2001; Bibbo *et al.,* 2003; Murphy *et al.,* 2003; Negri *et al.,* 2003; Zielinski *et al.,* 2003; Wang *et al.*, 2004b).

Klaes *et al.* (2002) showed improved inter-observer concordance in histological interpretation of p16-immunostained material (group kappa = 0.94; 95% CI 0.84–0.99) compared with haematoxylin–eosin stained material (group kappa = 0.71; 95% CI 0.65–0.78).

The potential use of p16 immunostaining in cytological smears, and the correlation with cytological and histological results, has been examined recently (Table 41). p16 immunostaining has been found to facilitate the retrieval of dysplastic cervical cells on a slide (Bibbo *et al.,* 2002, 2003; Sahebali *et al.,* 2004).

Strong nuclear and cytoplasmic p16 staining in conventional or liquid-based cervical smears gave a sensitivity for detection of HSIL or worse of about 90–100%; the specificity varied between 36% and 100% (Klaes *et al.,* 2001; Saqi *et al.,* 2002; Murphy *et al.,* 2003; Nassar *et al.,* 2003). Sporadic immunoreactivity in normal squamous metaplastic, inflammatory cells and more systematic staining of endometrial and tubal metaplastic cells and of bacteria has been reported (Bibbo *et al.,* 2002, 2003; Saqi *et al.,* 2002; Riethdorf *et al.,* 2002).

These accuracy measures have been computed from very small and highly selective series and cannot be considered representative for real screening or clinical situations, but the results are promising.

Possible advantages of immunostaining of protein markers include the higher reproducibility of microscopic interpretation, quicker detection of stained lesions and appropriateness for automated detection. A disadvantage is the presence of background staining and positive staining of endometrial or tubal cells, requiring the determination of criteria to define positivity, balancing the sensitivity against specificity.

Biochemical detection of p16 protein in lysates of cervical swab samples using a sandwich ELISA assay is a potentially simple approach for resource-poor settings (Herkert *et al.,* 2004).

Ki-67

Expression of the Ki-67 protein occurs in proliferating cells and its presence is normally confined to the basal or suprabasal epithelial cell layers. Expression of Ki-67 allows distinction of atrophic cells (negative for Ki-67) from neoplastic cells (positive for Ki-67) in menopausal women (Ejersbo *et al.,* 1999; Mittal *et al.,* 1999; Bulten *et al.,* 2000). Expression beyond the inner third of the cervical epithelium is observed in case of CIN and cancer (Bulten *et al.,* 1996; Keating *et al.,* 2001). Several authors have found a significant correlation between the presence or intensity of Ki-67 and the severity of cytological abnormality in cytological preparations (Dunton *et al.,* 1997; Sahebali *et al.,* 2003). Dunton *et al.* (1997) found a sensitivity of 89% for Ki-67 immunostaining in a set of selected abnormal smears for detection of histologically confirmed CIN 2+ lesions, whereas the specificity was 65%.

Other proliferation or cell cycle regulating markers

Several other proteins are overexpressed in proliferating cells and certain cell progression regulators have been proposed as potential markers for cervical neoplasia, such as proliferating cell nuclear antigen (PCNA) (Demeter *et al.,* 1994; Mittal *et al.,* 1993), Mcm5 and Cdc6 (Williams *et al.,* 1998) and cyclin E (Altiok, 2003).

Proliferation markers are physiologically present in basal or para-basal epithelial cells, and are an objective indicator of neoplasia when observed beyond the lower cell layers. In cervical smears lacking architectural information, the presence of proliferation markers is less informative and can easily yield false positive results.

Telomerase

Telomeres are repeated arrays of six nucleotides (TTAGGG) at the chromosome ends that protect chromosomes against degradation and aberrant fusion or recombination (Collins &

Table 40. Overview of p16 immunoreactivity in histological material by severity of lesion and by HPV status

Reference	Detection system: primary and antibody and chromogen	p16 positivity	Lesion, HPV status	*N*	% p16+
Sano *et al.* (1998)	-Mouse monoclolnal antibody (JC8) -Biotinylated horse anti-mouse antibody -Chromogen: DAB	Diffuse	HPV 16 Other hrHPV+ HPV 6 or 11	60 28 34	100.0% 96.4% 0.0%
Keating *et al.* (2001)	-Clone G175-405 (Pharmigen, -Biotinylated goat anti-mouse antibody -Chromogen: DAB	Diffuse	Normal LSIL HSIL hrHPV+ and CIN	24 24 37 40	0.0% 37.5% 70.3% 70.0%
Klaes *et al.* (2001)	-Clone E6H4 (MTM Lab., Heidelberg) -Biotinylated horse anti-mouse antibody -Chromogen: aminoethylcarbazole with hydrogen peroxide in acetate buffer	Diffuse	Normal, hrHPV– Normal, hrHPV+ Inflammation, hrHPV– Inflammation, hrHPV+ Reserve cell hyperplasia, hrHPV- Reserve cell hyperplasia, hrHPV+ CIN1, hrHPV– CIN1, hrHPV+ CIN2, hrHPV– CIN2, hrHPV+ CIN3, hrHPV– CIN3, hrHPV+ Invasive cancer, hrHPV– Invasive cancer, hrHPV+	32 10 30 18 13 8 32 15 14 18 9 51 5 55	0.0% 0.0% 0.0% 0.0% 0.0% 12.5% 46.9% 86.7% 100.0% 100.0% 100.0% 100.0% 100.0% 96.4%
Bibbo *et al.* (2002)	-Clone E6H4 (MTM Lab., Heidelberg) -Mouse non-avidin-biotin Envision+ polymer (Dako) -Chromogen: DAB	Focal or diffuse	Normal CIN 1 CIN 2 CIN 3	3 19 11 14	0.0% 73.7% 90.9% 100.0%
Bibbo *et al.* (2003)	-Clone E6H4 (MTM Lab, Heidelberg) -Mouse non-avidin-biotin Envision+ polymer (Dako) -Chromogen: DAB	Focal or diffuse	Chronic cervicitis Squamous metaplasia CIN 1 CIN 2 CIN 3	5 2 5 4 11	0.0% 0.0% 40.0% 100% 100%
Murphy *et al.* (2003)	-Clone G175-405 (Pharmingen, San Diego) -Biotinylated universal antibody, avidin–biotin complex (Vector Laboratories, Burlingame) -Chromogen: DAB	>10% positive staining	Normal cGIN CIN 1 CIN 2 CIN 3 Squamous-cell carcinoma Adenocarcinoma	21 5 38 33 46 8 2	0.0% 100.0% 92.1% 72.7% 91.3% 100.0% 100.0%
Negri *et al.* (2003)	-Clone E6H4 (MTM Lab., Heidelberg) -Avidin–biotin kit (Lab. Vision Corp., Fremont) -Chromogen: aminoethylcarbazole	Diffuse	Reactive cells Endocervical glandular atypia Adenocarcinoma in situ Adenocarcinoma	15 4 8 18	0.0% 0.0% 100.0% 94.4%
Zielinksi *et al.* (2003)	-Clone E6H4 (MTM Lab., Heidelberg) -Biotinylated rabbit anti-mouse antibody -Chromogen: DAB or aminoethylcarbazole	Diffuse & strong	Adenocarcinoma, hrHPV– Adenocarcinoma, hrHPV+ Adenocarcinoma endom., hrHPV-	5 20 15	20.0% 95.0% 0.0%

Abbreviations: DAB, 3,3'-diaminobenzidine; hr, high risk; cGIN, cervical glandular intraepithelial neoplasia

Table 41. Overview of p16 immunoreactivity in cervical smears material by severity of cytological abnormality

Reference	Detection system: primary antibody and chromogen	Preparation	Cytological lesion/histological lesion in corresponding biopsy	*N*	% p16+
Klaes *et al.* (2001)	-Clone E6H4 (MTM Lab., Heidelberg -Biotinylated horse anti-mouse antibody -Chromogen: aminoethylcarbazole	Conventional smears	Pap I/II Pap IIID+ (LSIL+)	36 7	0.0% 100.0%
Bibbo *et al.* (2001)	-Clone E6H4 (MTM Lab., Heidelberg) -Mouse non-avidin-biotin Envision+ polymer (Dako) Chromogen: DAB	ThinPrep LBC	Within normal limits LSIL HSIL	2 26	0.0% 96.2%
Saqi *et al.* (2002)	-p16 antibody (Neomarkers, Fremont) -Envision + system (Dako)	SurePath LBC	Within normal limits AGUS LSIL HSIL Squamous-cell carcinoma Adenocarcinoma	25 5 30 10 1 2	4.0% 60.0% 80.0% 90.0% 100% 100.0%
Bibbo *et al.* (2003)	-Clone E6H4 (MTM Lab., Heidelberg) - Mouse non-avidin-biotin Envision+ polymer (Dako) - Chromogen: DAB	ThinPrep LBC	Chronic cervicitis Squamous metaplasia CIN 1 CIN 2 CIN 3	5 2 5 6 12	0.0% 0.0% 40.0% 83.3% 100.0%
Murphy *et al.* (2003)	-Clone G175-405 Pharmingen, San Diego -Biotinylated universal antibody, avidin-biotin complex (Vector Laboratories, Burlingame) -Chromogen: DAB	ThinPrep LBC	Normal cGIN CIN 1–3	12 1 20	0.0% 100.0% 100.0%
Nassar *et al.* (2003)	-Monoclonal antibody (Neomarkers) Mouse non-avidin-biotin Envision+ polymer (Dako)	Surepath LBC	Not neoplastic Benign cell changes ASCUS LSIL HSIL	10 9 14 4 1	50.0% 11.1% 14.3% 50.0% 100.0%
Negri *et al.* (2003)	-Clone E6H4 (MTM Lab, Heidelberg) -Avidin-biotin kit (Lab Vision Corp., Fremont) -Chromogen: aminoethyl-carbazole	ThinPrep	AGUS	10	100.0%
Nieh *et al.* (2003)	-Clone E6H4 (MTM Lab., Heidelberg) -Mouse non-avidin-biotin Envision+ polymer (Dako) -Chromogen: DAB	ASCUS Pap smears	Reactive CIN 1 CIN 2/3 Squamous carcinoma Adenocarcinoma *in situ*	21 24 17 2 2	0.0% 8.3% 94.1% 100.0% 100.0%

Abbreviations: DAB, 3,3'-diaminobenzidine; LBC, liquid-based cytology; cGIN, cervical glandular intraepithelial neoplasia; AGUS, atypical glandular cells of undetermined significance

Mitchell, 2002). They become progressively shorter as cells multiply, resulting in chromosomal instability and senescence when a critical short length is reached (Counter *et al.,* 1992). The enzyme telomerase is a ribo-nucleoprotein composed of an RNA part (hRT) and a catalytic part (hTERT), which controls telomere length and is believed to play a role in immortalization of cells (Mathon & Lloyd, 2001; Blasco, 2002). Its activity is increased in CIN and cancer. The intensity of telomerase activity is reported to be correlated with the severity of the abnormality in biopsies and in cervical scrapings, but reliable detection of hTR, hTERT and telomerase activity is still limited by analytical deficiencies (Oh *et al.,* 2001; Jarboe *et al.,* 2002; Fu *et al.,* 2003).

Detection of viral oncogene transcripts

Viral mRNA can be detected using (nested) real-time PCR or nucleic acid sequence-based amplification assay (NASBA) (Smits *et al.,* 1995; Sotlar *et al.,* 1998; Deiman *et al.,* 2002). Presence of viral mRNA transcripts coding for the E6 and E7 proteins from high-risk HPV might be a more specific predictor of progressive infection than simple presence of HPV DNA (Nakagawa *et al.,* 2000; Cuschieri *et al.,* 2004). A commercial kit exists (PreTect HPV-Proofer, NorChip AS, Klokkarstua, Norway) which detects E6 mRNA from HPV16 and E7 mRNA from HPV types 18, 31, 33 and 45.

The presence of E6 and E7 mRNA and absence of viral L1 DNA (negative test result on consensus PCR) indicate integration of viral DNA in the human genome, yielding enhanced transcription of the E6–E7 sequence. Molden *et al.* (2004) found that rates of HPV-Proofer positivity and presence of HPV DNA (measured with GP5+/6+ consensus PCR and type-specific PCR) increased with the severity of cytological or histological cervical abnormality. Nevertheless, lower proportions of mRNA-positive results were observed in normal cases, ASCUS and LSIL (see Table 42).

Viral DNA integration markers

Testing for HPV integration appears to increase the predictive value that an HPV-positive sample is derived from tissue containing progressive CIN or cervical cancer (Klaes *et al.,* 1999). Viral integration often occurs at the E2 gene of the HPV genome. Disruption of the E2 gene is believed to result in more intensive transcription of the oncogenes E6 and E7. In the episomal state, E2 and E6 DNA are present in equal amounts, while in the integrated form, less intact E2 is present (zur Hausen, 2002). A decrease in E2/E6 DNA ratio assessed with real-time PCR is another potential progression marker. However, other authors have reported exclusively episomal HPV DNA in tumours (Cullen *et al.,* 1991; Pirami *et al.,* 1997).

Micro-array technology

It is believed that profiles of multiple host–virus interaction factors will reveal possibilities for accurate individualized risk assessment and prognosis prediction. By the use of DNA microarray technology or DNA chips, expression of many genes can be analysed at once using only a small amount of sample (Hughes & Shoemaker, 2001). The first step consists in extraction of mRNA from a tissue sample. Using reverse transcriptase, complementary DNA (cDNA) is synthesized, which is labelled with a fluorescent molecule. This labelled cDNA is subsequently divided over a slide or membrane where hundreds or thousands of known target DNA sequences are fixed. Hybridization of the labelled cDNA with target DNA is detected as a coloured light signal at a particular locus on the array, which indicates expression of a particular gene.

Post-translational changes also play an important role in pathogenesis, and can be studied using protein arrays or proteomics techniques (Wulfkuhle *et al.,* 2003; Lee *et al.,* 2004).

Table 42. Positivity rate for HPV DNA and HPV mRNA in a series of 4136 women presenting at an outpatient gynaecological service, Oslo, Norway

	Normal	ASCUS	LSIL	CIN 2	CIN 3	Squamous cancer
N	3950	57	20	5	12	1
HPV mRNA+	95	12	6	2	9	1
	2.4%	*21.1%*	*30.0%*	*40.0%*	*75.0%*	*100.0%*
HPV DNA+	368	27	15	2	10	1
	9.3%	*47.4%*	*75.0%*	*40.0%*	*83.3%*	*100.0%*
p-value	<0.0001	0.08	0.009	1.00	0.62	1.00

From Molden *et al.* (2004)

General comments

Research on molecular markers has so far been largely restricted to correlation studies documenting the presence or absence or the intensity of the considered marker in cytological or histological material from selected women. These test accuracy measures can be assessed for detection of CIN 2+ but are not representative for real screening, triage or follow-up settings.

Potential advantages from the use of molecular markers in future clinical practice include: triage of women with minor cytological abnormalities (ASCUS and LSIL) with higher specificity than HPV DNA detection; improvement of the accuracy of histology as gold standard for screening test assessment, by more accurate and reproducible classification of histological squamous and glandular cervical lesions and clearer distinction between cervical and endometrial glandular lesions; selection of best treatment procedures; prognosis prediction; and last but not least, more accurate primary screening for cervical progressive cancer precursors.

Combinations of different modalities

As the previous sections have demonstrated, no single currently available screening test for cervical cancer provides an optimal trade-off between sensitivity and specificity. Because various screening techniques are available, applying them in combinations might be advantageous. Although combinations of tests necessarily require extra resources, the added testing accuracy might increase the detection of treatable disease and allow lengthened screening intervals. Research is in progress to find the combinations that are most complementary, to determine how these tests should be combined, and to clarify how to interpret and manage the increased complexity of results. The results so far are promising but far from complete.

Screening with more than one technique

The main classes of available screening techniques are cytology, visual inspection and HPV DNA testing. It is possible to consider combinations of two techniques within a class (e.g., conventional and liquid-based cytology), but most interest has focused on combining two techniques of different classes in the hope of gaining benefit from complementarity. Thus, researchers have examined cytology plus visual techniques, cytology plus HPV DNA testing and, to a limited degree, HPV DNA testing plus visual techniques. Because of the practical limitations of resources, there has been only occasional interest in combining more than two techniques (Reid *et al.,* 1991).

Whenever two screening techniques are combined, with abnormal results from either test taken to constitute an overall positive result, the sensitivity will be higher than that of either test alone (Franco & Ferenczy, 1999). However, the key question is whether the increase in sensitivity is sufficiently greater than random to merit consideration. Increased sensitivity will typically lead to an offsetting decrease in specificity and the trade-off must be examined to determine the overall effect of the combination on screening accuracy. Various statistical methods for evaluating the added value of adding a second test have been suggested, but none has been fully accepted. The best statistical methods generate roughly equivalent conclusions (Macaskill *et al.,* 2002; Ferreccio *et al.,* 2003), although the interpretations depend on varying regional standards of acceptable safety and cost.

The studies on combination of different modalities have been run with designs that provide estimates of sensitivity and specificity that are not totally comparable. Most of the designs were cross-sectional without correction for verification bias (see Chapter 4). Therefore, the results and conclusions depend on the number of tests applied. If two tests only are considered, the cross-sectional sensitivity estimate for the test combination is bound to be 100%. The same applies, albeit not as a logical consequence, if too few women with dual negative tests are subjected to a commonly accepted reference standard such as colposcopic examination with guided biopsy. Colposcopy itself is not sufficiently sensitive to rule out missed disease and therefore its own errors must be recognized when considering results that depend upon it. In general, the sensitivity estimate of any combination of two tests is smaller if more tests are used for detection of disease and those who tested negative on every test are not subjected to the reference standard (colposcopy). Only one study (Sherman *et al.,* 2003b) has been based on interval cancer incidence, the ideal to estimate the true sensitivity (see Chapter 4). In the absence of data on the expected incidence, the risk in screen positives versus that in screen negatives was used as an indicator of sensitivity.

Cytology plus HPV testing

The residual cytology specimen from liquid-based cytology or a co-collected specimen can be tested for oncogenic HPV types. There is much evidence that screening of women with both cytology and HPV DNA tests increases sensitivity for detection of prevalent CIN 3 or cancer sufficiently to permit longer screening intervals than with cytology alone. After consideration of the accumulated evidence regarding increased sensitivity, decreased specificity and the possibility of lengthened screening intervals using the combination, the US FDA

approved HC2 for HPV DNA testing as an adjunct to cytological testing for women aged 30 years and older. The supporting evidence has been summarized by Franco (2003), with additional recent support (Cuzick *et al.,* 2003; Ferreccio *et al.,* 2003). However, in two other recent studies (de Cremoux *et al.,* 2003; Coste *et al.,* 2003), HC2 performed worse than in other published studies and conventional cytology performed considerably better than is usually reported. Several of the supportive studies are summarized in Table 43. The increase in sensitivity from adding HPV testing was generally greater than the decrease in specificity and, in some studies, cytology actually added little to the performance of HPV testing.

In practice, the introduction of HPV testing into routine screening in combination with cytology produces multiple risk strata ranging from very low to very high absolute risk (positive predictive value) of prevalent or incipient CIN 3 or cancer. A woman with HSIL cytology (especially with a positive oncogenic HPV test) has a high risk of underlying CIN 3. In contrast, negative cytology (whether conventional or liquid-based) with a negative HPV test is associated with a risk of CIN 3 within two years that approaches zero (Ferreccio *et al.,* 2003). These extremes of risk stratification are easily managed. However, strategies to manage the very large numbers of women who are HPV-positive and cytology-negative need to be developed and evaluated. Repeating viral and cytological tests between six and twelve months has been proposed (Wright *et al.,* 2004) as an interim measure until more data are available to develop truly evidence-based guidelines.

The negative predictive value of adding an HPV test to cytology is its major utility. HPV infection is so common that a positive test conveys only a moderate positive predictive value for prevalent or incipient CIN 3 and cancer (Sherman *et al.,* 2003b). However, because persistent infection with oncogenic types of HPV is the necessary cause of virtually all cases of cervical cancer, a negative test for oncogenic HPV has unusually high negative predictive value, or reassurance, in the context of negative or even mildly abnormal cytology. The uncommon combination of a negative HPV test and an HSIL cytological result merits further evaluation because of its rarity and the possibility that one of the two results is in error.

Cytology plus visual techniques

The interest in combining cytology with some kind of visual inspection is natural, since cytology and colposcopy have comprised the strategy responsible for a half century of successful screening for cervical cancer. Population screening using cytology and colposcopy concurrently has been explored (Olatunbosun *et al.,* 1991), but is much too expensive and demanding of limited expertise for use in most regions. The search for an easier alternative to colposcopy has led to studies of cytology and cervicography and of cytology and direct visual inspection. Two representative studies to examine these combinations are summarized in Table 44. Cervicography is no longer available commercially, but it is worth noting that studies of cervicography as an adjunct have suggested that a visual technique might complement conventional cytology. The two kinds of technique tend to detect different groups of women with CIN 2 or 3 without an obviously unacceptable loss in specificity (Ferreccio *et al.,* 2003), resulting in overall increased accuracy. The cost-effectiveness of combining cytology with some kind of visual inspection (Shastri *et al.,* 2004) should be compared with alternative strategies and evaluated more fully and formally on a regional basis. Real-time cervical scanning based optico-electrical devices might improve the sensitivity of cytology (Singer *et al.,* 2003), but the influence on specificity of adding such new techniques is not yet clear.

HPV plus visual inspection

In many developing countries, approaches that do not rely on an extensive infrastructure of highly trained personnel must be considered. Because first-rate cytological screening programmes are difficult to create and maintain, there is interest in establishing programmes that rely on more easily performed and standardized techniques. It may be feasible to combine an HPV test for primary screening with triage modalities other than cytology, such as direct visual assessment by non-physician health-care providers. The use of HPV testing plus visual inspection is best considered as a sequential strategy, to maximize sensitivity with acceptable specificity (see below).

Combining two techniques from the same class

The combination of two cytological techniques can be seen as a logical extension of re-screening of slides or of computer-assisted screening, as discussed in the first section of this chapter. The combination of conventional and liquid-based cytology is not particularly complementary (Ferreccio *et al.,* 2003) and holds little promise because of the expense of conducting the two tests for each woman screened. Similarly, there is probably no reason to combine two visual techniques (e.g., cervicography and direct inspection) because of the correlated nature of the results and limitations of the techniques (Shastri *et al.,* 2004). HPV testing with multiple techniques could reveal some testing errors, but not enough to justify the high cost.

Table 43. Combined use of cytology and HPV DNA testing for screening

Study	No. of tests	Case identification	Colposcopy for negatives	No. of women	No. of CIN	CIN threshold	Test	Cut-points	Sens. %	Spec. %	PPV %	NPV %
Ratnam *et al.* (2000)	2	P	8	2098	30	CIN2/3	Cytol.	ASCUS	40	92	11	98.4
							HC1, HC2	10,1 pg/ml	68	91	15	99.1
							Cytol.+HC	Either +	76	86	12	99.3
Wright *et al.* (2000)	4	P	No	1356	56	CIN2/3	Cytol.	ASCUS	68	88	18	96.9
							HC2	1 pg/ml	84	85	17	98.3
							Cytol.+HC2	Either +	86	76	13	98.6
Belinson *et al.* (2001)	5	P	No	1997	86	CIN2/3	LBC	ASCUS	94	78	16	99.7
							HC2	1 pg/ml	95	85	23	99.8
							LBC+HC2	Either +	92	93	39	99.6
Salmeron *et al.* (2003)	3	P	No	7732	93	CIN3	Cytol.	ASCUS	59	98	36	99.5
							HC2	1 pg/ml	93	92	15	99.9
							Cytol.+HC2	Either +	98	91	12	99.9
Petry *et al.* (2003)	2	P	5	7908	37	CIN3	Cytol.	ASCUS	46	98	10	99.7
							HC2	1 pg/ml	97	95	9	100
							Cytol.+HC2	Either +	100	95	8	100
Sherman *et al.* (2003a)	2	P	No	20810	171	CIN3	Cytol.	ASCUS	34	97	10	98.6
							HC2	1 pg/ml	64	86	7	99.1
							Cytol.+HC2	Either +	72	85	7	99.2
Ferreccio *et al.* (2003)	4	P+I	No	8551	110	CIN3	Cytol.	ASCUS	63	94	12	99.5
							LBC	ASCUS	86	88	9	99.8
							PCR	+/–	85	88	9	99.8
							Cytol.+PCR	Either +	91	88	9	99.9
							LBC+PCR	Either +	91	88	9	99.9
Cuzick *et al.* (2003)	2	P+I	5	10358	90	CIN2/3	Cytol.	Borderline	77	96	16	
								Mild	70	99	34	
							HC2	1 pg/ml	97	93	13	
								2 pg/ml	96	94	15	
							Cytol.+HC2	Mild/2 pg	100	94	14	

P, detected at screening; I, interval cancers; Cytol, conventional cytology; LBC liquid-based cytology; PCR, polymerase chain reaction; HC, hybrid capture; PPV, positive predictive value; NPV, negative predictive value

Table 44. Combined use of cytology and visual-based methods for screening: two representative studies

Study	No. of cases	No. of CIN	CIN	Test	Cut-points	Sensit. (%)	Specif. (%)	PPV (%)	NPV (%)
Ferreccio *et al.* (2003)	8551	110	CIN3	Cytol.	ASCUS	63	94	12	99.5
				LBC	ASCUS	86	88	9	99.8
				Cervicog.	A	62	85	5	99.4
				Cytol. + Cervicog.	LSIL or P	75	91	10	99.6
				LBC + Cervicog.	ASCUS or P	93	84	7	99.9
Shastri *et al.* (2004)	4039	57	CIN2/3	Cytol.	LSIL	57	99	38	99.4
				VIA	P	60	88	7	99.3
				VILI	P	75	84	7	99.6
				Cytol. + VIA	LSIL or P	83	87	9	99.7
				Cytol. + VILI	LSIL or P	89	83	7	99.8

Abbreviations: Cytol. = conventional cytology; LBC = liquid-based cytology; A = equivocal cervigram; P = positive cervigram; PPV, positive predictive value; NPV, negative predictive value; VIA, visual inspection with acetic acid; VILI, visual inspection with Lugol's iodine

Sequential tests (triage)

The classical scheme of secondary cancer prevention consists of three steps: sensitive screening of asymptomatic individuals to identify those at risk of disease, specific diagnosis of the disease state and treatment of those with cancer or a cancer precursor (see Figure 45). Triage is an additional step interposed between screening and diagnosis to further stratify individuals with positive primary screening results according to their risk for the disease state. In other words, a second test is performed only if the first test is neither completely normal nor definitely indicative of need for treatment. In this respect, triage is conceptually related to the consideration of residual error in stepwise prediction models.

The utility of a triage test in the context of a cervical cancer screening programme will depend not only on the performance characteristics of the test itself in relation to the primary test, but also on the target screening population, the prevalence of disease, the cost of follow-up, the available resources (logistic and monetary) and patient compliance (Solomon, 2003). Triage is of most value when the screening test lacks specificity and/or the diagnostic procedure is expensive or a limited resource. An efficient triage test should reduce overtreatment, patient anxiety and inconvenience, as well as overall management costs, usually by reducing the number of diagnostic procedures performed—all without excessively sacrificing sensitivity for detection of disease. However, the sequential use of imperfect tests tends invariably to reduce sensitivity somewhat.

Triage of cytology

In cervical cancer screening, the primary test has traditionally been a programme of repeated cytological tests, which generally succeeds because of the typically long natural history of HPV persistence leading to cervical carcinogenesis. A single cytological test is not sufficiently sensitive to serve as an adequate screening test. In an effort to maximize sensitivity and negative predictive value, a test finding of ASCUS or above is used as the threshold for referral for additional follow-up in the USA. ASCUS is a common cytological interpretation, applied in approximately 5% of screening cytological tests. Similarly, in the United Kingdom, approximately 4% of all smears show borderline or mildly dyskaryotic changes. The threshold of test positivity at equivocal cytology substantially increases sensitivity for identifying histological CIN 2/3 (Kinney *et al.,* 1998), but at the cost of repeated cytology or referring millions of women for colposcopy and biopsy. Many of these women are not infected with oncogenic HPV types, and some 90% do not have prevalent CIN 2 or CIN 3 and are not destined to develop it in the immediate future. In this setting of lower specificity, a triage test that further stratifies women according to cancer risk is appealing.

A multicentre, randomized clinical trial was conducted by the US National Cancer Institute to compare different strategies for managing the 2–3 million women with ASCUS and 1.25 million women with LSIL cytological results in the USA each year (Schiffman & Adrianza, 2000). About 40–50% of women with ASCUS are HPV-positive (Manos *et al.,* 1999; Solomon *et al.,* 2001), the actual proportion depending on the patient population and the

cytomorphological threshold utilized. Importantly, virtually all of the occult CIN 2 or 3 associated with ASCUS is found in the HPV-positive fraction. Therefore, in the context of ASCUS cytology, triage by HPV testing can save approximately 50% of women from unnecessary colposcopy without compromising sensitivity.

This result has been supported in a recent meta-analysis of 15 comparable studies of HPV DNA testing as an alternative to repeat cytology in women who had equivocal results on a previous cytological test (Arbyn *et al.,* 2004b). The pooled sensitivity and specificity to detect histologically confirmed CIN 2 or worse were 84.4% (95% CI 77.6–91.1%) and 72.9% (95% CI 62.5–83.3%), respectively, for overall HPV testing. Restriction to the nine studies (Table 45) where the HC2 assay was used yielded a pooled sensitivity of 94.8% (95% CI 92.7–96.9%) and a pooled specificity of 67.3% (95% CI 58.2–76.4%).

Consensus management guidelines for follow-up of an ASCUS cytology result, based on the accumulated evidence, were developed for the USA under the sponsorship of the American Society for Colposcopy and Cervical Pathology (ASCCP). Acceptable options following an ASCUS cytological interpretation include repeat cytology, immediate colposcopy or HPV testing (Wright *et al.,* 2002c; American College of Obstetricians and Gynecologists, 2003).

Triage by HPV DNA testing of women with ASCUS is now very common in the USA, where, if initial liquid-based cytology is used, 'reflex' HPV testing (see Glossary) is considered the preferred triage approach, as it obviates the need for a repeat visit (Wright *et al.,* 2002c). HPV testing has also been recommended to be introduced in the United Kingdom for borderline cytological cases on a pilot basis (Cuzick *et al.,* 1999a, b) and the HART study in women over the age of 30 years has confirmed the validity of this approach (Cuzick *et al.,* 2003). With conventional cytology smears, there is no residual sample available for HPV testing, but in the USA, an additional specimen is now often co-collected to be used for triage if an ASCUS interpretation is obtained (otherwise the co-collected specimen is discarded). When oncogenic HPV is detected in conjunction with an ASCUS cytological interpretation, the tendency at present is to report both findings, rather than to upgrade the cytological interpretation to SIL (Levi *et al.,* 2003).

The ASCUS-LSIL Triage Study (ALTS) and other studies have shown that cytologically identified LSIL, when interpreted stringently (as in, for example, France, Sweden and the USA) is so highly associated with HPV that an HPV triage test (as a sequential test) is not useful, due to low specificity (ASCUS-LSIL Triage Study (ALTS) Group, 2000, 2003a; Arbyn *et al.,* 2002; Scott *et al.,* 2002). In a meta-analysis of studies based on HC2 for detection of CIN 3 (Arbyn *et al.,* 2002), the estimates of relative sensitivity, specificity, positive predictive value and negative predictive value were 95.7% (95% CI 91.2–100), 32.9% (95% CI 17.8–48.0), 32.4% (95% CI 13.4–51.3) and 98.8% (95% CI 97.1–100), respectively. Cuzick *et al.* (2003) suggested that HPV testing might be useful for triage of mild dyskaryosis based on high negative predictive value, despite the high HPV prevalence associated with this cytological interpretation.

Atypical glandular endocervical cells (AGC), the glandular counterpart of ASC, is a much less common cytological interpretation with a higher risk for underlying precancerous lesions or cancer than ASCUS. Current ASCCP guidelines recommend colposcopic evaluation with endocervical sampling for all women with AGC or 'atypical endocervical cells' (Wright *et al.,* 2002c). One study has suggested the possible utility of HPV triage following an AGC result (Ronnett *et al.,* 1999). Finally, colposcopic referral is recommended for another relatively uncommon equivocal interpretation, ASC-H, because of the high risk of underlying CIN 2 or 3 (Wright *et al.,* 2002c).

International variation in cytological terminology, compounded by the use of different morphological criteria for similarly termed diagnoses, might imply that results could not be generalized between countries (Scott *et al.,* 2002). However, use of an atlas of cytology images, with known HPV status and disease outcome, should allow the performance of HPV triage to be transferred between classification systems and screening programmes without the need for costly repetition of trials (Solomon, 2003).

In regions where expert colposcopic services are limited or expensive, the possibility of triage with another visual technique is attractive. However, several evaluations of triage by cervicography or visual inspection after cytological testing (Costa *et al.,* 2000; Denny *et al.,* 2000b; Mould *et al.,* 2000; Blumenthal *et al.,* 2001; Ferris *et al.,* 2001b) have indicated lower accuracy than HPV DNA testing, with inadequate sensitivity.

HPV first, then triage by cytology or visual inspection

The combination of HPV as an adjunct to cytology may be an interim strategy in an evolution that ultimately leads to primary screening by HPV with triage by cytology. In fact, HPV testing followed by cytology is a rational approach for older women, given the higher sensitivity of HPV testing and the greater specificity of cytology (Sasieni & Cuzick, 2002).

The performance of cytology as a triage test might be very different from

Table 45. Triage of ASCUS cytology by HPV DNA testing (Hybrid Capture 2) for detection of histologically confirmed CIN 2+

Study	Sensitivity	Specificity	PPV	NPV	Test positivity	Prevalence of CIN 2+
Manos *et al.* (1999)	0.892	0.641	0.151	0.988	0.395	0.067
Bergeron *et al.* (2000b)	0.833	0.616	0.208	0.968	0.432	0.108
Fait *et al.* (2000)	0.857	0.971	0.906	0.954	0.235	0.248
Lin *et al.* (2000)	1.000	0.745	0.692	1.000	0.527	0.365
Shlay *et al.* (2000)	0.933	0.739	0.230	0.993	0.313	0.077
Morin *et al.* (2001)	0.895	0.742	0.162	0.992	0.292	0.053
Rebello *et al.* (2001)	0.857	0.759	0.581	0.932	0.413	0.280
Solomon *et al.* (2001)	0.959	0.484	0.196	0.989	0.568	0.116
Zielinski *et al.* (2001)	0.917	0.687	0.149	0.993	0.347	0.056

PPV, positive predictive value; NPV, negative predictive value
Modified from Arbyn *et al.* (2004)

its characteristics as a screening test (Solomon, 2003). If used as a triage test, there would be a dramatic reduction in the number of tests overall and a marked increase in the yield of positive results, altering the ratio of negative to abnormal specimens. It is unclear how this would affect the sensitivity and specificity of cytological testing.

HPV then visual inspection

It might be possible to combine an inexpensive, rapid HPV test of the kinds now under development with simplified visual inspection to produce screening strategies with good characteristics. Evidence supporting this possibility comes from a very few studies where HPV testing and cervicography were both analysed (Ferreccio *et al.,* 2003; Jeronimo *et al.,* 2003). Screening, triage and even treatment services could be combined in the same visit and thereby reduce loss to follow-up in areas remote from health clinics. Simple visual assessment categories (e.g., normal versus lesion treatable by cryotherapy versus lesion requiring a gynaecologist) could be calibrated to maximize reliability and sensitivity if screening was done only once or twice in a woman's lifetime. The use of this approach depends upon the development of a widely applicable inexpensive and rapid HPV test. For such a combined approach, revised (more sensitive) criteria might be needed for visual assessment, to prevent a serious loss of overall sensitivity (Denny *et al.,* 2000b). A disadvantage would be the loss of the limited ability of cervical cytology to detect non-cervical neoplasia, particularly endometrial cancer, in a strategy restricted to HPV and visual inspection.

Triage following visual inspection

Some investigators have considered two-stage cervical cancer screening in which visual inspection would be followed by a second test (Denny *et al.,* 2000b). However, in a simple application requiring both results to be positive before treatment, the overall sensitivity would be limited by the numbers of cases missed by either test, regardless of the order in which they are applied.

Follow-up of positive test results

The conventional confirmatory test following an abnormal primary screening result has been colposcopically directed biopsy. Certain procedures could appear in several of the categories depicted in Figure 45. Thus colposcopy is used in some settings, particularly in Europe, as an adjunctive screening test, but it can also be categorized as a triage modality.

Although colposcopically directed biopsy has been used as the gold standard for diagnosis, recent findings suggest that it misses about a quarter or more of prevalent CIN 3 (ASCUS-LSIL Triage Study (ALTS) Group, 2003b). This implies that women with an apparently negative diagnosis on colposcopy remain at increased cancer risk, possibly requiring more than resumption of routine screening (Viikki *et al.,* 2000). Moreover, the three-stage strategy of screening, requiring a return for histological diagnosis and a third visit for treatment, leads in many regions to unacceptable loss to follow-up. Consequently, there is reason to explore other ways of combining the first test, triage, diagnostic test and management. Women diagnosed with less than CIN 2 by colposcopy are at approximately 10% risk of CIN 2 or CIN 3 within two years; this risk is similar regardless of whether the colposcopically directed biopsy result was 'negative' or 'CIN 1' (Cox *et al.,* 2003). There is insufficient evidence regard-

SCREENING	Triage	DIAGNOSIS	Diagnosis and risk clarification	TREATMENT	*Post-treatment monitoring*
Cytology LBC Smear **** HPV DNA **** VIA Cervico-graphy Colposcopy	Repeat cytology **** HPV DNA **** Colposcopy	Colposcopically directed biopsy and histological diagnosis	Molecular markers of risk ?	Cryotherapy *** Laser therapy **** LEEP **** Cold-knife conization	Cytology **** HPV DNA *** Colposcopy

Figure 45 Sequence of cervical cancer screening and prevention
The classical steps of cancer screening and prevention are screening, diagnosis and treatment, in bold capitals. Triage and Diagnosis and Risk Clarification are steps to clarify the risk of respective subpopulations (adapted from Solomon, 2003).

ing the optimal management of women diagnosed with less than CIN 2 by colposcopically directed biopsy. In ALTS, various re-triage strategies combining follow-up cytology and HPV testing were compared (Guido *et al.,* 2003). A single HPV test at 12 months gave the best trade-off of sensitivity and referral percentage. As an alternative, semi-annual cytological sampling appeared to be useful. Further studies are needed to find assays or strategies that more efficiently identify women with occult CIN 3 and permit the majority of women to safely return to routine screening.

More sensitive screening and triage strategies that translate into increased detection of 'early' and often very small CIN 3 lesions may lead to earlier treatment, but the impact on cancer outcomes has not been established. Many small high-grade lesions might regress, and others could be detected later, when larger but still intraepithelial (Sherman *et al.,* 2002). Very early detection leads to a greater likelihood of overtreatment of lesions, particularly CIN 2, that might otherwise regress. Identifying markers of risk of progression to cancer is a priority in order to reduce unnecessary treatment and attendant complications and costs associated with treating all cases of CIN 2 or 3. Novel approaches were considered in the previous section of this chapter.