

***Computer-assisted
image analysis for
cervical screening***

May 2003

MSAC reference 12c

Assessment report

© Commonwealth of Australia 2003

ISBN 0 64282250 6

ISSN (print) 1443-7120

ISSN (online) 1443-7139

First printed September 2003

This work is copyright. Apart from any use as permitted under the *Copyright Act 1968* no part may be reproduced by any process without written permission from AusInfo. Requests and inquiries concerning reproduction and rights should be directed to the Manager, Legislative Services, AusInfo, GPO Box 1920, Canberra, ACT, 2601.

Electronic copies of the report can be obtained from the Medical Service Advisory Committee's Internet site at:

<http://www.msac.gov.au/>

Hard copies of the report can be obtained from:

The Secretary
Medical Services Advisory Committee
Department of Health and Ageing
Mail Drop Point 107
GPO Box 9848
Canberra ACT 2601

Enquiries about the content of the report should be directed to the above address.

The Medical Services Advisory Committee is an independent committee that has been established to provide advice to the Commonwealth Minister for Health and Ageing on the strength of evidence available on new and existing medical technologies and procedures in terms of their safety, effectiveness and cost-effectiveness. This advice will help to inform Government decisions about which medical services should attract funding under Medicare.

This report was prepared by the Medical Services Advisory Committee with the assistance of Dr Claire Harris, Dr Paul Fennessy and Associate Professor Jeremy Anderson (Centre for Clinical Effectiveness, Monash Institute of Health Services Research) and Ms Liliana Bulfone and Mr Anthony Harris (Health Economics Unit, Centre for Health Program Evaluation), Monash University. The report was endorsed by the Commonwealth Minister for Health and Ageing on 8 August 2003.

Publication approval number: 3250 (JN7563)

MSAC recommendations do not necessarily reflect the views of all individuals who participated in the MSAC evaluation.

Contents

Executive summary	iii
Introduction	1
Background	2
Computer-assisted image analysis for cervical screening.....	2
The procedure	2
Clinical need/burden of disease	5
Existing procedures and comparators	12
Marketing status of the device/technology.....	13
Current reimbursement arrangement.....	13
Approach to assessment	14
Review of literature.....	14
Expert advice.....	22
Results of assessment	23
Is it safe?.....	24
Is it effective?.....	24
What are the economic considerations?	31
Conclusions	32
Safety.....	32
Effectiveness.....	32
Cost-effectiveness	32
Recommendation	34
Appendix A MSAC's terms of reference and membership	35
Appendix B Supporting Committee	37
Appendix C HTA and other websites and databases searched	38
Appendix D Studies included in the review	40
Appendix E Studies excluded from the review	41
Abbreviations	45
References	46

Tables

Table 1	Number of low and high-grade intraepithelial abnormalities on histology for women aged 20-69 years, 1999	7
Table 2	Categories of reporting of cervical smears in Australia.....	9
Table 3	Electronic databases accessed for this review	14
Table 4	Medline search.....	15
Table 5	Criteria and definitions for assessing validity of included articles	17
Table 6	Levels of evidence for diagnostic tests	17
Table 7	Evidence dimensions.....	18
Table 8	Levels of evidence	18
Table 9	Validity criteria according to study design	19
Table 10	The generic relationship between results of the diagnostic test and disease status.....	20
Table 11	Appraisal of existing secondary studies (AHTAC 1998)	25
Table 12	Appraisal of existing secondary studies (Broadstock 2000).....	26
Table 13	Appraisal of secondary studies.....	28

Figures

Figure 1	AutoPap® Primary Screening System process.....	3
Figure 2	Current NHMRC guidelines for the management of women with screen detected abnormalities (NHMRC 1994)	11

Executive summary

The procedure

Conventional cytology involves examination of cervical smears in the laboratory by trained personnel using light microscopy. With recent advances in technology this manual system can now be augmented by computer-assisted image analysis. Computerised systems can be used in both the primary screening phase and the quality control re-screening process. There are two main systems available. One uses algorithm-driven software to classify cervical smears. The other is an interactive system based on neural network technology that assists the cytologist to classify the smear by highlighting areas of the slide most likely to contain an abnormality.

The Medical Services Advisory Committee – role and approach

The Medical Services Advisory Committee (MSAC) is a key element of a measure taken by the Commonwealth Government to strengthen the role of evidence in health financing decisions in Australia. The MSAC advises the Commonwealth Minister for Health and Ageing on the evidence relating to the safety, effectiveness and cost-effectiveness of new and existing medical technologies and procedures, and under what circumstances public funding should be supported.

A rigorous assessment of the available evidence is thus the basis of decision making when funding is sought under Medicare. A team from the Centre for Clinical Effectiveness, Monash Institute of Health Services Research, and the Health Economics Unit, Centre for Health Program Evaluation, at Monash University was engaged to conduct a systematic review of the literature on computer-assisted image analysis for cervical screening cytology. A supporting committee with expertise in this area then evaluated the evidence and provided advice to the MSAC.

MSAC's assessment of computer-assisted image analysis for cervical screening cytology

Clinical need

Cervical cancer is one of the most common cancers in women throughout the world. Although the incidence is decreasing, it remains the ninth most common form of cancer in Australian women (Commonwealth Department of Health and Aged Care 1998). Almost 1,000 new cases of cervical cancer are identified each year, with 269 women dying from the disease in 1998 (Australian Institute of Health and Welfare 2000).

For a number of reasons, cervical cancer is well suited to a screening program. The pre-cancerous state is readily detected prior to the patient developing symptoms; a suitable test exists that is acceptable to the target population of women; effective treatments are available and generally accessible. Screening plays a crucial role in reducing morbidity and mortality from this disease. Routine Pap smears taken every two years can prevent up to 90 per cent of squamous cervical cancers (Australian Institute of Health and Welfare 2000).

Safety

Computer-assisted image analysis is conducted in the laboratory on the same slides as prepared for conventional cervical cytology as it involves only slide reading and interpretation. The method for sample collection is identical to that used for conventional cytology. Slides are prepared in the usual way for the traditional Pap smear or for liquid-based cytology and then read either manually with a light microscope or using one of the computer-assisted systems.

Hence, the safety issues with this technology are the same as those for conventional cytological assessment methods.

Effectiveness

Assessment of the accuracy of diagnostic or screening tests requires that study subjects receive both the test under investigation and a reference test or 'gold standard'. The inclusion criteria for this review required that measures of test accuracy were based on a reference standard of cervical histology for high-grade lesions (CIN II and above) and appearance at colposcopy for lower grade lesions (CIN I and below) undertaken within six months of cytology. Long-term follow-up of patient outcomes was also acceptable.

Two systematic reviews of computer-assisted image analysis for cervical screening cytology have been undertaken previously: New Zealand Health Technology Assessment (NZHTA) Broadstock 2000, Australian Health Technology Advisory Committee (AHTAC) 1998. Both reviews found that there was insufficient evidence to support firm conclusions about the performance of these technologies and that estimates of test sensitivity and specificity could not be reliably determined.

The only study identified as meeting the inclusion/ exclusion criteria for the current review was Hartmann et al (2001). As with the earlier AHTAC (1998) and NZHTA (Broadstock 2000) reviews, this paper reported that the main shortcoming of relevant studies was the lack of an appropriate reference standard. Hartmann et al were unable to find any studies that met their systematic review protocol. They concluded that current evidence was inadequate to gauge whether new technologies were 'better' than conventional cytology.

Critical appraisal of this study revealed that the restricted search strategy used by Hartmann et al was a limitation of their methodology, potentially contributing to the lack of evidence identified. However the more comprehensive search strategy

employed in the present MSAC review also failed to reveal any studies that would have met their criteria.

The main reasons for exclusion of studies from the present review were: lack of an appropriate reference standard, populations that included both screening and high-risk patients, and circumstances that were not generalisable to the Australian setting (eg where the protocol for conventional cytology by light microscopy was not comparable to Australian practice).

In summary, there is a lack of evidence that computer-assisted image analysis is as effective as conventional manual screening for cervical screening cytology.

Cost effectiveness

Due to the lack of evidence of clinical effectiveness based on patient outcomes and the lack of information on test parameters such as sensitivity and specificity, an economic evaluation of computer-assisted image analysis for cervical screening could not be performed.

Recommendation

The MSAC recommends that as there is insufficient evidence to draw conclusions on the appropriate use of computer-assisted image analysis versus manual cytology screening, there are no grounds to change current funding arrangements.

The Minister for Health and Ageing accepted this recommendation on 8 August 2003.

Introduction

The Medical Services Advisory Committee (MSAC) has reviewed the use of computer-assisted image analysis, a diagnostic technology for cervical screening. The MSAC evaluates new and existing health technologies and procedures for which funding is sought under the Medicare Benefits Scheme in terms of their safety, effectiveness and cost-effectiveness, while taking into account other issues such as access and equity. The MSAC adopts an evidence-based approach to its assessments, based on reviews of the scientific literature and other information sources, including clinical expertise.

The MSAC's terms of reference and membership are at Appendix A. The MSAC is a multidisciplinary expert body, comprising members drawn from such disciplines as diagnostic imaging, pathology, surgery, internal medicine and general practice, clinical epidemiology, health economics, consumer health and health administration.

This report summarises the assessment of current evidence for computer-assisted image analysis for cervical screening.

Background

Computer-assisted image analysis for cervical screening

The procedure

Conventional cytology involves the examination of cervical smears in the laboratory by trained personnel using light microscopy. With recent advances in technology, this manual system can now be augmented by computer-based image analysis. Computerised systems can be used in two stages of the screening process, these being primary screening and re-screening.

Primary screening is the initial assessment conducted on all slides. Based on primary screening, slides considered to have no abnormality are defined as 'within normal limits' (WNL). All remaining slides are referred for further evaluation.

Re-screening is a quality control measure that can be done in a variety of ways. However, the various methods fall into the following two broad categories:

- All slides can be re-screened by either comprehensive or rapid review; or
- A sample of slides can be comprehensively reviewed. Samples can be selected at random from the whole slide population or by targeting slides designated as high-risk.

Computer-assisted image analysis can be used to perform either or both of these tasks. The three commercial products developed to date are AutoPap® Primary Screening System, PAPNET® Testing System and AutoCyte SCREEN. The latter two systems are no longer commercially available to new purchasers but are in current use in some Australian laboratories. The MSAC decided for this report, to review all technologies utilising computer-assisted image analysis.

AutoPap

The AutoPap® Primary Screening System has recently been renamed the Focalpoint™ slide profiler. However, as the term 'Focalpoint' does not appear in any of the published literature, 'AutoPap' will be used throughout this report. The following information was taken from literature provided by the manufacturer (Tripath Imaging 1999).

In the AutoPap System, cervical cytology slides are analysed by high-speed video microscope and image interpretation software. Software algorithms rank each slide based on the likelihood that an abnormality is present. The process entails computer evaluation of visual patterns and localisation of significant objects and groupings, followed by systematic comparison of these findings against thousands of reference images. The algorithms detect both squamous and glandular lesions and also assess the adequacy of the specimen.

The slides are ordered by probability of abnormality, and numbered one to 'n', where 'n' is the total number in the batch. In addition, the slides are allocated to one of five groups with Group 1 samples most likely to be abnormal and Group 5 the least likely to be abnormal. The percentage of the total batch of slides to be assigned to each grouping is set before screening.

The slides with the lowest likelihood of abnormality are designated 'No further review'. The manufacturer recommends that up to 25 per cent of slides can be assigned to this group which is then archived. The remaining slides that have been ranked and grouped according to the probability of abnormality are then reviewed by a cytologist.

AutoPap can also be used for quality control re-screening. The manufacturer recommends that 15 per cent of all successfully processed slides be reviewed by a cytologist. The selection of samples for this group is based on the highest probability of being abnormal (ie likely to be false negatives).

Earlier versions of AutoPap did not perform primary screening and were only used for quality control purposes to re-screen slides classified as WNL by the initial manual primary screen. The AutoPap® Primary Screening System process is outlined in Figure 1.

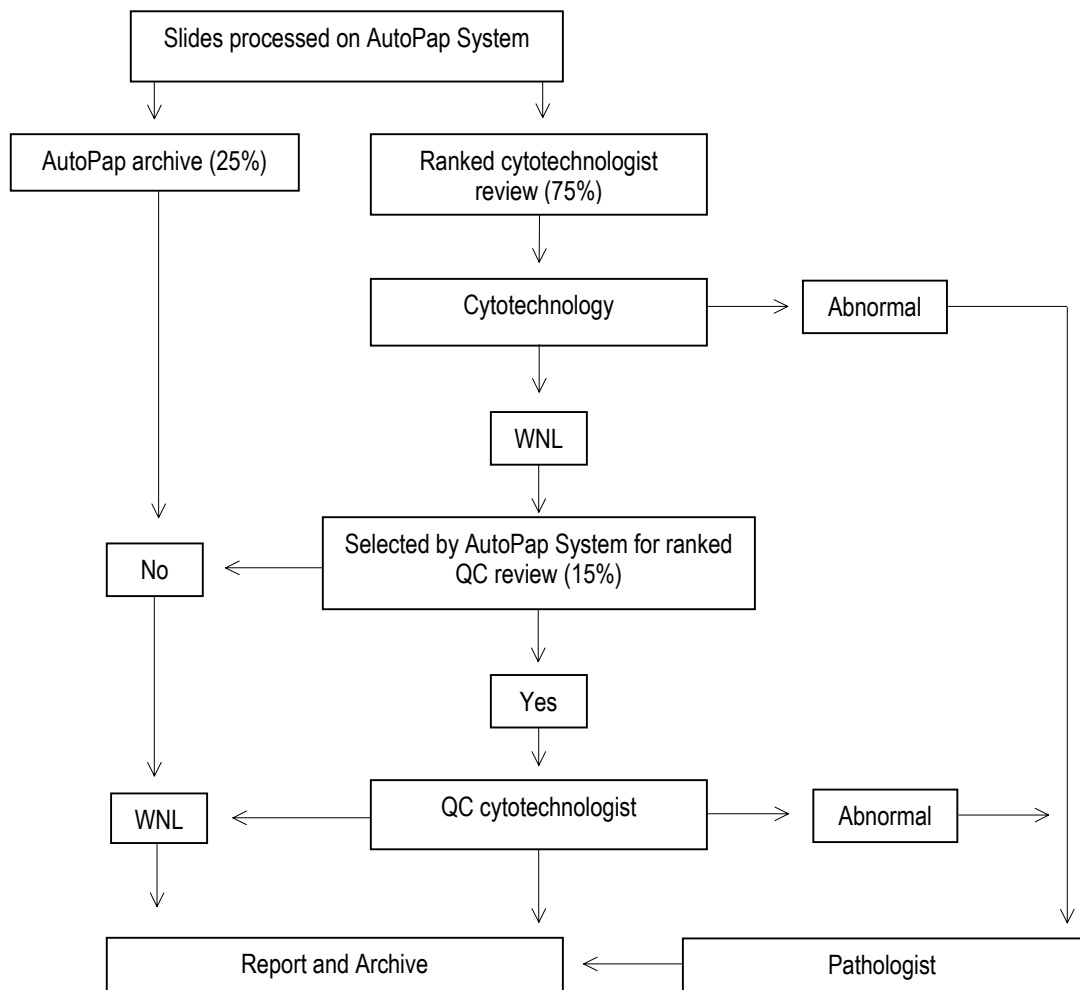


Figure 1 AutoPap® Primary Screening System process

AutoPap also evaluates the adequacy of Pap smear specimens. A satisfactory Pap smear should contain sufficient numbers and types of cells to indicate adequate sampling and the cells should not be obscured by blood, discharge or vaginal creams or other products. The criteria used by AutoPap to assess adequacy of the specimen include detection of both squamous and endocervical components, and estimation of the degree of any inflammation or obscuration present. Slides are classified as satisfactory, satisfactory but limited, or unsatisfactory.

The AutoPap® Primary Screening System is not indicated for the evaluation of slides identified as 'high risk'. Smears from patients with a past history of cervical disease who are currently symptomatic or have an abnormality on clinical examination should be referred directly for review by a cytologist. The manufacturer also acknowledges limitations in detection and/or evaluation of endometrial cells in post-menopausal women, reactive changes associated with radiation, atrophy with inflammation, and rare malignant neoplasms such as sarcomas or extra-uterine and metastatic carcinomas.

The intended users of the system are trained cytology laboratory personnel operating under the direct supervision of a qualified cytology supervisor or laboratory manager.

PAPNET

Unlike AutoPap, the PAPNET® Testing System does not diagnose abnormalities on slides. It is an interactive system that assists the cytologist to identify abnormal cells. Based on neural network technology, the program recognises suspicious cells and cell clusters and presents them to the cytologist for further assessment.

PAPNET consists of a scanning apparatus and a review station. Slides processed in the scanning apparatus are viewed through an automated microscope with an attached video camera. Selected images are assessed by a primary classifier and sent to the neural network computing unit which identifies the 128 images (64 cells and 64 cell clusters) most likely to be abnormal on each slide. Once processed, the colour images are transferred to digital tape. The scanning apparatus may be within the laboratory or at a remote location. For instance, some Australian laboratories send slides to Hong Kong while others have their own scanning apparatus. The review station is located in the local laboratory and the cytologist examines the recorded images on the high-resolution computer screen. If an abnormality is suspected, the original slide is reviewed manually (Koss et al 1997).

In the re-screening mode, PAPNET scans all slides classified in the manual primary screen as WNL. Hence, each slide designated as WNL is assessed twice by the cytologist, with the second examination confined to the areas of the slide highlighted by PAPNET. Review of these computer images by a cytologist determines if further manual examination of the slide is required. Although PAPNET has been approved by the Food and Drug Administration (FDA) in the USA for use in primary screening, it is no longer marketed commercially and information from the manufacturer was unavailable.

AutoCyte SCREEN

AutoCyte SCREEN is a computerised image analysis system for primary screening of slides prepared by the liquid-based cytology method, AutoCyte PREP. This differs from AutoPap and PAPNET which can be used with both conventional Pap smears and liquid-based preparations.

Like PAPNET, AutoCyte SCREEN selects the most abnormal cellular findings and presents them to the cytologist for further review. In addition, like AutoPap, the system makes an independent judgement on whether a slide is WNL. Slides for which there is disagreement in classification between automated and manual screening are subject to further cytological review (Bishop et al 2000).

Intended purpose

Cytology involves the detection and classification of abnormal cells. In Australia, less than 10 per cent of cervical smears contain an abnormality. In those that do, the abnormality may only be present in less than one per cent of cells on the slide. Detecting the visual contrast between normal and abnormal cells in large populations of predominantly normal slides is very demanding for human operators and could potentially be performed more effectively by machines.

Manual screening may be affected by operator fatigue after a prolonged period of assessing slides. To address this issue, US guidelines prevent cytologists from reviewing more than 100 slides per day and the more stringent Australian guidelines allow only 70 per day.

Operator fatigue, visual habituation and variability in screening techniques are sources of error in the manual screening process that could potentially be reduced by automation. Proponents of computer-assisted image analysis argue that automation will increase detection rates of abnormalities by improving accuracy.

There are additional arguments that computerisation can increase productivity due to 24-hour operational capability and decrease costs in terms of cytologists' time. However the significant establishment costs for these systems must also be considered.

Clinical need/burden of disease

Cervical cancer is one of the most common cancers among women throughout the world. Since cervical cancer has a pre-cancerous state that is readily detected and amenable to treatment, screening plays a crucial role in reducing morbidity and mortality from this disease.

The cervix is the neck of the uterus and the central cervical canal links the vagina with the uterine cavity. The cells lining the vagina are squamous cells that appear flat and lie in layers. Those lining the upper endocervical canal are columnar, named because they sit side by side to form a column.

The junction of the vagina and the upper endocervical canal where columnar cells undergo metaplasia to squamous cells is called the transformation zone. During the normal process of transformation, cells in this area are vulnerable to mutagenic agents such as chemicals or viruses that cause abnormal changes that sometimes lead to pre-cancerous or cancerous lesions.

Natural history

Squamous cell carcinoma is the most frequent type of cervical cancer, comprising about 70 per cent of currently reported cases (Australian Institute of Health and Welfare 2002). The other 30 per cent comprises adenocarcinomas, adenosquamous carcinomas and other rare types. Squamous cell cervical cancers almost always arise in the transformation zone of the cervix.

Pre-cancerous stage

Cervical cancer is a slow-growing lesion with a recognised pre-cancerous stage, cervical intraepithelial neoplasia (CIN), in which dysplastic changes are confined to the epithelium. CIN is graded from I to III according to the severity of the lesion, with CIN III being the most serious. The natural history of CIN is variable. It may progress to invasive cancer, persist without progression, or even regress (McIndoe et al 1984). Approximately one-third to one-half of cases of CIN I and CIN II will spontaneously regress (Channen 1990). In general, the more severe the abnormality, the less likelihood of regression, although women with CIN of any level have an increased risk of developing cervical cancer compared with women who have a normal Pap smear.

Symptomatic disease

The major symptoms of invasive cervical cancer include discomfort or bleeding during or after sexual intercourse, intermenstrual bleeding and vaginal discharge. Symptoms of more advanced disease include pelvic pain, blood in the urine, constipation, excessive tiredness, swollen legs and backache.

Role of Human Papillomavirus

The central role of human papillomavirus (HPV) in the pathogenesis of cervical cancer has been firmly established. Fourteen of the HPV subtypes are associated with cervical cancer and four of these (16, 18, 31 and 45) account for 80 per cent of all cases (Bosch et al 1995). One or more of the 14 high-risk subtypes are found in 99.7 per cent of cervical cancers worldwide (Bosch et al 1995; Cuzick et al 1999; Division of STD Prevention & Services, CDC 1999; Walboomers et al 1999). Women infected with high-risk HPV types have relative risks ranging from 12 to 350 for the development of high-grade cervical disease, with the higher risks pertaining to those with persistent HPV infection (Cuzick et al 1999, Liaw et al 1999, Nobbenhuis et al 1999). Up to 80 per cent of sexually active women are exposed to HPV at some point in their lives. At any one time 20 to 30 per cent of women between the ages of 18 and 30 are positive for HPV (Melkert et al 1993). For the vast majority of women, the HPV infection will clear naturally within 12 months and only the small proportion of women with persistent HPV infection are at increased risk of cervical cancer. Of women aged over 35, only 5 to 10 per

cent are persistent carriers of cancer-associated HPV types (Shah and Howley 1996).

Incidence and prevalence

One in 101 Australian women will develop cancer of the cervix in their lifetime. Although the incidence of cervical cancer is reducing in Australia, it still remains the ninth most common form of cancer in Australian women (Commonwealth Department of Health and Family Services 1998). In 1998, there were 868 new cases of cervical cancer in Australia with an age-standardised incidence rate of 4.4 cases per 100,000 amongst all women. The same year there were 269 deaths from cervical cancer, representing an age-standardised mortality rate of 2.5 cases per 100,000 women and 3,693 person-years life lost (Australian Institute of Health and Welfare & Australian Association of Cancer Registries 2000).

The prevalence of histologically verified high-grade epithelial abnormalities (HGEAs) in women screened in the target age group, ie 20-69 years, in Australia was 7 per 1,000 (Australian Institute of Health and Welfare 2000).

Table 1 Number of low and high-grade intraepithelial abnormalities on histology for women aged 20-69 years, 1999

Abnormalities	NSW	Vic	WA	SA	Tas	NT	Australia
LGEA ^a	6,207	4,197	2,563	1,767	640	158	15,753
HGEA ^b	4,523	3,546	1,509	1,237	470	179	11,642
Ratio	1.37	1.18	1.70	1.43	1.36	0.88	1.35

Source: <http://www.aihw.gov.au>. ^aLGEA=low-grade epithelial abnormalities; ^bHGEA=high-grade epithelial abnormalities

Although cervical cancer can affect women of all ages, it is more prevalent among women in their fifth or sixth decades, with a mean age of 54 years at diagnosis (Cannistra & Niloff 1996). In contrast, the intraepithelial lesions that are precursors of invasive cervical cancer more often occur in women under the age of 40.

The age-standardised incidence rate for cervical cancer declined by an average of 5.4 per cent annually between 1990 and 1999 (Australian Institute of Health and Welfare 2002). Mortality rates have fallen by an average of 5.5 per cent each year since 1990. These gains are predominantly due to the success of the National Cervical Screening Program.

Detection

The Pap smear is the currently used screening method to detect cervical abnormalities in Australia. The test is undertaken by a doctor or nurse and involves insertion of a speculum into the vagina to allow visualisation of the cervix. A specimen is then collected by gently scraping the surface and outer canal. Cells collected in this way are transferred to a microscope slide that is then forwarded to a pathology laboratory for assessment.

Screening for cervical cancer by routine Pap smears taken every two years can prevent up to 90 per cent of squamous cervical cancers (Australian Institute of

Health and Welfare 1998). The National Cervical Screening Program (NCSP) coordinates these activities in Australia (Australian Department of Health and Ageing 2002).

The current recommendations for the NCSP screening program are:

- Routine screening with Pap smears should be carried out every two years for women who have no symptoms or history suggestive of cervical pathology.
- All women who have ever been sexually active should commence having Pap smears between the ages of 18 to 20 years, or one or two years after first sexual intercourse, whichever is later. In some cases, it may be appropriate to start screening before 18 years of age.
- Pap smears may cease at the age of 70 years for women who have had two normal Pap smears within the last five years. Women over 70 years who have never had a Pap smear, or who request a Pap smear, should be screened.
- This policy applies only to women without symptoms which could be due to cervical pathology, past history of high-grade cervical lesions, or who are being followed up for a previous abnormal smear.

Australian laboratories performing cervical cytology apply a standardised reporting terminology and are assessed against a set of performance measures as part of the formal accreditation process undertaken by the Australian Government.

Cervical Cytology Registries which exist in each State and Territory play a key role in quality assurance by providing reminders to women for routine screening and follow-up of abnormal Pap smears. They also provide clinical information and performance data to medical practitioners and pathology laboratories.

The Australian NCSP has been very successful, reducing mortality from cervical cancer by 40 per cent in the past 10 years. However there remains a population of Australian women at risk who are not adequately screened. One study of Australian women with invasive cancer found that at least a third had never been screened (Wain et al 1992) and a recent Victorian Cervical Cytology Registry report provided evidence that only one quarter of the women with micro-invasive cancer had been adequately screened (Mitchell et al 2002). The need to improve recruitment strategies for unscreened and under-screened women is one of the key findings of an evaluation of the NCSP (Australian Department of Health and Ageing 2002).

Classification

The Australian system of classification for cervical cytology separates the lesions into two main groups, low- and high-grade abnormalities (LGEA and HGEA), with detailed subcategories (NHMRC 1994). The US-based Bethesda method has a different basis for classification that includes low- and high-grade squamous

intraepithelial lesion (LSIL, HSIL), atypical squamous cells of undetermined significance (ASCUS), atypical glandular cells of undetermined significance (AGUS) and within normal limits (WNL). The Australian classification system is outlined in Table 2.

Table 2 Categories of reporting of cervical smears in Australia

<p>High-grade epithelial abnormalities (HGEA)</p> <p>1. Intraepithelial lesions (HSIL*)</p> <ul style="list-style-type: none"> ▪ Squamous cell changes indicating moderate dysplasia (CIN II) and severe dysplasia or carcinoma in situ (CIN III) ▪ Squamous cell changes of severe dysplasia or carcinoma in situ (CIN III) with features of possible invasion ▪ Glandular cell changes indicating significant columnar cell dysplasia and adenocarcinoma in situ ▪ Mixed intraepithelial lesions with squamous and glandular components <p>2. Invasive lesions</p> <ul style="list-style-type: none"> ▪ Squamous cell carcinoma ▪ Adenocarcinoma ▪ Undifferentiated carcinoma of small-cell or large-cell type ▪ Mixed carcinoma <p>High-grade, non-epithelial abnormalities</p> <p>Sarcomas, lymphomas, malignant melanoma, etc.</p> <p>Inconclusive</p> <p>Abnormal cells suggesting the possibility of a high-grade abnormality, but in which a confident cytological diagnosis is not possible (ASCUS*)</p> <p>Low-grade epithelial abnormalities (LGEA)</p> <p>1. Squamous cell changes</p> <ul style="list-style-type: none"> ▪ Non-specific minor changes with some features that may be associated with HPV or minimal dysplasia but lacking stringent criteria of HPV effect or dysplasia (ASCUS*) ▪ Smears showing stringent criteria of HPV effect (LSIL*) ▪ Mild dysplasia (CIN I) with or without criteria of associated HPV effect (LSIL*) <p>2. Glandular cell changes</p> <ul style="list-style-type: none"> ▪ Minor changes in endocervical glandular cells including changes attributable to HPV effect (AGUS*) <p>Negative – no cytological evidence of dysplasia or malignancy</p> <p>Those smears in which no abnormal cells are detected</p> <p>Those smears with changes which are readily attributable to reactive processes (WNL*)</p> <p>Technically unsatisfactory</p> <p>Those smears which cannot be assessed at all due to the paucity of the sample, or a covering of blood or inflammatory exudate, or poor fixation</p>
--

*Bethesda Classification System. Source: NHMRC (1994).

Management

For a screening program to be effective, it is imperative that there is an effective and acceptable treatment for the condition and appropriate services where treatment is readily accessible.

There is a range of treatments for pre-cancerous changes including cryosurgery, cauterisation, laser surgery or loop or cone biopsies. In a small number of cases a hysterectomy may be necessary. Treatment for invasive cancer usually involves

extensive surgery. The current NHMRC guidelines for the management of cervical abnormalities are outlined in Figure 2.

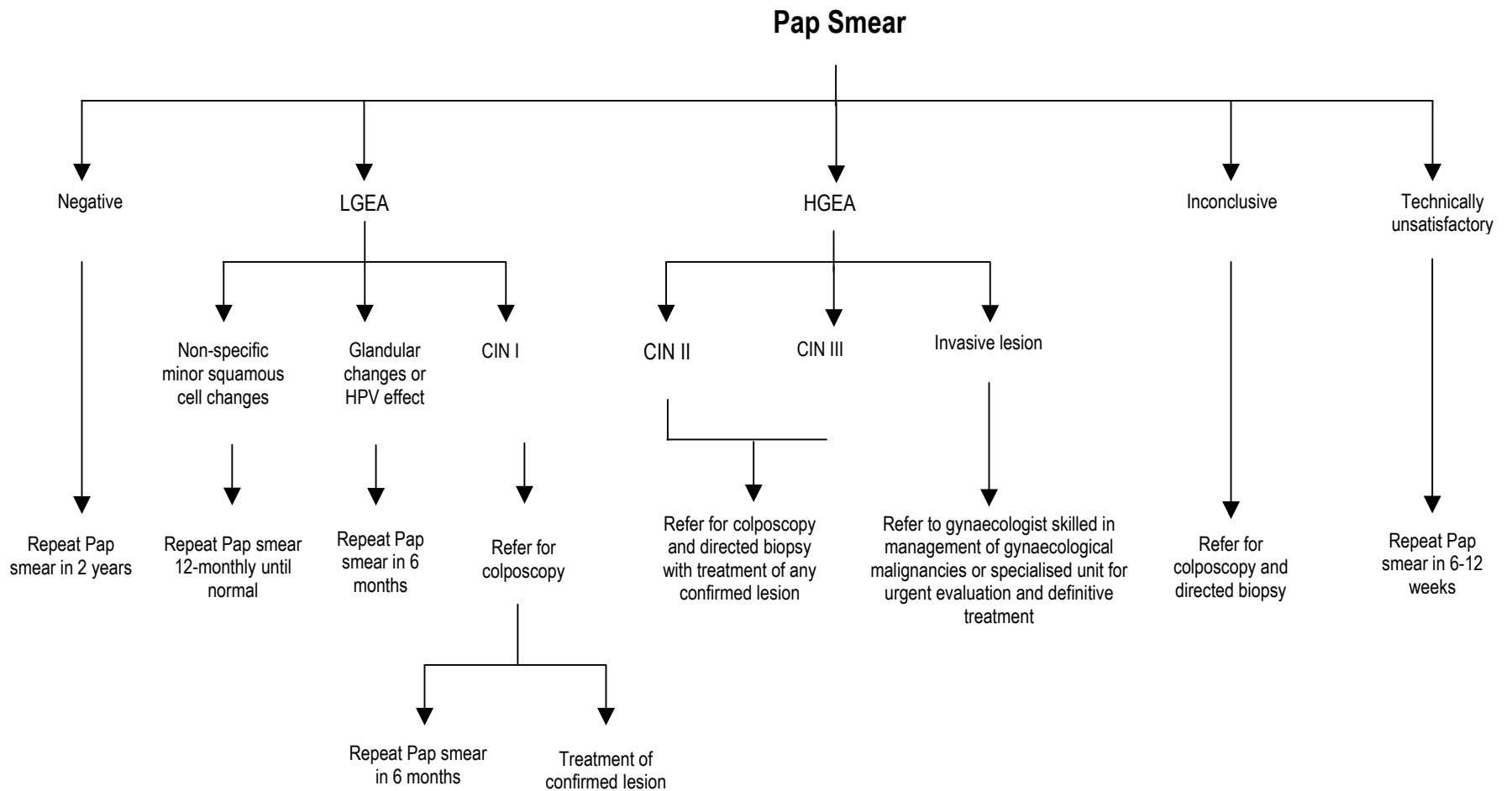


Figure 2 Current NHMRC guidelines for the management of women with screen detected abnormalities (NHMRC 1994)

Existing procedures and comparators

The comparator for computer-assisted image analysis is conventional manual screening in which trained laboratory personnel examine the slides using light microscopy.

Primary screening

Primary screening is conducted on all slides. A cytologist reviews all slides. Those considered to have no abnormality are defined as WNL. All other slides are referred for detailed examination.

Re-screening

As indicated above, re-screening is a quality control measure that can be done in a variety of ways:

- Re-screening all slides by either comprehensive or rapid review; or
- Comprehensive review of a sample of slides. Samples can be selected at random from the whole slide population or by targeting slides designated as high-risk.

In Australia, the usual practice for re-screening involves the following two components.

- All slides with a clinical indication (ie the patient has symptoms or signs suggestive of abnormality) undergo full comprehensive review; and
- All remaining WNL slides undergo rapid review.

In other centres, particularly in the USA where it is mandated practice, a 10 per cent sample of WNL slides is selected to be re-screened. This is not favoured in Australia as it is recognised that only 10 per cent of any missed abnormalities can be detected this way. Rapid review involves sampling the cells on each slide, rather than sampling the slides themselves. Even if rapid review is only 50 per cent effective it will identify 50 per cent of the missed abnormalities, whereas random sampling of 10 per cent of the slides will only ever identify a maximum of 10 per cent of missed abnormalities.

Further review

Those slides designated for further review are examined again. Those in which an abnormality is detected are referred to a pathologist for definitive diagnosis.

Marketing status of the device/technology

AutoCyte SCREEN and PAPNET are no longer commercially available but are currently used in some Australian laboratories. AutoPap is now marketed under the name Focalpoint.

These devices fall into the category of in vitro diagnostics. Diagnostic instruments that do not interact directly with the patient are exempt from listing in the Australian Register of Therapeutic Goods.

The Australian National Pathology Accreditation Advisory Council has issued 'Guidelines for the use of fluid based collection systems and automated and semi-automated screening devices in the practice of gynaecological (cervical) cytology' (NPAAC 2001) for use in laboratory accreditation.

Current reimbursement arrangement

There is currently no listing in the Medicare Benefits Schedule (MBS) specifically relating to computer-assisted image analysis.

Approach to assessment

Review of literature

Search strategy

Reviews of this topic have previously been undertaken by the Australian Health Technology Advisory Committee (AHTAC 1998) and the New Zealand Health Technology Assessment group (Broadstock 2000). The MSAC Supporting Committee decided that the current review would update the findings of these reports rather than duplicating previous work. All three devices were reviewed by AHTAC in 1997 but only AutoPap was evaluated in the Broadstock 2000 report.

To update the existing reviews, the medical literature was searched to identify relevant studies for AutoCyte SCREEN and PAPNET from 1997, and for AutoPap from 2000. Table 3 presents the electronic databases that were used to provide a list of citations.

Table 3 Electronic databases accessed for this review

Database	Period covered
Cochrane Library including: The Cochrane Database of Systematic Reviews (CDSR) Database of Abstracts of Reviews of Effectiveness (DARE) The Cochrane Controlled Trials Register (CCTR) Health Technology Database (HTA) NHS Economic Evaluation Database (NHS EED)	2002, Issue 3
Medline	1996 to Sept Week 2002
PreMedline	Sept 18 2002
Current Contents	1993 Week 27 to 2002 Week 35
Biological Abstracts	1980 to Sept 2002
CINAHL	1982 to Sept Week 1 2002
EBM Reviews: Cochrane, ACP Journal Club, CCTR and DARE	4th Quarter 2002
CancerLit	September 2002

Additionally, an Internet search of health technology assessment (HTA) databases and HTA agency websites was undertaken. The Internet sites that were searched are listed in Appendix C.

The search statements used to identify relevant literature in Medline are presented in Table 4. The search was repeated on the additional databases after adaptation, eg to the appropriate syntax, MeSH headings.

Table 4 Medline search

1. autopap.tw.	18. vaginal smears/
2. papnet.tw.	19. vaginal smear\$.tw.
3. autocyte cscreen.tw	20. (cervi\$ adj (screen\$ or smear\$)).tw.
4. focalpoint.tw.	21. or/17-20
5. leytas.tw.	22. mass screening/
6. automated cytol\$.tw.	23. cytol\$ screen\$.tw.
7. automated screen\$.tw.	24. cytopath\$.tw.
8. autoanalysis/	25. cytodiagnosis/
9. automation/	26. cytological techniques/
10. cytophotometry/	27. or/22-26
11. image cytometry/	28. cervix Uteri/
12. diagnostic imaging/	29. cervi\$.tw.
13. diagnosis, computer-assisted/	30. 28 or 29
14. image processing, computer-assisted/	31. 27 and 30
15. "neural networks (computer)"/	32. 21 or 31
16. or/1-15	33. 16 and 32
17. ((pap or pap\$) adj (smear\$ or test\$)).tw.	34. limit 33 to (human & female & yr=1997-2002)

Selection criteria

The following criteria were developed *a priori* to determine eligibility of relevant studies:

Subject characteristics

Inclusion: Women undergoing cervical screening.

Exclusion: Not specified.

Characteristics of the test/intervention

Inclusion: Cervical screening cytology using computer-assisted image analysis as an adjunct to, or replacement for, primary manual screening.

Exclusion: Not specified.

Characteristics of the comparison test/intervention

Inclusion: Cervical cytology screening by manual reading alone.

Exclusion: Not specified.

Characteristics of the outcome (screening test)

Inclusion: Measures of diagnostic accuracy based on cervical histology for high-grade lesions (CIN II and above) and, as a minimum, appearance at colposcopy for lower grade lesions (CIN I and below) that was undertaken within six months of cytology.

Exclusion: Not specified.

Characteristics of the outcome (screening program)

Inclusion: Clinically relevant patient outcomes, safety and cost-effectiveness.

Exclusion: Not specified.

Characteristics of the study design (screening test)

Inclusion: Health technology assessments, systematic reviews, meta-analyses and cross-sectional prospective comparative studies. If these were unavailable, other comparative studies, cohort studies and case series are evaluated.

Exclusion: Case series in less than 10 patients, case reports, narrative reviews, editorials and letters.

Characteristics of the publication (date, language, specific journals)

Inclusion: Studies published after July 1997 (the final search date of the 1998 AHTAC review).

Exclusion: Languages other than English.

Assessment of validity

Part 1: Screening test

The most rigorous study design for assessing the diagnostic accuracy of a screening test is considered to be a prospective blind comparison of the test against a reference or 'gold' standard in a consecutive series of patients from a relevant clinical population (Jaeschke et al 1994, Sackett et al 2000). The Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests (1996) expanded on this definition and recommended six criteria for assessing the validity of evidence. Based on these criteria, a checklist was developed (Table 5).

Table 5 Criteria and definitions for assessing validity of included articles

Validity criteria	Definition
Test is compared with a reference (gold) standard	Patients in the study should have undergone both the test in question and a reference test that would provide confirmatory proof that they do or do not have the target disorder.
Appropriate spectrum of consecutive patients	Study included patients for whom the test would normally be used in clinical practice, ie patients covering the spectrum of mild to severe cases of the target disorder, early and late cases, and patients with other, commonly confused diagnoses. An inappropriate spectrum compares patients already known to have the disorder with a group of normal non-diseased patients (case-control) or with patients diagnosed with another condition.
Masked assessment of study and reference tests results	The study test and the reference test should be interpreted separately by persons unaware of the results of the other (avoidance of review bias).
All study subjects tested with both study and reference tests	The reference test should be applied regardless of a positive or negative result from the study test (avoidance of work-up/verification bias).
Study test measured independently of clinical information	The person interpreting the test should be masked to clinical history and results of any other tests performed previously.
Reference test measured prior to any interventions	No treatment interventions should be initiated prior to the application of the reference (or study) test.

Included studies are also classified according to a hierarchy of evidence. At present the NHMRC does not have a system for assigning a hierarchy of evidence to studies of screening and diagnostic tests. The system developed by the Centre for Evidence Based Medicine (2001) has been adapted for use here (Table 6). The levels of evidence reflect the methodological rigour of the studies. A study assigned as Level I evidence is considered the most rigorous and least susceptible to bias while a study deemed to contain Level IV evidence is considered the least rigorous and most susceptible to bias.

Table 6 Levels of evidence for diagnostic tests

Level of Evidence	Criteria
I	Independent blind comparison of an appropriate spectrum* of consecutive patients, all of whom have undergone both the study test and the reference standard.
II	Independent, blind or objective comparison but in a set of non-consecutive patients, or confined to a narrow spectrum of study individuals (or both), all of whom have undergone both the study test and the reference standard.
III	Independent blind comparison of an appropriate spectrum, but the reference standard was not applied to all study patients.
IV	Either of: Reference standard was not applied blinded or not applied independently, or No reference test applied (case series).

* An appropriate spectrum is a cohort of patients who would normally be tested for the target disorder. An inappropriate spectrum compares patients already known to have the disease with patients diagnosed with another condition, or with a separate group of normal patients (case-control).

Part 2: Screening program

The most rigorous study design for assessing the validity of a screening test implemented within a screening program is a randomised controlled trial (RCT) (Guyatt et al 1993, Sackett et al 2000). The evidence is assessed and classified using the dimensions defined by the NHMRC (2000). These dimensions (Table 7) include three domains: strength of the evidence, size of the effect and relevance of the evidence. The first domain is derived directly from the identified studies. The last two require expert clinical input as part of the determination.

Table 7 Evidence dimensions

Type of evidence	Definition
Strength of the evidence	The study design used, as an indicator of the degree to which bias has been eliminated by design*. The methods used by investigators to minimise bias within a study design. The <i>p</i> -value or, alternatively, the precision of the estimate of the effect. It reflects the degree of certainty about the existence of a true effect.
Level	
Quality	
Statistical precision	
Size of effect	The distance of the study estimate from the 'null' value and the inclusion of only clinically-important effects in the confidence interval
Relevance of evidence	The usefulness of the evidence in clinical practice, particularly the appropriateness of the outcome measures used.

*see Table 8

The three sub-domains (level, quality and statistical precision) collectively measure the strength of the evidence. The designations of the levels of evidence are shown in Table 8.

Table 8 Levels of evidence

Level of evidence	Study design
I	Evidence obtained from a systematic review of all relevant randomised controlled trials
II	Evidence obtained from at least one properly-designed randomised controlled trial
III-1	Evidence obtained from well-designed pseudo-randomised controlled trials (alternate allocation or some other method)
III-2	Evidence obtained from comparative studies (including systematic reviews of such studies) with concurrent controls and allocation not randomised, cohort studies, case-control studies, or interrupted time series with a control group
III-3	Evidence obtained from comparative studies with historical control, two or more single arm studies, or interrupted time series without a parallel control group
IV	Evidence obtained from case series, either post-test or pre-test/post-test

Modified from NHMRC 2000

All accepted articles are assessed for study validity (Table 9) based on criteria that focus on important aspects of study design (Chalmers and Altman 1995 and Sackett et al 2000 for systematic reviews; Schulz et al 1995 and Jadad et al 1996 for RCTs; and NHS Centre for Reviews and Dissemination 2001 for other study designs).

Table 9 Validity criteria according to study design

Study design		Validity criteria
Primary studies*	RCT	Randomised method Allocation concealment Similar groups at baseline Specified eligibility criteria Blinding of patients, investigators and outcome assessors Proportion lost to follow-up Point estimates and measure of variability presented for the primary outcome measure Intention to treat analysis
	Cohort	Prospective/retrospective Comparable groups at inception Intervention/treatment reliably ascertained Identification and adjustment for confounding factors Blind outcome assessment Sufficient duration of follow-up Proportion lost to follow-up
	Case-control	Explicit definition of cases Adequate details of selection of controls Comparable groups with respect to confounding factors Interventions and other exposures assessed in same way for cases and controls Possibility of over-matching ie cases and controls matched according to factors related to exposure Appropriate statistical analysis
	Case series	Explicit description of patients Explicit inclusion/exclusion criteria All patients included Sufficient follow-up Outcomes assessed objectively Explicit description of techniques
Secondary studies#	Systematic reviews	Focused research question Explicit inclusion/exclusion criteria Explicit and comprehensive search strategy Validity of included trials appraised Homogeneity between studies assessed Summary of main results Strengths and limitations

*Primary study criteria modified from NHS Centre for Reviews and Dissemination (2001); #Secondary study criteria modified from Evidence Based Medicine Toolkit, University of Alberta (<http://www.med.ualberta.ca/ebm/ebm.htm>, Accessed August 2002)

Data extraction

Data were extracted from included studies using standardised instruments created for the assessment. Two reviewers examined each article and any discrepancies in evaluation were discussed and resolved through consensus.

Part 1: Screening test

The accuracy of a screening test is primarily determined by its ability to identify the target disorder compared to the recognised 'gold standard' test. The gold standard is used as a proxy for true disease status ie individuals who test positive using the gold standard are assumed to have the disease and those who test negative are assumed to be disease-free. This can be summarised in two-by-two tables as in Table 10. Individuals who test positive for the disease in both the study test and the reference test are represented in cell 'a' and are called true positives (TP). Individuals without the disease who test negative in both tests (the 'd' cell) are called true negatives (TN).

There may be discordance between the test result and the true disease status of the subject. When this occurs a false result is reported. In cell 'b', the test is positive in individuals without the disease; in cell 'c', the test is negative in diseased individuals. These two sets of false results are called false positives (FP) and false negatives (FN), respectively.

Table 10 The generic relationship between results of the diagnostic test and disease status

Study Test Results	True Disease Status (Reference test)		Total
	Diseased	Not Diseased	
Positive	a	b	a+b
Negative	c	d	c+d
Total	a+c	b+d	a+b+c+d

Abbreviations: a=number of diseased individuals detected by the test; b=number of individuals without disease detected by the test; c=number of diseased individuals not detected by the test; d=number of individuals without disease not detected by the test; a+b=total number of individuals testing positive; c+d=total number of individuals testing negative; a+c=total number of diseased individuals; b+d=total number of individuals without disease; a+b+c+d=total number of individuals studied.

The accuracy of a screening test is assessed using the concepts of sensitivity, specificity and likelihood ratio.

Sensitivity is the proportion of diseased individuals who test positive. It is a measure of the probability of correctly diagnosing a case or the probability that any given case will be identified by the test. Referring to Table 10, sensitivity is calculated as:

$$Sen = \frac{a}{a+c} = \frac{TP}{TP+FN}$$

Specificity is the proportion of individuals without disease who test negative. It is the probability of correctly identifying a non-diseased person with the study test and is calculated as:

$$Spe = \frac{d}{b + d} = \frac{TN}{TN + FP}$$

The complement of specificity is the false positive rate (FPR), calculated as:

$$FPR = 1 - Spe$$

Likelihood ratios (LR) indicate by how much a given test result will raise or lower the pre-test probability of the target disorder. They express the odds that a given level of a test result would be expected in a patient with the condition compared to one without the condition. The likelihood ratio for a positive test result ($LR+$) is related to sensitivity and the false positive rate and is calculated as:

$$LR+ = \frac{Sen}{FPR}$$

The likelihood ratio for a negative test result ($LR-$) is calculated as:

$$LR- = \frac{1 - Sen}{Spe}$$

Note that large positive likelihood ratios of 10 or more and small negative likelihood ratios less than 0.1 indicate large changes in disease likelihood. If the likelihood ratio for a positive test result is less than two and the likelihood ratio for a negative test result is greater than 0.5, then there is little or no change in diagnosing presence of disease after taking the test.

Part 2: Screening program

The ability of a test to detect pathology accurately is not in itself a good indicator of its usefulness in a screening program. Application of the test should also enhance patient outcomes. The ideal method for assessing enhanced outcomes is an RCT that compares outcomes between patients who have had the test and those who have not had the test. This comparison should demonstrate that early detection and treatment (due to the test) produce better patient outcomes than delaying treatment until the disease becomes symptomatic and would be recognised without screening. At the very least, the study should discuss patient management options based on the test result.

Expert advice

A supporting committee with expertise in cervical cytology was established to evaluate the evidence and provide advice to the MSAC from a clinical perspective. In selecting members for supporting committees, the MSAC's practice is to approach the appropriate medical colleges, specialist societies and associations and consumer bodies for nominees. Membership of the supporting committee is provided at Appendix B.

Limitations of this review

This review was conducted using the systematic review methodology outlined above. However, due to the time frame available, the search was limited to English language articles only. Members of the supporting committee were available to provide advice on current work or publications not accessible via electronic search methods, however no attempt was made to systematically search the 'grey literature'.

Results of assessment

Search results

An initial assessment of the abstracts allowed for the exclusion of articles that clearly failed to meet the selection criteria. Ambiguous or unclear citations were included in the next assessment stage in which the full text was examined. Two reviewers examined the articles for inclusion at each stage. Discrepancies in selection were discussed and resolved through consensus. A final decision to include or exclude a study was based on a thorough reading of the complete article. All studies included in this report passed through each stage of this process.

An initial search for studies on computer-assisted image analysis for cervical screening identified 325 articles of which 270 were rejected, leaving 55 articles for assessment of full text. Of these, only one met the inclusion criteria. Appendix E lists articles that were excluded after assessment of the full text together with the reason for exclusion.

There were three main reasons for excluding papers from the present review. Most of the excluded studies did not use the appropriate reference standard of histology or colposcopy. Instead, they generally used consensus of cytological reporting as the reference standard. Without an independent measure representing the true condition of the cervix, sensitivity, specificity and positive and negative predictive values cannot be calculated. Some studies had histological data available for a subgroup of patients, but the presentation of results did not allow separate interpretation of these data. The inclusion criteria for the current review required that measures of diagnostic accuracy were based on cervical histology for high-grade lesions (CIN II and above) and appearance at colposcopy for lower grade lesions (CIN I and below) that was undertaken within six months of cytology.

The second main reason for exclusion was that many studies contained in their test population, a mixture of both screening and referred patients. Referred patients are those being investigated as a follow-up to a previous abnormality or because they currently have symptoms or clinical signs. These women are at higher risk of having a cervical abnormality than the group presenting for screening. The increased prevalence of abnormality means that the sensitivity of the test in the referred group is different to that in asymptomatic women presenting for screening. One of the inclusion criteria for this review was that the population should be in women of average risk presenting for cervical screening.

The third main reason for excluding studies was that the circumstances were not generalisable to the Australian setting and involved mainly high-risk populations (eg those with a known high prevalence of disease) or studies where the protocol for conventional cytology by light microscopy was not comparable to Australian practice.

Several studies were excluded for more than one of these three main reasons.

Is it safe?

Computer-assisted image analysis is conducted in the laboratory on the same slides as prepared for conventional cervical cytology as it involves only slide reading and interpretation. The sample collection method is identical to that used for conventional cytology. Slides are prepared in the usual way for the traditional Pap smear or for liquid-based cytology and then read either manually with a light microscope or using one of the computer-assisted systems.

Hence, the safety issues with this technology are the same as those for the conventional cytological assessment methods in current use.

Is it effective?

Findings of existing secondary studies

Reviews of this topic have previously been undertaken by the Australian Health Technology Advisory Committee (AHTAC 1998) and the New Zealand Health Technology Assessment group (Broadstock 2000). It was decided that this review would update the findings of these reports rather than duplicating the previous work. The AHTAC and Broadstock reports were critically appraised and found to be both of high quality and applicable to the current research question. Details of the critical appraisals are outlined in Tables 11 and 12.

Both systematic reviews found that there was insufficient evidence to support firm conclusions about the performance of these technologies. This was because estimates of test sensitivity and specificity could not be reliably determined.

AutoCyte SCREEN was reviewed in the AHTAC report only. Peer-reviewed literature was scarce and clinical trial results were not available at the time.

PAPNET was also reviewed by the AHTAC only, resulting in a small positive finding. Prospective biopsy-confirmed studies indicated that review by PAPNET after conventional screening may increase the detection of biopsy-proven low-grade changes by between zero and seven per cent and high-grade changes by between three and six per cent. Most additional abnormalities detected were atypical or low-grade squamous intraepithelial lesions (SIL), the clinical significance of which is difficult to interpret.

AutoPap was reviewed in both reports. The results suggested a possible improvement in detection of low-grade abnormalities relative to 10 per cent random re-screening. However there was no evidence to suggest an increase in detection of high-grade abnormalities.

The implications of these incremental benefits in the Australian setting are not clear. While laboratories with lower accuracy rates would benefit most from these technologies, Australian standards are already high compared to many overseas services. In addition, Australian laboratories do not use 10 per cent random re-screening as it is recognised to be less efficient in detecting false negatives.

Table 11 Appraisal of existing secondary studies (AHTAC 1998)

Reference	Australian Health Technology Advisory Council, 1998. 'Review of automated and semi-automated cervical screening devices', Commonwealth Department of Health and Family Services, Canberra.
Is there a focused research question? (ie PICO elements)	Not explicitly stated. Patients: Women (screening). Intervention/Diagnostic test: Image analysis devices. Comparison: Conventional cytology. Outcomes: Number of additional abnormalities detected, number of additional cancers or pre-cancers detected, likely number of extra cancers detected or prevented, cost per additional cancer detected, cost per life-year gained, cost benefit of automated devices. Reference Standard: Unspecified.
Are inclusion and exclusion criteria, including <i>a priori</i> , for selected studies stated?	Experimental studies with appropriate purpose, nature and comparators (details not specified).
Is there an explicit and comprehensive search strategy?	Search terms included 'automated cytology', 'AutoPap', 'PAPNET', 'Pap smear screening', 'cervical screening', 'cytology screening', 'cytopathology', 'automated screening' and 'vaginal instrumentation'. Databases searched included Medline and CancerLit. Internet websites were also searched. Restricted to articles published between 1990 and July 1997.
Are the included trials appraised for validity? Are validity criteria stated?	Studies were assessed according to design criteria, study type, sample size and outcomes reported. Factors included prospective versus retrospective, tests applied to same group of people and same slides, tests compared with reference diagnosis, definition of reference diagnosis, blinded assessment of slides, random sampling of slides, generalisability of screening conditions, definition of abnormality, reporting of sensitivity and specificity, statistical analysis and whether primary screening or re-screening. Studies were rated according to NHMRC (1995) quality guidelines.
Are results consistent from study to study? Is homogeneity assessed?	No high level studies were identified and no statistical combination of results was undertaken. Results were discussed qualitatively.
Summary of main results	AutoPap: There were few peer-reviewed clinical studies of AutoPap found for evaluation. Results suggest an improvement in detection of abnormalities relative to 10% random re-screening. PAPNET: Most studies of PAPNET reviewed were retrospective studies, involving re-screening of archival smears. Some reports of sensitivity and specificity in the literature are limited by comparison not being made with the gold standard of biopsy confirmation. In many studies which examined re-screening of false negative results, the reviewers were not blinded to the outcome. Prospective biopsy-confirmed studies indicated that review by PAPNET after conventional screening may increase the detection of biopsy-proven low-grade changes by between 0% and 7% and high-grade changes by between 3% and 6%. Most additional abnormalities detected were atypical or low-grade SIL, the clinical significance of which is difficult to interpret. AutoCyte SCREEN: There is little peer-reviewed literature on this device. Clinical trial results are not yet available.
Strengths	Although some of the details of the review process were not clearly specified, it appears that the process was undertaken using an evidence-based approach with systematic critical appraisal of retrieved studies based on NHMRC guidelines. There was little high quality material to work with so the findings were essentially qualitative discussions of the available literature.
Limitations	The inclusion/exclusion criteria were not clearly specified. From the text, it appeared that the process was more likely to include rather than exclude studies so this may not be a significant limitation. The search strategy limited to Medline and CancerLit may have missed information in journals not listed in these databases.

Table 12 Appraisal of existing secondary studies (Broadstock 2000)

Reference	Broadstock M., 2000. 'Effectiveness and cost-effectiveness of automated and semi-automated cervical screening devices. A systematic review of the literature', New Zealand Health Technology Assessment, Report 2000; 3(1), Christchurch.
Is there a focused research question? (ie PICO elements)	Yes, explicitly stated in the review. Patients: All women (screening). Intervention/Diagnostic test: Semi-automated imaging devices for primary screening and re-screening. Comparison: Conventional cytology. Outcomes: Sensitivity, specificity, relative true and false positive rates, positive predictive value at a threshold of HSIL+ for both cytology and reference standard. Reference standard: Histology (biopsy), cytology by expert panel review.
Are inclusion and exclusion criteria, including <i>a priori</i> , for selected studies stated?	Included studies comparing the clinical or cost effectiveness of semi-automated imaging devices. Used either histology or cytology by adjudicated panel review as the reference standard. Excluded non-English studies, correspondence, abstracts, single case studies and those with poor description of methods and results. Studies evaluating slides taken from samples were excluded (such as women with previous abnormal samples, never been screened, pregnant, or those less than 18 years).
Is there an explicit and comprehensive search strategy?	Search terms included MeSH headings: 'cervix neoplasms', 'vaginal smears', 'mass screening', 'cytological techniques', 'image processing-computer assisted', 'cervix neoplasms-diagnosis' and Key words: 'Papanicolaou', 'autopap', 'Papnet', 'autocyte screen', 'rescreen*', '((vagina* or cervi*) near (screen* or smear*))', 'cervi* near cytology', 'pap* near (screen* or smear*)'. Databases searched included Medline, EMBASE, HealthSTAR, Current Contents, Science Citation Index, CancerLit, EconLit, Cochrane Library, Database of Abstracts of Reviews of Effectiveness, NHS Economic Evaluation and HTA database, US National Library of Medicine, North Thames Regional Library (UK), and World Health Organisation. The NZ National Bibliographic database, Ministry of Health website and library, university and medical library catalogues and the NZHTA in-house collection. Internet websites were also searched. 'Grey' literature was accessed via personal contact. Restricted to English articles (January 1, 1997 – May 31, 2000), to update the July 1997 Australian Health Technology Advisory Committee (AHTAC) review.
Are the included trials appraised for validity? Are validity criteria stated?	The review critically appraised the methodological qualities of the studies. Secondary studies were critiqued in terms of search strategy, inclusion/exclusion criteria, data synthesis and interpretation. Primary studies were coded on recruitment, blind verification, reference standard used and extent of verification. Studies were rated according to a revised hierarchy of evidence adapted from the recommended methods of the Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests. Rigorous assessment of validity was undertaken.
Are results consistent from study to study? Is homogeneity assessed?	Five systematic reviews, one meta-analysis and one primary study were identified and appraised (in addition to the AHTAC review). There was general agreement amongst the secondary studies about the limitations of the available evidence. The primary studies reviewed varied in terms of methodology, reference standard used and generalisability to standard practice.
Summary of main results	AutoPap: Only AutoPap was investigated since the other commercial brands were not available in NZ. Estimates of test sensitivity and specificity could not be reliably determined. Although there may potentially be increases in detection of low-grade abnormalities for AutoPap compared with conventional screening followed by 10% random re-screening, there is no evidence to suggest an increase in detection of high-grade abnormalities. NB: 10% random re-screening is not usual practice in Australia.
Strengths	A comprehensive literature search was performed. The aims, inclusion criteria and appraisal strategy were clearly stated. The review was rigorous, explicit and systematic.
Limitations	Only AutoPap reviewed. Restricted to English language publications.

Secondary studies

Two systematic reviews were identified (Hartmann et al 2001, Mango & Radensky 1998). The paper by Mango and Radensky (1998) was excluded from further evaluation. Although the authors stated that they employed a 'systematic review protocol', their paper does not fulfil the accepted criteria for systematic reviews. There were no inclusion/exclusion criteria for studies included in the review (all available studies were included) and included papers were not evaluated against validity criteria. Results were pooled according to a taxonomy developed *a posteriori* that linked studies with similar outcomes. However, without critical appraisal of the validity of the studies, these results cannot be interpreted.

Critical appraisal of the Hartmann et al (2001) systematic review was undertaken using established validity criteria. A summary of this detailed assessment is presented in Table 13. This paper is an update of an earlier review undertaken by the Agency for Health Care Policy and Research (AHCPR) (McCroly et al 1999).

The major limitation of the review by Hartmann et al (2001) is the search strategy which was restricted to Medline only. The authors justify this on the basis that they could identify all the studies in the previous AHCPR report from a search of Medline alone, although the papers were originally found by searching Medline, EMBASE, HealthSTAR, CancerLit and CINAHL. Even though Medline identified all of the articles found previously by a search of several databases, this does not mean that the more recent publications were also confined to Medline. A broader search strategy may have uncovered additional papers not cited in Medline. In addition, the search terms used only MeSH headings, omitting text words. As a result, any paper not indexed against these specified MeSH headings would be missed. The search was also restricted to English language articles and there was no reference to attempts to locate unpublished studies. These factors suggest that relevant publications may not have been identified.

Other than the limited search strategy, the methodology for this systematic review appeared to be rigorous. A focused research question was stated and inclusion/exclusion criteria were explicit and appropriate.

The authors found no papers that met their inclusion criteria. As with the earlier AHTAC and Broadstock reviews, the main shortcoming of relevant studies was the lack of an appropriate reference standard. Their conclusion was 'that current evidence is inadequate to gauge whether new technologies are "better" than conventional cytology'.

Table 13 Appraisal of secondary studies

Reference	Hartmann KE, Nanda K, Hall S & Myers E, 2001. 'Technologic advances for evaluation of cervical cytology: Is newer better?', <i>Obstetrical & Gynecological Survey</i> , 56 (12), 765-774.
Is there a focused research question? (ie PICO elements)	Yes, explicitly stated in the review. Patients: Women undergoing screening for cervical cancer Intervention/Diagnostic test: Computerised re-screening Comparison: Conventional cytology Outcomes: Sensitivity, specificity, predictive values, likelihood ratios Reference standard: Colposcopy and/or cervical biopsy
Are inclusion and exclusion criteria, including <i>a priori</i> , for selected studies stated?	Included studies where the test being evaluated was: <ul style="list-style-type: none"> • obtained as a screening test or as an adjunct to screening (ie not as follow-up of prior abnormality) • compared with a reference standard of colposcopy/biopsy • verified by colposcopy and/or biopsy within a 3-month interval from screening sample • reported in a fashion that allowed completion of 2x2 tables relating findings to colposcopy/biopsy results Two individuals independently reviewed the titles and abstracts. If they disagreed, the full article was reviewed and a final decision was made.
Is there an explicit and comprehensive search strategy?	Searched forward for new articles subsequent to the 1999 AHCPR report. Searched MEDLINE using MeSH headings 'cervix neoplasm', 'cervix dysplasia', 'vaginal smears/and screening'. Restricted to human, English language and years 1995 - March 2001
Are the included trials appraised for validity? Are validity criteria stated?	The investigators worked closely with the authors of the 1999 AHCPR report and applied the same criteria and data extraction techniques to allow combination of results. A methodologist and a clinician-researcher both read each full article and entered data independently into evidence tables that were later compared for agreement. Discrepancies were resolved by joint review. Validity assessment was not explicitly described but was implied by the fact that study design, test methods, location, patient population, outcome measures, prevalence of documented lesions and test characteristics were considered. No articles met the study criteria
Are results consistent from study to study? Is homogeneity assessed?	No articles met the study criteria
Summary of main results	The literature about computer-assisted image analysis is fundamentally limited by the lack of histologically-confirmed performance measures
Strengths	The aims and inclusion criteria were clearly stated and applied. Critical appraisal strategies were discussed, but no studies met the inclusion criteria
Limitations	The authors restricted the search to Medline only. The justification for this was that a search of Medline alone found all of the studies included in the AHCPR report which were originally identified from a search of Medline, EMBASE, HealthSTAR, CancerLit and CINAHL. Although Medline alone could identify all the articles found by the previous search of several databases, this does not mean that more recent publications were confined to Medline. This finding could also be a function of the search terms used, and a broader search strategy may uncover additional papers outside Medline. The search terms used only MeSH headings, omitting text words. This means that any paper not indexed with these MeSH headings would have been missed. The search was restricted to English language articles and there is no reference to attempts to locate unpublished studies. These factors indicate that the search may have missed relevant publications.

Primary studies

The search strategy identified 53 primary studies, none of which met the inclusion criteria.

Primary screening

Of the six studies identified that used computer-assisted image analysis for primary screening, five were excluded for one or more of the following reasons: populations contained both screening and high-risk patients; the results from patients with available histology could not be separated from the whole; the comparator was not the screening protocol in Australia.

One study reviewing the performance of AutoCyte SCREEN did not present sufficient information to allow a clear decision regarding inclusion or exclusion (Minge et al 2000). The patient population was drawn from three obstetric-gynaecologic specialty practices in Nebraska. It is likely, though not specified, that this group contained referred as well as screening patients. The prevalence of SIL in this clinical population is reported as 6.5 per cent, which seems high. A related rate in Australia is 6.7 per cent of all smears being reported as 'abnormal' (RCPA 2001), although the Australian figure includes slides with cells of uncertain significance in addition to those with defined abnormalities equivalent to the SIL group reported in the US data.

This paper was further evaluated on the basis that the study population may have been drawn from women presenting for screening and therefore satisfy our inclusion criteria. Smears from 2,156 women were processed with conventional Pap or AutoCyte PREP liquid based methods. Both sets of slides were read manually and the AutoCyte PREP specimens were also submitted to AutoCyte SCREEN computerised analysis. Histology was available for 134 of the patients. Sensitivity and specificity for all three methods are reported based on the biopsy findings, however these calculations cannot be verified from the data presented in the tables. Calculations of sensitivity based on the available raw data were quite different from those presented in the text.

Given the ambiguity about the composition of the study population and the lack of verification of test parameters, this study is not presented as evidence.

Re-screening

Two large Australian studies were identified but neither met the criteria for inclusion in the present review.

The study by Mitchell and Medley (1998) aimed to measure the sensitivity of PAPNET for detecting abnormal slides that had been missed on a first manual screen. One hundred and ninety five slides were selected based on: a classification of 'negative' on the initial manual screen; a change in classification to 'abnormal' following manual re-screening; and confirmation by available histology of a diagnosis of LGEA, HGEA or carcinoma in situ. These 195 slides were seeded into a total of 20,000 slides where all of the additional samples had been reported as showing no abnormality following two manual screens.

The length of time to histology was not reported for the LGEA and HGEA groups, and was stated to be 'within 24 months' for the women with carcinoma in situ. This is outside the inclusion criteria for this review which requires histology to be within six months of cytology. In addition, the reporting of results for PAPNET-assisted review of only the abnormal slides meant that the false positive rate of the PAPNET process could not be determined.

Halford (1998) examined 27,014 slides by routine manual screening followed by rapid re-screening of all slides classified as 'normal' at the initial review. Those designated as 'normal' after the second manual screen were then subjected to the PAPNET review process. Histology and colposcopy data were available for some of the 102 slides that were re-classified by the PAPNET process to 'abnormal', but these data cannot be separated from those for whom this information is not available. Some patients had results from follow-up cytology, but no time period was given to indicate whether this was a reliable indicator of cervical status. In addition, the patient population was referred specifically for PAPNET screening at the request of patient or doctor. The authors suggest that the lower than expected rates of detected abnormalities indicate a biased sample due to the selection process. There was no follow-up of patients in the group classified as 'normal'.

Triage

The present review identified a number of studies in which computer-assisted image analysis was being used for further assessment of slides classified as ASCUS (atypical squamous cells of undetermined significance) or AGUS (atypical glandular cells of undetermined significance) according to the Bethesda System. Smears are classified as ASCUS or AGUS when cellular irregularities exceed the criteria for reactive or inflammatory change but are not diagnostic of pre-neoplastic or neoplastic conditions.

It has been well demonstrated that significant numbers of patients with ASCUS or AGUS actually have high-grade lesions on histology (Emerson et al 2002, Hammoud et al 2002, Kaufman et al 1998). Referring all women with ASCUS or AGUS for colposcopy and/or biopsy to detect those with more serious pathology would mean large numbers of women having unnecessary, invasive procedures with identifiable risks and financial cost to them and the community. At present, patients are advised to return for a repeat smear at a shorter interval. If the findings persist they are then referred for colposcopy. Identifying the women at higher risk of underlying abnormality at the time of the first smear could be very useful.

In recognition of the fact that diagnosis of high-risk individuals is not an approved use for computerised image analysis systems, the identified studies noted that their work was experimental. The inclusion and exclusion criteria for the present review disallow further examination of these studies. Computer-assisted image analysis for triaging of patients with ASCUS is discussed here as a potential future use of this technology.

What are the economic considerations?

Due to insufficient evidence of clinical effectiveness based on patient outcomes and lack of information on test parameters such as sensitivity and specificity, an economic evaluation of computer-assisted image analysis for cervical screening could not be performed.

Conclusions

Safety

Computer-assisted image analysis is conducted in the laboratory on the same slides as conventional cervical cytology. Hence, the safety issues with this technology are the same as those for conventional cytological assessment methods.

Effectiveness

Three systematic reviews of computer-assisted image analysis for cervical screening cytology were identified (AHTAC 1998, Hartmann et al 2001, Broadstock 2000). All three found that there was insufficient evidence to make firm conclusions about the performance of this technology and that estimates of test sensitivity and specificity could not be reliably determined. The most recent review (Hartmann et al 2001) was unable to identify any studies that met their review protocol. Similarly, no primary studies were identified in the literature search that satisfied the inclusion/exclusion criteria for the present review.

There is a lack of evidence that computer-assisted image analysis for cervical screening cytology is as effective as conventional, manual screening.

Cost-effectiveness

Due to their being insufficient evidence of clinical effectiveness based on patient outcomes and a lack of information on test parameters such as sensitivity and specificity, an economic evaluation could not be performed.

Implications for Australian research

There is a lack of high quality evidence of effectiveness for many of the new technologies relating to cervical screening. In addition to computer-assisted image analysis, the MSAC has reviewed liquid-based cytology and human papillomavirus testing for screening and triage and found insufficient evidence to recommend any change to current practice.

International studies

Australia's National Cervical Screening Program follows standardised guidelines. It has high population participation rates and very high standards of performance at both clinical and technical levels. Many international studies are not generalisable to the Australian setting as they follow very different protocols, use different terminology for reporting cytology, are based on populations with higher prevalence of disease or have a lower standard of clinical and technical performance.

In addition, the sample size of identified studies currently underway on these new technologies will be too small to provide evidence to inform policy decisions on population screening.

Lack of an appropriate reference standard

Lack of an appropriate reference standard is the main shortcoming of the studies that evaluate cervical screening technologies. To calculate measures of diagnostic accuracy (sensitivity, specificity, positive predictive value, negative predictive value) the test under investigation must be compared to a reference test or 'gold standard'.

Implementation of reference tests in healthy populations undergoing screening can be problematic. This is particularly so in the case of cervical screening where the gold standard of histology is an invasive test. The inclusion criteria for the present review required that measures of test accuracy were based on a reference standard of cervical histology for high-grade lesions (CIN II and above) and, as a minimum, appearance at colposcopy for lower grade lesions (CIN I and below) that was undertaken within six months of cytology. While long-term follow-up of patient outcomes would have been acceptable, no relevant studies met these criteria.

Australian research

Although international studies are currently underway, it is anticipated that there will be considerable limitations to generalisability of the results within the Australian context. Large, adequately powered studies within the Australian cervical screening program using rigorous methodology and incorporating long-term follow-up are required to accurately assess the potential of these new technologies.

Recommendation

The MSAC recommends that as there is insufficient evidence to draw conclusions on the appropriate use of computer-assisted image analysis versus manual cytology screening, there are no grounds to change current funding arrangements

The Minister for Health and Ageing accepted this recommendation on 8 August 2003.

Appendix A MSAC's terms of reference and membership

The MSAC's terms of reference are to:

- advise the Minister for Health and Ageing on the strength of evidence pertaining to new and existing medical technologies and procedures in relation to their safety, effectiveness and cost-effectiveness and under what circumstances public funding should be supported;
- advise the Minister for Health and Ageing on which new and existing medical technologies and procedures should be funded on an interim basis to allow data to be assembled to determine their safety, effectiveness and cost-effectiveness;
- advise the Minister for Health and Ageing on references related either to new or existing medical technologies and procedures; and
- undertake health technology assessment work referred by the Australian Health Ministers' Advisory Council (AHMAC) and report its findings to the AHMAC.

The membership of the MSAC comprises a mix of expertise covering pathology, nuclear medicine, surgery, specialist medicine and general practice, clinical epidemiology and clinical trials, health economics, consumer issues, and health administration and planning:

Member	Expertise or Affiliation
Dr Stephen Blamey (Chair)	general surgery
Professor Bruce Barraclough	general surgery
Professor Syd Bell	pathology
Dr Paul Craft	clinical epidemiology and oncology
Professor Jane Hall	health economics
Dr Terri Jackson	health economics
Ms Rebecca James	consumer health issues
Professor Brendon Kearney	health administration and planning
Associate Professor Richard King	internal medicine
Dr Ray Kirk	health research
Dr Michael Kitchener	nuclear medicine
Mr Lou McCallum	consumer health issues
Dr Ewa Piejko	general practice
Mr Chris Sheedy	Assistant Secretary Diagnostics and Technology Branch Commonwealth Department of Health and Ageing
Professor John Simes	clinical epidemiology and clinical trials
Professor Richard Smallwood	Chief Medical Officer Commonwealth Department of Health and Ageing
Dr Robert Stable	representing the Australian Health Ministers' Advisory Council
Professor Bryant Stokes	neurology
Professor Ken Thomson	radiology
Dr Douglas Travis	urology

Appendix B Supporting Committee

Supporting Committee for MSAC Reference 12c: Computer-assisted image analysis for cervical screening

Professor Brendon Kearney (Chair) MBBS, FRACP, FRACMA Executive Director South Australian Department of Human Services	MSAC member
Professor Sydney Bell MBBS, FRCPA, MD, FAFPHM Area Director of Microbiology South East Sydney Area Health Service	co-opted MSAC member
Dr Paul Craft MBBS, FRACP, MPH Director, Medical Oncology Unit The Canberra Hospital Garran, ACT	co-opted MSAC member
Dr Ray Kirk BSc, PhD Director, New Zealand Health Technology Assessment Unit Clinical Senior Lecturer in Public Health Christchurch School of Medicine and Health Sciences University of Otago, New Zealand	co-opted MSAC member
Dr Alistair Lochhead MBBS, FRCPA, FIAC Staff Specialist, Southern Pathology Staff Specialist in Anatomical Pathology The Wollongong Hospital, NSW	nominated by the Royal College of Pathologists of Australasia
Ms Kathleen Mazzella Health Consumers Council WA Gynaecological Awareness Information Network	nominated by the Consumers' Health Forum of Australia
Dr Heather Mitchell MBBS, MD, MSc, FRACP, FAFPHM Deputy Director, Victorian Cytology Service	co-opted epidemiologist
Dr Marion Saville MBChB, Am Bd, FIAC, Grad Dip Med (Clin Epi) Director, Victorian Cytology Service	nominated by the Australian Society of Cytology
Dr Judy Straton MD, MPH, FAFPHM Senior Medical Advisor, Primary Care Division Department of Health and Ageing	observer
Dr Lynnette Wray MBBS, FPA Cert., FACSH, MMedVen General Practitioner Coral Lloyd Family Planning Clinic Sydney, NSW	nominated by the Royal Australian College of General Practitioners
Linda Marshall BA, BSc, MBA MSAC Project Manager	Health Technology Section Department of Health and Ageing

Appendix C HTA and other websites and databases searched

Health technology assessment (HTA) agency websites

Agence d'Evaluation des Technologies et des Modes d'Intervention en Santé (AÉTMIS) <http://www.aetmis.gouv.qc.ca/> [Accessed 4 November 2002]

Alberta Heritage Foundation for Medical Research (AHFMR) <http://www.ahfmr.ab.ca/index.html> [Accessed 4 November 2002]

Agency for Healthcare Research and Quality <http://www.ahrq.gov/> [Accessed 4 November 2002]

The Centre for Health Services and Policy Research (CHSPR) <http://www.chspr.ubc.ca/> [Accessed 4 November 2002]

Canadian Coordinating Office for Health Technology Assessment (CCOHTA) <http://www.ccohta.ca/> [Accessed 4 November 2002]

Danish Institute for Health Technology Assessment (DIHTA) <http://www.dihta.dk/> [Accessed 4 November 2002]

Finnish Office for Health Care Technology Assessment <http://www.stakes.fi/finohta/e/> [Accessed 4 November 2002]

Health Council of the Netherlands <http://www.gr.nl/engels/welcome/frameset.htm> [Accessed 4 November 2002]

Minnesota Health Technology Advisory Council <http://www.health.state.mn.us/htac/> [Accessed 4 November 2002]

Institute for Clinical Systems Improvement <http://www.icsi.org/talist.htm> [Accessed 4 November 2002]

International Network of Agencies for Health Technology Assessment (INAHTA) <http://www.inahta.org/> [Accessed 4 November 2002]

International Society of Technology Assessment in Health Care (ISTAHC) <http://www.istahc.org/> [Accessed 4 November 2002]

Medical Technology & Practice Patterns Institute (MTPPI) <http://www.mtppi.org/> [Accessed 4 November 2002]

National Health Service Centre for reviews and dissemination, University of York <http://nhscrd.york.ac.uk/welcome.html> [Accessed 4 November 2002]

The National Coordinating Centre for Health Technology Assessment (NCCHTA) <http://www.hta.nhsweb.nhs.uk/> [Accessed 4 November 2002]

National Horizon Scanning Centre
<http://www.bham.ac.uk/PublicHealth/horizon/> [Accessed 4 November 2002]

National Institute for Clinical Excellence (NICE) <http://www.nice.org.uk/>
[Accessed 4 November 2002]

New Zealand Health Technology Assessment (NZHTA)
<http://nzhta.chmeds.ac.nz/> [Accessed 4 November 2002]

The Swedish Council on Technology Assessment in Health Care
<http://www.sbu.se/admin/index.asp> [Accessed 4 November 2002]

The Norwegian Centre for Health Technology Assessment
<http://www.oslo.sintef.no/smm/news/FramesetNews.htm>
[Accessed 4 November 2002]

Veterans' Affairs Technology Assessment Program (VATAP)
<http://www.va.gov/resdev/ps/pshsrd/mdrc.htm#HealthCareTechnologyAssessment>
[Accessed 4 November 2002]

WHO Health Technology Assessment Programme (Collaborating Centres)
http://www.who.int/pht/technology_assessment/index.html
[Accessed 4 November 2002]

Clinical Trial Register websites

Centre Watch Clinical Trials Listing Service <http://www.centerwatch.com/>
[Accessed 4 November 2002]

Clinical Trials.com <http://clinicaltrials.com/> [Accessed 4 November 2002]

Clinical Trials.gov <http://clinicaltrials.gov/> [Accessed 4 November 2002]

Current Controlled Trials <http://www.controlled-trials.com/>
[Accessed 4 November 2002]

FDA Clinical trials links <http://www.fda.gov/oc/oha/default.htm#clinical>
[Accessed 4 November 2002]

NHMRC Clinical Trial Registry
<http://www.ctc.usyd.edu.au/6registry/reg1.shtml> [Accessed 4 November 2002]

Society for Clinical Trials <http://www.sctweb.org/> [Accessed 4 November 2002]

Trials Central <http://www.trialscentral.org/> [Accessed 4 November 2002]

UK National Research Registry <http://www.update-software.com/National/>
[Accessed 4 November 2002]

Appendix D Studies included in the review

Australian Health Technology Advisory Committee, 1998. 'Review of automated and semi-automated cervical screening devices'. Commonwealth Department of Health and Family Services, Canberra.

Broadstock M., 2000. 'Effectiveness and cost-effectiveness of automated and semi-automated cervical screening devices. A systematic review of the literature'. NZHTA Report 2000, 3(1) Christchurch.

Hartmann, K.E., Nanda, K., Hall S. & Myers, E., 2001. 'Technologic advances for evaluation of cervical cytology: Is newer better?', *Obstetrical & Gynecological Survey*, 56 (12), 765-774.

Mango, L.J. & Radensky, P.W., 1998. 'Interactive neural-network-assisted screening. A clinical assessment', *Acta Cytologica*, 42 (1), 233-245.

Minge, L., Fleming, M., VanGeem, T. & Bishop, J.W., 2000. 'Autocyte prep (r) system vs conventional cervical cytology - comparison based on 2,156 cases', *Journal of Reproductive Medicine*, 45 (3), 179-184.

Appendix E Studies excluded from the review

Narrative Review

Anonymous, 1998. 'Monolayer slide preparation and automated slide reading systems for cervical cancer screening – clinical-effectiveness analysis', *Tecnologica MAP Supplement*, 34-39.

Anonymous, 1998. 'New pap test screening techniques', *International Journal of Gynecology & Obstetrics*, 63 (3), 312-314.

Bishop, J.W., Marshall, C.J. & Bentz, J.S., 2000. 'New technologies in gynecologic cytology', *Journal of Reproductive Medicine*, 45 (9), 701-719.

Bondi, A. & Ghidoni, D., 2001. 'Computer assisted cytology', *Tumori*, 87 (5), S7-8.

Cohn, E. & Herzog, T.J., 2001. 'New innovations in cervical cancer screening', *Clinical Obstetrics & Gynecology*, 44 (3), 538-549.

Coleman, D.V., 1998. 'Evaluation of automated systems for the analysis of cervical smears', *Cytopathology*, 9 (6), 359-368.

Halford, J., 1998. 'Cervical cytology - a review of evolving technologies', *Australian Journal of Medical Science*, 19 (1), 8-19.

Lisboa, P.J.G., 2002. 'A review of evidence of health benefit from artificial neural networks in medical intervention', *Neural Networks*, 15 (1), 11-39.

Mango, L.J., 1997. 'Clinical validation of interactive cytologic screening - automating the search, not the interpretation', *Acta Cytologica*, 41 (1), 93-97.

Mango, L.J., 1998. 'Neural network-assisted cervical cancer screening', *Journal of Clinical Ligand Assay*, 21 (2), 203-207.

Meijer, C.J. & Walboomers, J.M., 2000. 'Cervical cytology after 2000: Where to go?', *Journal of Clinical Pathology*, 53 (1), 41-43.

Nuovo, J., Melnikow, J. & Howell, L.P., 2001. 'New tests for cervical cancer screening', *American Family Physician*, 64 (5), 780-786.

Rosenthal, D.L., 1997. 'Computerized scanning devices for pap smear screening - current status and critical review', *Clinics in Laboratory Medicine*, 17 (2), 263.

Russo, J.F., 2000. 'Controversies in the management of abnormal pap smears', *Current Opinion in Obstetrics & Gynecology*, 12 (5), 339-343.

Solomon, H.M. & Frist, S., 1998. 'Papnet testing for HSILS - the few cell small cell challenge', *Acta Cytologica*, 42 (1), 253-259.

Valente, P.T. & Schantz, H.D., 2001. 'Cytology automation: An overview', *Laboratory Medicine*, 32 (11), 686-690.

Wilbur, D.C. & Norton, M.K., 2002. 'The primary screening clinical trials of the Tripath Autopap® system', *Epidemiology*, 13 (3 Suppl S), S30-S33.

Letter

Baird, P.J., 1997. 'Evaluation of the Papnet system in a general pathology service', *Medical Journal of Australia*, 167 (5), 285-286.

Maughan, K.L., 1998. 'Rescreening of Pap smears', *Journal of Family Practice*, 46 (4), 279-280.

Rosenthal, D.L. & American Society of Cytopathology., 2000. 'ASC statement on new technologies in cervical cytology screening (June 2000). American Society of Cytopathology', *Diagnostic Cytopathology*, 23 (6), 430.

Schechter, C.B., 1998. 'Rescreening of cervical Papanicolaou smears using Papnet', *Journal of the American Medical Association*, 279 (22), 1787.

Shaver, J.L., 1998. 'Increased cervical cytologic smear access through increased capacity', *Journal of Reproductive Medicine*, 43 (11), 1005-1006.

Editorial, position paper or discussion paper

American Society of Cytopathology, 2000. 'American Society of Cytopathology statement on new technologies in cervical cytology screening. June 2000', *Analytical & Quantitative Cytology & Histology*, 22 (4), 351.

Bartels, P.H., Bibbo, M., et al, 1998. 'Computerized screening devices and performance assessment - development of a policy towards automation - IAC Task Force summary', *Acta Cytologica*, 42 (1), 59-68.

Boronow, R.C. & Cavett, J.R., 3rd, 1998. 'When your patient asks: "Doctor, I read there have been serious Pap smear errors. Shouldn't I get one of those new computer Pap smears?!", *Journal of the Mississippi State Medical Association*, 39 (4), 136-141.

Frale, W.J., 2001. 'Cervical screening by neural networks', *Cancer Cytopathology*, 93 (3), 171-172.

Krieger, P. & Bibbo, M., 1997. 'Our journey towards improved accuracy in cytology: The role of new technologies', *Acta Cytologica*, 41 (1), 11-14.

Wilson, P.O., 1997. 'Automation in cytopathology', *Cytopathology*, 8 (5), 295-297.

Patient information

Robb-Nicholson, C., 1999. 'By the way, doctor. I read with interest your article regarding conventional Pap smears, Papnet, and Autopap. Lately I've been reading about another test called Thinprep. I'm not sure which one to request. Which of these techniques is better for detecting cervical cancer?', *Harvard Women's Health Watch*, 6 (12), 8.

Not generalisable to Australian screening setting (high-risk or referred patients; and/or not compared with conventional manual screening protocol)

Alasio, L.M., Alphandery, C., et al, 2001. 'Performance of the Autopap primary screening system in the detection of high-risk cases in cervicovaginal smears', *Acta Cytologica*, 45 (5), 704-708.*

Bergeron, C., Masseroli, M., et al, 2000. 'Quality control of cervical cytology in high-risk women - Papnet system compared with manual rescreening', *Acta Cytologica*, 44 (2), 151-157.

Kok, M.R., Boon, M.E., et al., 2000. 'Cytological recognition of invasive squamous cancer of the uterine cervix: Comparison of conventional light-microscopical screening and neural network-based screening', *Human Pathology*, 31 (1), 23-28.*

Kok M.R., Boon M.E., et al, 2001. 'Less medical intervention after sharp demarcation of grade 1-2 cervical intraepithelial neoplasia smears by neural network screening', *Cancer*, 93 (3), 173-178.

Rozendaal, L., Westerga, J., et al, 2000. 'PCR based high risk HPV testing is superior to neural network based screening for predicting incident CIN III in women with normal cytology and borderline changes', *Journal of Clinical Pathology*, 53 (8), 606-611.

Sherman, M.E., Schiffman, M., et al, 1998. 'Performance of a semiautomated Papanicolaou smear screening system - results of a population-based study conducted in Guanacaste, Costa Rica', *Cancer Cytopathology*, 84 (5), 273-280.*

Tench, W.D., 2002. 'Validation of Autopap® primary screening system sensitivity and high-risk performance', *Acta Cytologica*, 46 (2), 296-302.*

Veneti, S., Papaefthimiou, M., et al, 1999. 'Papnet for cervical cytology screening - experience in Greece', *Acta Cytologica*, 43 (1), 30-33.*

No histology or colposcopy; histology available for sub-sample but unable to separate in results; or histology available but not within 6 months of cytology

Chang A.R., Lin W.F., et al, 2002. 'Can technology expedite the cervical cancer screening process? A Hong Kong experience using the Autopap primary screening system with location-guided screening capability', *American Journal of Clinical Pathology*, 117 (3), 437-443.

Doornewaard, H., van der Schouw, Y.T., et al, 1999. 'The diagnostic value of computer-assisted primary cervical smear screening: A longitudinal cohort study', *Modern Pathology*, 12 (11), 995-1000.*

Koss, L.G., Sherman, M.E., et al, 1997. 'Significant reduction in the rate of false-negative cervical smears with neural network-based technology (Papnet testing system)', *Human Pathology*, 28 (10), 1196-1203.*

Patten, S.F., Lee, J.S.J., et al, 1997. 'The Autopap 300 QC system multicenter clinical trials for use in quality control rescreening of cervical smears. I: A prospective intended use study', *Cancer Cytopathology*, 81 (6), 337-342.

Patten, S.F., Lee, J.S.J., et al, 1997. 'The Autopap 300 QC system multicenter clinical trials for use in quality control rescreening of cervical smears. II: Prospective and archival sensitivity studies', *Cancer Cytopathology*, 81 (6), 343-347.

Prismatic Project Management Team, 1999. 'Assessment of automated primary screening on Papnet of cervical smears in the prismatic trial (vol 353, pg 1381, 1999)', *Lancet*, 353 (9169), 2078.

Renshaw, A.A., 2002. 'Measuring sensitivity in gynecologic cytology - a review', *Cancer Cytopathology*, 96 (4), 210-217.

Vassilakos, P., Petignat, P., et al, 2002. 'Primary screening for cervical cancer precursors by the combined use of liquid-based cytology, computer-assisted cytology and HPV DNA testing', *British Journal of Cancer*, 86 (3), 382-388.

Computer-assisted image analysis used for triage (not screening)

Chieng, D.C., Elgert, P.A., et al, 2000. 'Use of computer-assisted rescreening as an ancillary tool to subclassify AGUS cervical smears', *Diagnostic Cytopathology*, 23 (3), 165-170.

Kaufman, R.H., Schreiber, K. & Carter, T., 1998. 'Analysis of atypical squamous (glandular) cells of undetermined significance smears by neural network-directed review', *Obstetrics & Gynecology*, 91 (4), 556-560.

Kok, M.R., Habers, M.A., et al, 1998. 'New paradigm for ASCUS diagnosis using neural networks', *Diagnostic Cytopathology*, 19 (5), 361-366.*

Lerman, E., Colomo, L., et al, 1998. 'Rescreening of atypical cervicovaginal smears using Papnet', *Cancer Cytopathology*, 84 (6), 361-365.

Stastny, J.F., Remmers, R.E., et al, 1997. 'Atypical squamous cells of undetermined significance - a comparative review of original and automated rescreen diagnosis of cervicovaginal smears with long term follow-up', *Cancer Cytopathology*, 81 (6), 348-353.

* Several studies were excluded for more than one reason.

Abbreviations

AGUS	Atypical glandular cells of undetermined significance
AHCPR	Agency for Health Care Policy and Research
AHMAC	Australian Health Ministers Advisory Council
AHTAC	Australian Health Technology Advisory Committee
AIHW	Australian Institute of Health and Welfare
ASCUS	Atypical squamous cells of undetermined significance
CIN	Cervical intraepithelial neoplasia
HGEA	High-grade epithelial abnormality
HPV	Human papillomavirus
HSIL	High-grade squamous intraepithelial lesion
LGEA	Low-grade epithelial abnormality
LSIL	Low-grade squamous intraepithelial lesion
MSAC	Medical Services Advisory Committee
NCSP	National Cervical Screening Program
NHMRC	National Health and Medical Research Council
NPV	Negative predictive value
NZHTA	New Zealand Health Technology Assessment
Pap smear	Papanicolaou smear
PPV	Positive predictive value
RCPA	Royal College of Pathologists of Australia
RCT	Randomised controlled trial
SIL	Squamous intraepithelial lesion
WNL	Within normal limits

References

Australian Health Technology Advisory Council, 1998. 'Review of automated and semi-automated cervical screening devices', Commonwealth Department of Health and Family Services, Canberra.

Australian Department of Health and Ageing 2002, *National cervical screening program* [Internet]. Australian Department of Health and Ageing, Available from: <http://www.health.gov.au/pubhlth/strateg/cancer/cervical/> [Accessed 28 November 2002]

Australian Institute of Health and Welfare, 1998. 'Breast and cervical cancer screening in Australia 1996–1997', Australian Institute of Health and Welfare, Canberra.

Australian Institute of Health and Welfare, 2000. 'Cervical screening in Australia 1997-1998', Australian Institute of Health and Welfare, Canberra.

Australian Institute of Health and Welfare, 2002. 'Cervical screening in Australia 1998-1999', Australian Institute of Health and Welfare, Canberra.

Australian Institute of Health and Welfare & Australian Association of Cancer Registries, 2000. Cancer in Australia 1997: 'Incidence and mortality data for 1997 and selected data for 1998 and 1999', Australian Institute of Health and Welfare, Canberra.

Bishop, J.W., Marshall, C.J. & Bentz, J.S., 2000. 'New technologies in gynecologic cytology', *Journal of Reproductive Medicine*, 45 (9), 701-719.

Bosch, F.X., Manos, M.M., et al, 1995. 'Prevalence of human papillomavirus in cervical cancer: A worldwide perspective. International biological study on cervical cancer (IBSCC) study group', *Journal of the National Cancer Institute*, 87 (11), 796-802.

Broadstock M., 2000. 'Effectiveness and cost-effectiveness of automated and semi-automated cervical screening devices. A systematic review of the literature', New Zealand Health Technology Assessment, Report 2000; 3(1), Christchurch.

Cannistra, S.A. & Niloff, J.M., 1996. 'Cancer of the uterine cervix', *New England Journal of Medicine*, 334 (16), 1030-1038.

Centre for Evidence Based Medicine, 2001. *Levels of evidence and grades of recommendation*. [Internet] Oxford - Centre for Evidence Based Medicine. Available from: http://www.cebm.net/levels_of_evidence.asp. [Accessed September 2003)

Chalmers, I. & Altman, D.G., 1995. *Systematic reviews*, BMJ Publishing Group, London.

Channen, W., 1990. 'The CIN saga – the biological and clinical significance of cervical intraepithelial neoplasia', *Australian and New Zealand Journal of Obstetrics and Gynaecology*, 30 (1), 18-23.

Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests, 1996, *Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests: Recommended methods* [Internet]. Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests, Available from: <http://www.cochrane.org.au> [Accessed 4 November 2002].

Commonwealth Department of Health and Family Services (1998). 'Screening for the prevention of cervical cancer', Commonwealth Department of Health and Family Services, Canberra.

Cuzick, J., Beverley, E., et al, 1999. 'HPV testing in primary screening of older women', *British Journal of Cancer*, 81 (3), 554-558.

Division of STD Prevention, CDC Department of Health and Human Services (1999). 'Prevention of genital HPV infection and sequelae: Report of an external consultants' meeting', Centers for Disease Control and Prevention, Atlanta, Georgia.

Emerson, R.E., Puzanov, A., et al, 2002. 'Long-term follow-up of women with atypical squamous cells of undetermined significance (ASCUS)', *Diagnostic Cytopathology*, 27 (3), 153-157.

Guyatt, G.H., Sackett, D.L. & Cook, D.J., 1993. 'Users' guides to the medical literature, II: How to use an article about therapy or prevention. Are the results of the study valid? Evidence-based medicine working group', *Journal of the American Medical Association*, 270 (21), 2598-2601.

Hammoud, M.M., Haefner, H.K., et al, 2002. 'Atypical glandular cells of undetermined significance. Histologic findings and proposed management', *Journal of Reproductive Medicine*, 47 (4), 266-270.

Hartmann, K.E., Nanda, K., et al, 2001. 'Technologic advances for evaluation of cervical cytology: Is newer better?', *Obstetrical & Gynecological Survey*, 56 (12), 765-774.

Jadad, A., Moore, R., et al, 1996. 'Assessing the quality of reports of randomized clinical trials: Is blinding necessary?', *Controlled Clinical Trials*, 17 (1), 1-12.

Jaeschke, R., Guyatt, G. & Sackett, D., 1994. 'Users' guide to the medical literature, III: How to use an article about a diagnostic test: Are the results of the study valid?', *Journal of the American Medical Association*, 271, 389-391.

Jelfs, P.L., (1995). 'Cervical cancer in Australia', Australian Institute of Health and Welfare, Canberra.

Kaufman, R.H., Schreiber, K. & Carter, T., 1998. 'Analysis of atypical squamous (glandular) cells of undetermined significance smears by neural network-directed review', *Obstetrics & Gynecology*, 91 (4), 556-560.

Koss, L.G., Sherman, M.E., et al, 1997. 'Significant reduction in the rate of false-negative cervical smears with neural network-based technology (Papnet testing system)', *Human Pathology*, 28 (10), 1196-1203.

Liaw, K.L., Glass, A.G., et al, 1999. 'Detection of human papillomavirus DNA in cytologically normal women and subsequent cervical squamous intraepithelial lesions', *Journal of the National Cancer Institute*, 91 (11), 954-960.

Mango, L.J. & Radensky, P.W., 1998. 'Interactive neural-network-assisted screening. A clinical assessment', *Acta Cytologica*, 42 (1), 233-245.

McCrary, D.C., Mather, D.B., et al, 1999. 'Evaluation of cervical cytology', Evidence Report/Technology Assessment No 5, AHCPR Publication No 99-E010, Agency for Healthcare Evaluation and Policy, Rockville Maryland. February 1999.

McIndoe, W.A., McLean, M.R., et al, 1984. 'The invasive potential of carcinoma in situ of the cervix', *Obstetrics and Gynecology*, 64 (4), 451-458.

Melkert, P.W., Hopman, E., et al, 1993. 'Prevalence of HPV in cytomorphologically normal cervical smears, as determined by the polymerase chain reaction, is age-dependent', *International Journal of Cancer*, 53 (6), 919-923.

Minge, L., Fleming, M., et al, 2000. 'Autocyte Prep® system vs conventional cervical cytology - comparison based on 2,156 cases', *Journal of Reproductive Medicine*, 45 (3), 179-184.

Mitchell, H. & Medley G., 1998. 'Detection of laboratory false negative smears by the PAPNET cytologic screening system', *Acta Cytologica*, 42 (1), 265-270.

Mitchell, H., Higgins, V. & Burrows, C., 2002. 'Victorian Cervical Cytology Register: Statistical Report 2001', Victorian Cytology Service. April 2002

NHMRC, 1994. 'Screening to prevent cervical cancer: Guidelines for the management of women of screen detected abnormalities', Commonwealth Department of Human Services and Health, Canberra.

NHMRC, 1999. 'A guide to the development, implementation and evaluation of clinical practice guidelines', National Health and Medical Research Council, Canberra.

NHMRC, 2000. 'How to use the evidence: assessment and application of scientific evidence', National Health and Medical Research Council, Canberra.

NPAAC, 2001. 'Guidelines for the use of fluid based collection systems and automated and semi-automated screening devices in the practice of gynaecological (cervical) cytology', National Pathology Accreditation Advisory Council, Department of Health and Aged Care, Canberra.

NHS Centre for Reviews and Dissemination 2001, *Undertaking systematic reviews of research on effectiveness. CRD's guidance for those carrying out or commissioning reviews* [Internet]. NHS Centre for Reviews and Dissemination, Available from: <http://www.york.ac.uk/inst/crd/report4.htm> [Accessed 28 November 2002].

Nobbenhuis, M.A., Walboomers, J.M., et al, 1999. 'Relation of human papillomavirus status to cervical lesions and consequences for cervical-cancer screening: A prospective study', *Lancet*, 354 (9172), 20-25.

RCPA, 2001. 'Performance standards for Australian laboratories reporting cervical cytology', Royal College of Pathologists Australasia, Sydney.

Sackett, D., Strauss, S., et al, 2000. *Evidence-based medicine: How to practice and teach EBM*, 2nd ed, Churchill Livingstone, Edinburgh.

Schulz, K., Chalmers, I., et al, 1995. 'Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials', *Journal of the American Medical Association*, 273 (5), 408-412.

Shah and Howley *Papilloma Virus*, in Fields, D., et al, (Eds) *Fields Virology* 3rd Ed. Philadelphia, Lippencott Raven, 1996. Quoted by Bosch, F. X. *The worldwide prevalence of HPV* Paper presented at HPV Summit99, Chamonix France, 1999. Proceedings published by National HPV and Cervical Cancer Resource Center, American Social Health Association, Research Triangle Park, North Carolina. <http://www.ashstd.org/>

Tripath Imaging 1999. 'Cancer detection and prevention'. Burlington, North Carolina.

Wain, G.V., Farnsworth, A. & Hacker, N.F., 1992. 'The Papanicolaou smear histories of 237 patients with cervical cancer', *Medical Journal of Australia*, 157 (1), 14-16.

Walboomers, J.M., Jacobs, M.V., et al, 1999. 'Human papillomavirus is a necessary cause of invasive cervical cancer worldwide', *Journal of Pathology*, 189 (1), 12-19.