In response to an enquiry from the Cervical Cytology Laboratory Review Group

Advice: The SHTG advise that image-assisted cervical cytology using the ThinPrep® imager (Hologic™, Inc) is accepted as a method of cervical cytology screening for use in NHSScotland.

The evidence from the UK MAVARIC trial, which undertook cost effectiveness modeling, predicted that automated reading would dominate manual reading of liquid based cytology (LBC) slides as it was associated with cost-savings, a small gain in quality adjusted life years (QALYs) and an incremental cost-effectiveness ratio of -£7,592 per QALY.

SHTG noted that the evidence on accuracy of image-assisted cytology compared with manual reading of LBC samples is variable.

Advice context
No part of this advice may be used without the whole of the advice being quoted in full.

This advice represents the view of the SHTG at the date noted. It is provided to inform NHS boards in Scotland when determining the place of health technologies for local use. The content of this Advice Statement was accurate and based upon the most up-to-date evidence available at the time of publication. Readers are asked to consider that new trials and technologies may have emerged since first publication and the evidence presented may no longer be current.

SHTG Advice Statements will be reviewed on a 2-yearly basis. The evidence will be updated if requested by the clinical community, dependent on new published reports.

This advice does not override the individual responsibility of health professionals to make decisions in the exercise of their clinical judgment in the circumstances of the individual patient, in consultation with the patient and/or guardian or carer.

Chairman
Scottish Health Technologies Group
SUMMARY OF EVIDENCE

Image-assisted cytology in cervical screening August 2011

Question

What is the clinical and cost effectiveness of the ThinPrep® imager (Hologic™, Inc) for assisted cervical cytology compared with manual reading of liquid based cytology slides?

Evidence Note 34 is available for download from the Healthcare Improvement Scotland website.

This Evidence Note highlights published literature on the clinical and cost effectiveness of the ThinPrep® imager (Hologic™, Inc) for assisted cervical cytology compared with manual reading of liquid based cytology (LBC) slides.

Key points

- A systematic review in 2009 found three robustly designed studies showing no significant difference between image-assisted and manual reading of liquid based cytology (LBC) slides in the number of true cases of CIN2+ identified at any threshold.

- A systematic review found three robustly designed studies showing inconsistent results regarding false positives for CIN2+. Using high grade thresholds: two studies found that image-assisted reading generated significantly fewer false positives than manual reading of LBC slides, and one study found no significant difference. There were no significant differences in false positives for CIN2+ using low grade thresholds in the two studies which examined this.

- In 2011, the MAVARIC (Manual Assessment Versus Automated Reading in Cytology) trial found that image-assisted reading with ThinPrep® was associated with significantly reduced sensitivity for the detection of CIN2+ pathology compared with manual reading, but there was no significant difference between the two methods for detection of CIN3+.

- The Scottish Cervical Cytology ThinPrep® Imager Feasibility Study found similar levels of sensitivity and specificity between image-assisted and manual reading for the detection of all lesions. In phase II only, there was some indication of improved sensitivity in the detection of high grade lesions with image-assisted reading (statistical significance not reported).

- Differences in screening protocol, study design, reference standard, cytologist training and outcome measure make comparison between studies difficult.

- Productivity is improved with image-assisted screening.
Cost-effectiveness modeling in the MAVARIC trial predicted that automated LBC would dominate manual LBC, based on estimates using aggregated data from two automated devices: ThinPrep® and BD SurePath™. Automated reading was associated with cost savings, a small gain in quality adjusted life years (QALYs) and a combined incremental cost-effectiveness ratio (ICER) of £7,592 QALY gained, compared with manual reading, of LBC slides. However, trial authors urged caution in the interpretation of this result.

Health technology description

LBC has been in use in Scotland for cervical screening since 2003. A sample of cells is brushed from the surface of the cervix and the head of the brush rinsed directly into the preserving fluid. The vial containing the sample is sent to the local processing laboratory. A thin layer of cells is applied to a microscope slide and stained. The slides can be read manually or with the aid of an imager.

Six out of 11 Scottish laboratories have been participating in a feasibility study of the ThinPrep® imager1. This uses a ‘near stoichiometric’ haematoxylin as part of the Papanicolaou stain, which makes cells with higher DNA content appear darker. Dyskaryotic cells typically have raised DNA content. The imager guides the cytologist to the 22 fields of view most likely to contain cells of interest. If any abnormal cells are identified the whole slide is viewed manually.

Additional training is required for cytologists to become familiar with the very different presentation of dyskaryosis using the near stoichiometric stain, compared with conventional staining techniques.

Epidemiology

Cervical cancer is caused by sexually transmitted infection with the human papilloma virus (HPV) (although the majority of HPV infections do not lead to cancer)2. Cervical cancer is newly diagnosed in around 300 women per year in Scotland and accounts for around 100 deaths per year3. Smokers are at increased risk due to genetic and immune response changes, and reduced effectiveness of treatment2. Women from more deprived areas have an increased chance of getting or dying from cervical cancer4, due to their increased exposure to risk factors and lower uptake of screening.

The aim of cervical screening is to reduce mortality and morbidity from cervical cancer. Cervical cancer has a long preinvasive stage, during which screening can prevent progression from pre-cancerous squamous cell changes to invasive disease. Screening detects cervical intraepithelial neoplasia (CIN). These are graded 1, 2 or 3, depending on how deeply the surface layer is affected:

- CIN1 (affecting a depth of one third)
- CIN2 (affecting a depth of two thirds), or
- CIN3 (affecting the full thickness).
A national cervical screening programme operates in Scotland. Women aged 20–60 years are invited to attend for a smear test every 3 years. Uptake is around 74% (based on those who have a previous recorded smear within 3.5 years), and is lowest among younger women.

Laboratories process around 420,000 tests per year. Around 3% of samples are not able to be analysed, and are described as unsatisfactory. This occurs for a number of reasons, including insufficient cells in the specimen, infection, excessive blood, and sample preparation artefacts. Samples that can be screened are graded according to the severity of abnormality. Table 1 illustrates the breakdown of results for the most recent year.

<table>
<thead>
<tr>
<th>Smear result</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>91.3</td>
</tr>
<tr>
<td>Borderline</td>
<td>5.2</td>
</tr>
<tr>
<td>Mild dyskaryosis</td>
<td>2.2</td>
</tr>
<tr>
<td>Moderate dyskaryosis</td>
<td>0.6</td>
</tr>
<tr>
<td>Severe dyskaryosis</td>
<td>0.6</td>
</tr>
<tr>
<td>Severe dyskaryosis/invasive</td>
<td>0.02</td>
</tr>
<tr>
<td>Glandular abnormality</td>
<td>0.03</td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
</tr>
</tbody>
</table>

Positive or abnormal smears range from borderline and mild dyskaryosis, which are sometimes described as 'low grade', to moderate dyskaryosis or more severe pathology, described as ‘high grade’. Inter-rater agreement in low grades is poor. The recently introduced HPV vaccination programme among teenage girls in Scotland is expected to lead to a reduction in the number of abnormal smears.

Following a positive smear, a woman is referred for colposcopy. A biopsy may be taken, which would indicate the presence of CIN. CIN3 is the immediate precursor to cervical cancer, and is treated. Lower grades of lesion may resolve spontaneously and not require clinical intervention. Differences exist in morphological diagnostic criteria among different countries’ screening programmes.

**Safety**

Concerns regarding physical and ergonomic discomfort for cytologists from older models of the ThinPrep® imaging system review microscope have been addressed in newer models. Problems of fatigue and decreasing cytologist competence from seeing fewer abnormal samples due to the changing epidemiology are countered by the shorter analysis time allowing more slides to be screened.

For the cervical screening population, an increase in false negative smears would result in increased levels of undiagnosed cervical cancer, and increased risk of death. An increase
in false positive smears would result in overtreatment as more people without cervical cancer would be subject to colposcopy.

**Clinical effectiveness**

This Evidence Note highlights key findings on the comparison of interest by reporting on a large UK trial from 2011\(^5\) and an Australian systematic review from 2009\(^6\). More recent primary studies would not have met the inclusion criteria for the systematic review, due to less robust study designs, and are not reported here. Inclusion of results from the Scottish Cervical Cytology ThinPrep® Imager Feasibility Study\(^1\) gives the Scottish context.

Screening studies require a large number of participants to generate sufficient pathological samples. Histological follow up may be available for women who screen positive; but for women who test negative it is usually unknown how many were correctly identified as not having CIN (true negative), and how many do in fact have the disease (false negative). Sensitivity and specificity of a new test are impossible to calculate in these circumstances. However positive predictive value (proportion of those who test positive who truly have CIN) can be calculated.

A possible strategy for boosting the number of abnormal tests in a screening study is to include seeded cases, which are extra known abnormal smears from colposcopy clinics or some other source. Where such boosting occurs the additional LBC samples are indicated as coming from non-routine sources.

The recently published MAVARIC trial in Manchester included a paired study arm of 24,867 samples (including 3,876 from non-routine sources) in a comparison of ThinPrep® image-assisted cytology with manual screening of LBC slides\(^5\). Training of study personnel was given by the manufacturer. It covered the ThinPrep® system, cell morphology recognition, and practical experience from 20 training slides and 25 test slides. Scores from the test slides ranged from 13 to 25 out of 25. There was no follow-up training or performance monitoring, therefore no opportunity for the correction of systematic errors in individual performance before or during the trial. The reference standard for positive results was HPV testing and colposcopy. Each sample was screened first with the imager, then manually; with the cytologist blinded to the study arm and the automated result at the time of the manual reading. Mathematical properties allow relative sensitivity to be calculated in these circumstances. However, relative specificity cannot be determined exactly. The sensitivity of ThinPrep® relative to manual reading (restricted to age 25–64 years and routine samples) was 0.92 (95% CI 0.87–0.98) for the detection of CIN2+ pathology and 0.97 (95% CI 0.91–1.03) for the detection of CIN3+ pathology. CIN2+ and CIN3+ indicate thresholds of ‘CIN2 or more severe’, and ‘CIN3 or more severe’ respectively. These results indicate that image-assisted reading was less sensitive than manual reading for the detection of CIN2+ pathology, but that there was no significant difference for the detection of CIN3+ pathology.

The Scottish Cervical Cytology ThinPrep® Imager Feasibility Study included approximately 180,000 smear samples in an unpaired study design, whereby smears from different women were allocated either to manual or automated screening\(^1\). Training of study
personnel was given by the manufacturer, and included practical experience with 45 slides. Stain and system validation took place in advance of the study, whereby two cytologists from each laboratory screened 100 slides manually and using the imager. All participants reached the required competency. Ongoing training was given by the manufacturer as required. Very similar proportions of negative, unsatisfactory, low grade, and high grade results were reported between the study arms. For calculations of sensitivity and specificity, the results of the primary screen were considered in relation to the final reported results (following rapid review or preview) as the reference standard, and presented as an unweighted mean over the six laboratories. Phase II results (in which the screening protocol was amended at the cytologists’ request) for the detection of low or high grade lesions were sensitivity 0.95, specificity 0.97 in the automated arm; compared with sensitivity 0.94 and specificity 0.96 in the manual arm (confidence intervals and statistical significance not reported). For the detection of high grade lesions only, mean sensitivity was 0.98 in the automated arm and 0.93 in the manual arm. Correlations were reported between histological follow up of positive cytology and the cytology results for positive smears, separately for each laboratory. These ranged between 0.76 and 0.96 in the automated arm of the trial, and between 0.72 and 0.90 for the manual arm (confidence intervals and statistical significance not reported).

The Australian Medical Services Advisory Committee (MSAC) published a systematic review of the safety, effectiveness and cost effectiveness of image-assisted LBC in 2009, which included a comparison of image-assisted with manual reading of ThinPrep® slides. Three studies met the inclusion criterion of a split sample design (Biscotti et al. (2005), Bolger et al. (2006), Roberts et al. (2007)), whereby the same LBC sample was divided, with one part screened manually and the other part screened using the ThinPrep® imager. The study by Biscotti et al. (2005) included seeded high grade squamous intraepithelial lesions (HSIL) cases, and used a reference standard of adjudicated cytology for all positive or discordant slides and a random 5% of negative slides. In the study by Bolger et al. (2006), the reference standard was adjudicated cytology for all positive slides. In the study by Roberts et al. (2007) there was incomplete verification against a reference standard of histology for high grade results. No significant differences were found in the number of true positives for CIN2+ detected at any threshold between image-assisted and manual reading in any of the studies; and there was no consistent tendency in either direction. Using a threshold of possible or definite HSIL or glandular, Roberts et al. (2007) found significantly fewer false positives for CIN2+ with image-assisted reading. Biscotti et al. (2005) found significantly fewer false positives for CIN2+ using a threshold of high grade, but no significant difference at other thresholds, with image-assisted reading. Bolger et al. (2006) found no significant difference in false positives for CIN2+ at any threshold.

Differences in screening protocols, reference standards, cytologist training and outcome measures render these studies difficult to compare.

Cost effectiveness

For reasons of commercial confidence, cost effectiveness was not reported separately for the two image-assisted screening devices included in the MAVARIC trial: ThinPrep® and
Overall, automated reading was found to be slightly less effective and slightly cheaper than manual reading\(^5\). Productivity was estimated to be 60–80% higher for both automated devices together compared with manual reading\(^5\). Cost per case detected was similar between automated and manual arms, both for CIN2+ and CIN3+ thresholds\(^5\). A Markov model predicted the lifetime (30 years) costs and effects of automated and manual LBC in a simulated cohort of 10,000 women, using both life years and quality adjusted life years (QALYs) as the outcome measure. The perspective was that of the NHS. Model results predicted that automated LBC would be associated with small cost-savings and fewer life years when compared to manual LBC. The cost savings were due to the reduction in treatment costs from less CIN detection and the decreased life years saved was driven by the finding that automated LBC was less sensitive than manual LBC. When quality of life was accounted for, the model predicted that automated LBC dominated manual LBC as it was associated with cost-savings (approximately £126,000) and a slight gain in QALYs (approximately 16) per 10,000 women. The increase in QALYs resulted from a small increase in specificity which reduces the disutility associated with false positive results. The incremental cost-effectiveness ratio (ICER) was -£7,592 per QALY gained by automated compared with manual LBC. However, the authors urged caution in the interpretation of the cost-effectiveness results, highlighting that the quality-of-life weights for health states used in the modeling were obtained from the international literature and may not generalise to the preferences of women in the UK and uncertainties around the duration of disutility for women associated with ‘overtreatment’ of pre-invasive cervical cancer lesions. Sensitivity analyses indicated, however, that the model results were robust in all scenarios tested (including choice of quality-of-life weights) with automated LBC associated with cost savings, a small loss in life years and a small gain in QALYs\(^5\).

The Australian MSAC review presented a cost-effectiveness model comparing both manual and automation-assisted reading of LBC slides to current practice of manual reading of conventionally prepared Pap smear cytology samples. The reported ICER was high and outside acceptable cost effectiveness thresholds in the Australian context. However, the generalisability of this result to Scotland is unclear given key differences between the Scottish and Australian screening programmes\(^6\).

In phase II of the Scottish Feasibility study, an overall efficiency gain of 28% (p=0.04) was reported for image-assisted reading compared with manual reading, approximating to the processing of an additional four slides per hour per person\(^1\). There was wide variation among laboratories (range 7–80%)\(^1\).
References


