Automation-assisted and liquid-based cytology for cervical cancer screening

March 2009

MSAC application 1122

Assessment report

ISBN (Print) 978-1-74241-017-3

ISBN (Online) 978-1-74241-018-0

ISSN (Print) 1443–7120

ISSN (Online) 1443–7139

First printed September 2009

Electronic copies of the report can be obtained from the Medical Service Advisory Committee's Internet site at http://www.msac.gov.au/

Printed copies of the report can be obtained from:

The Secretary
Medical Services Advisory Committee
Department of Health and Ageing
Mail Drop 106
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 (MSAC) is an independent committee which has been established to provide advice to the 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.

MSAC's advice does not necessarily reflect the views of all individuals who participated in the MSAC evaluation.

Publication approval number: P3-6049

Contents

Executive summary ..... **x**

Introduction **1**

Background **2**

Cervical cancer cytology 2
The National Cervical Screening Program 3
Intended purpose of LBC 5
Clinical need for LBC 5
Current treatment of cervical cancer and CIN 9
Existing procedures 9
New and emerging cervical cancer technologies 10
Potential impact of LBC 11
Reference standard 12
Comparator 12
Methodological issues 12
Marketing status of the technology 12
Current reimbursement arrangement 13
MSAC 2002 and 2003 reviews 13

Approach to assessment **15**

Research question 15
Assessment strategy 15
Review of the literature 16
Appraisal of the evidence 20
Data extraction and synthesis 24
Expert advice 24

Results of assessment **25**

Is it safe? 25
Is it effective? 25
Liquid-based cytology 25
Included studies 26
Test accuracy 29
Unsatisfactory rate 32
Test yield 33
Automated slide reading 33
Comparison to conventional cytology 34
Comparison to manual LBC slide reading 49
Direct comparison of LBC, automated LBC reading and conventional Pap cytology 58
Conclusions 61
What are the economic considerations? 62

Other relevant considerations 99
Discussion 100
LBC 100
Automated LBC 101
Cost-effectiveness analysis 103
Conclusions 106
Safety 106
Effectiveness 106
Economic considerations 107
Appendix A MSAC terms of reference and membership 109
Appendix B Advisory Panel and Health Technology Assessors 110
Appendix C FIGO Staging of cervical cancer 112
Appendix D MSAC 2002 and 2003 reviews 113
Appendix E Databases searched 115
Appendix F Included studies 117
Appendix G Clinical flowchart 136
Appendix H Existing HTA reports 2002–2008 138
Appendix I Davey et al. (2006) LBC review 139
Appendix J Detailed model data sources 140
Appendix K Derivation of test probability matrices for cytology 153
Abbreviations 171
References 174

Tables

Table 1 Comparison of the Australian Modified Bethesda System, 2004, and the US Bethesda System, 2001. 5

Table 2 Classification of histological abnormalities as grades of cervical intraepithelial neoplasia (CIN). 5

Table 3 Cumulative relative rates of survival of invasive cervical cancer at 5 and 10 years, by extent of disease and corresponding FIGO stage, NSW, 1991–2000. 7

Table 4 Electronic databases searched to identify primary studies of automated slide reading. 17

Table 5 Search terms for automated slide reading (EMBASE.com). 17

Table 6 Electronic databases searched to identify primary studies of LBC. 18

Table 7 Search terms for LBC (EMBASE.com). 18

Table 8 Inclusion/exclusion criteria for identification of relevant studies. 18

Table 9 Criteria used to assess the quality of diagnostic accuracy studies—the QUADAS tool (adapted from Whiting et al. 2003). 21

Table 10 Criteria used to assess the quality of effectiveness studies (adapted from NHMRC 2000 and CRD 2001). 21

Table 11 Grading system for the appraisal of studies evaluating diagnostic tests. 22

Table 12 NHMRC designations of levels of evidence. 23

Table 13 Characteristics and appraisal of systematic reviews and HTAs of LBC. 28

Table 14 Estimates of the relative accuracy of LBC versus CC for CIN 2+ from Arbyn et al. (2008). 30

Table 15 Subgroup analysis of relative accuracy estimates of LBC versus CC for CIN 2+ from Arbyn et al. (2008). 31

Table 16 Characteristics and quality of studies reporting on the relative accuracy of automated slide reading and conventional Pap smear cytology with manual slide reading. 38

Table 17 Estimates of the relative accuracy of automation-assisted versus conventional reading of cervical cytology for CIN 2+. 43

Table 18 Estimates of the relative accuracy of automation-assisted versus conventional reading of glandular cervical cytology for AIS+. 44

Table 19 Slide classifications of automation-assisted reading of LBC versus manual reading of CC slides. 46

Table 20 Yield of automation-assisted reading of LBC versus manual reading of CC (Davey et al. 2007). 46

Table 21 Reading times of automation-assisted versus conventional reading of cervical cytology. 47

Table 22 Characteristics and quality of studies reporting on the relative accuracy of manual and automated reading of LBC slides. 51

Table 23 Estimates of the relative TP and FP rates of ThinPrep Imager–assisted versus manual reading of LBC for CIN 2+. 55
Table 24 Estimates of the relative accuracy of manual versus ThinPrep Imager-assisted reading of LBC for HSIL/CIN 2+. 56
Table 25 Process outcomes of automation-assisted versus manual reading of cervical LBC samples. 57
Table 26 Process outcomes of LBC, automation-assisted reading of LBC and manual reading of conventional cervical cytology (Roberts et al. 2007). 59
Table 27 Direct comparison of the relative accuracy of LBC, automation-assisted reading of LBC and manual reading of conventional cervical cytology for CIN 2+ in *N* = 11 416 routine samples (Roberts et al. 2007). 60
Table 28 Implied sensitivity and specificity of conventional cytology for CIN 2+, based on the baseline test probability matrix. 69
Table 29 Implied sensitivity and specificity of manual and automated LBC for CIN 2+, relative to conventional cytology (test threshold of possible pLSIL)..... 69
Table 30 Model values for rates of unsatisfactory cytology, by test technology. 70
Table 31 Modelled test characteristics of Hybrid Capture II HPV test (based on international data). 70
Table 32 Cost items used in the modelled economic evaluation. 72
Table 33 Average costs of medical consultation (Medicare Australia 2007). 73
Table 34 Problems managed per GP encounter—single-service weighting (AIHW (2008)—General practice activity in Australia 2006–07 (Table 7.1)). 73
Table 35 Cost of screening tests (CC, LBC with manual reading and automated LBC)..... 74
Table 36 Cost of results or referral consultation. 74
Table 37 Cost of colposcopy, no biopsy. 75
Table 38 Cost of colposcopy, with biopsy. 75
Table 39 Cost of cytology done at colposcopy. 76
Table 40 Cost of treating CIN 2/3. 76
Table 41 Cost of treating CIN 2/3. 77
Table 42 Cost of CIN 2/3 follow-up (CC, LBC with manual reading and automated LBC) (based on 2005 NHMRC guidelines). 78
Table 43 Summary treatment and work-up costs by FIGO stage and disease extent. 79
Table 44 Summary treatment and work-up costs by FIGO stage. 80
Table 45 Cost of surgical managements. 81
Table 46 Cost of non-surgical managements. 81

Table 47 Unit costs of work-up procedures. 82
Table 48 Model prediction versus actual data for detection of histological high disease, and cervical cancer incidence and mortality.* 87
Table 49 Predicted and reported sensitivity and specificity of manual LBC, relative to CC, for CIN 2+. 88
Table 50 Predicted and reported additional TPs and FPs from automated LBC (test threshold of possible HSIL, excluding glandular where possible). 88
Table 51 Baseline results of economic evaluation. 89
Table 52 Predicted costs, effects, and incremental cost-effectiveness ratios, by cytology test technology. 91
Table 53 Predicted process outcomes in Australia, by cytology test technology (per annum). 91
Table 54 Predicted health outcomes in Australia, by cytology test technology (per annum). 91
Table 55 Additional interventions required to prevent one incident cancer case and one cancer death. 92
Table 56 Summary of range of parameter values used for sensitivity analysis. 94
Table 57 Financial implications of current and new test technologies—annual use and costs. 98
Table 58 Body-of-evidence matrix for accuracy of ThinPrep Imager vs conventional cytology 102
Table 59 Electronic databases and websites searched. 115
Table 60 Management recommendations for 'exception' cases under the 2005 NHMRC Guidelines—analysis of VCCR data. 143
Table 61 Model values for rates of unsatisfactory cytology, by test technology. 145
Table 62 Modelled test characteristics—colposcopy 146
Table 63 Probability that colposcopy will be unsatisfactory—modelled values 146
Table 64 Post-treatment HPV status by PCR at 12 months. 147
Table 65 Post-treatment CIN 2/3 detection. 147
Table 66 Proportion of detected lesions that are CIN 2 or 3. 147
Table 67 Baseline matrix for post-treatment recurrence 148
Table 68 Matrix for post-treatment recurrence (lower estimate for relative risk). 148
Table 69 Matrix for post-treatment recurrence (upper estimate for relative risk). 148
Table 70 Definition of model health states 153
Table 71 Correlation between model health states and histology categories in the VCCR Statistical Report, 2006, Table 5.1 155
Table 72 Sensitivity and specificity estimates at cytological thresholds of ASCUS, LSIL and HSIL (TBS1994 terminology). 156

Table 73 Sensitivity and specificity of LBC, relative to CC, for detection of CIN 2+ at cytological thresholds of ASCUS, LSIL and HSIL (TBS1994 terminology)..... 156
Table 74 Cytology report distributions, adjusted to exclude unsatisfactory smears..... 157
Table 75 Comparison of the Australian Modified Bethesda System (AMBS 2004) with previous Australian terminology and the Bethesda System (TBS2001)..... 158
Table 76 Approximations for model health states..... 158
Table 77 Data sources for various methods of deriving test probability matrices for CC or LBC..... 160
Table 78 Relative cytology test yields for automated LBC and LBC compared with CC..... 163
Table 79 Estimated test yields for automated LBC and LBC in women aged 25–29 years, baseline values..... 163
Table 80 Partial test probability matrix derived from Arbyn et al.'s meta-analysis and additional assumptions (%)..... 164
Table 81 Implied underlying disease distribution by various derivation methods..... 165
Table 82 Absolute and relative sensitivity and specificity for detection of CIN 2+ for cytology test characteristics based on Method 3..... 167

Figures

Figure 1 The natural history of HPV infection and cervical cancer. [a]..... 6
Figure 2 Summary of the process used to identify and select studies for the review..... 20
Figure 3 Relative sensitivity (A) and specificity (B) at test threshold of HSIL+ from Arbyn et al. (2008)..... 29
Figure 4 Relationship between cytologist reading speed and difference between conventional cytology and TPI-assisted LBC slide reading time..... 48
Figure 5 Linked models used to perform economic evaluation..... 66
Figure 6 Cumulative proportion of women ever screened by age..... 68
Figure 7 Predicted age-specific prevalence of oncogenic HPV infection, compared to a 2006 meta-analysis of international data (with 95% CI) (data from Smith et al. 2008)..... 83
Figure 8 Predicted age-specific incidence of cervical cancer in an unscreened population, compared to data from 25 developing countries (data from Parkin et al. 2002)..... 83
Figure 9 Predicted age-specific incidence of cervical cancer in Australia, compared with cancer registry data from 2002–2004..... 84
Figure 10 Predicted age-specific mortality in Australia, compared with cancer registry data from 2003 to 2005..... 85

Figure 11 Predicted age-specific rate of histologically confirmed high grades detected, compared with registry data (2006). 85
Figure 12 Predicted age-specific rate of histologically confirmed low grades detected, compared with registry data (2006). 86
Figure 13 Sensitivity analysis for cost-effectiveness ratios of automated LBC. 95
Figure 14 Sensitivity analysis for cost-effectiveness ratios of manual LBC (assuming $2.40 incremental cost for LBC)..... 96
Figure 15 Clinical flowchart of cervical cytology testing **Error! Bookmark not defined.**
Figure 16 Management of participants testing positive in screening program..... 137
Figure 17 Modelled cumulative proportion of women ever screened by age. 142
Figure 18 Management of a cytological prediction of possible or definite LSIL..... 149
Figure 19 Management of women with surgical cytology predicting possible high-grade squamous lesions..... 151
Figure 20 Age-specific test accuracy (sensitivity and specificity) of HPV triage for detection of CIN 2+ in women with pLSIL or dLSIL cytology..... 161
Figure 21 Age-specific percentages of women testing positive in HPV triage for pLSIL and dLSIL cytology. 162
Figure 22 Test characteristics associated with cytology in different populations (screening and colposcopy-referred populations). 166
Figure 23 Summary ROCs obtained via Method 2, compared to that described in Arbyn et al. 167
Figure 24 SROC obtained via Method 3, compared to that described in Arbyn et al. 168
Figure 25 SROC for Method 4, compared to that described in Arbyn et al. 169
Figure 26 SROC for Methods 5 & 6 compared to that described in Arbyn et al. 169

Executive summary

The procedure

In Australia, cervical cytology is conventionally undertaken using the Pap smear test. This involves the collection of cells from the uterine cervix. Cells are collected from the cervix using a small cytobrush/broom or spatula and smeared onto a glass slide for examination under the microscope by a cytologist. Cytological abnormalities are classified using the Australian Modified Bethesda System (AMBS) as high-grade squamous intraepithelial lesions (HSIL), possible HSIL (pHSIL), low-grade squamous intraepithelial lesions (LSIL) or possible LSIL (pLSIL).

Liquid-based cytology (LBC) involves collection of cervical cells in a similar way as for conventional Pap, but the head of the brush, broom or spatula is rinsed into a vial of liquid to produce a cell suspension. The cell sample is treated to remove other material, such as blood and mucus, so that a thin layer of cervical cells can be placed on a slide for microscopic examination. Automated cytology refers to the use of a computer imager to scan slides prepared using LBC or conventional techniques. The Two systems of automated LBC slide reading are marketed in Australia, the ThinPrep® Imager [Cytyc Pty Ltd] and the FocalPoint Imaging System [Becton Dickinson Pty Ltd]. These systems are used to direct cytologists to the areas on the slide most likely to contain abnormal cells.

Medical Services Advisory Committee—role and approach

The Medical Services Advisory Committee (MSAC) was established by the Australian Government to strengthen the role of evidence in health financing decisions in Australia. MSAC advises the Australian Government 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 NHMRC Clinical Trials Centre was engaged to conduct a systematic review of the literature and an economic evaluation of automation-assisted and liquid-based cytology for cervical cancer screening. This evaluation was conducted in collaboration with researchers from the NSW Cancer Council and the University of Sydney's Screening and Test Evaluation Program. An advisory panel with expertise in this area then evaluated the evidence presented and provided advice to MSAC.

MSAC's assessment

This report considers evidence published since the previous reviews of liquid-based and automation-assisted cervical cancer screening (Medical Services Advisory Committee 2002b; Medical Services Advisory Committee 2003).

The primary research question for this review was:

- What is the safety, effectiveness and cost-effectiveness of LBC using automated image analysis systems in comparison to manual reading of conventionally prepared Pap smear cytology samples for the screening and diagnosis of cervical cancer?

In addition, the following secondary research questions were addressed:

What is the safety, effectiveness and cost-effectiveness of

- LBC compared to conventionally prepared Pap smear cytology samples when manual reading of slides is used?

- automated image analysis systems in comparison to manual reading of conventionally prepared Pap smear cytology samples?

- LBC using automated image analysis systems compared to manual reading of LBC?

Clinical need

Approximately 1 in 179 Australian women will develop cervical cancer by the age of 85 years (Tracey et al. 2007). In 2005, there were 734 new cases of cervical cancer reported in Australia and 221 deaths (AIHW & AACR 2008).

The National Cervical Screening Program recommends 2-yearly cytology tests for women aged between 18–20 years and 69 years (Australian Government Department of Health and Ageing 2007). Since its introduction, new cases of cervical cancer among women of all ages almost halved from 13.2 new cancers per 100 000 women in 1991 to 6.9 new cancers per 100 000 women in 2005 (AIHW & AACR 2008). Mortality also halved from 4.0 deaths per 100 000 women in 1991 to 2.0 deaths per 100 000 women in 2005. These improvements can be attributed to the cervical screening program as well as to improvements in therapy.

Safety

LBC with manual or automation-assisted slide reading uses the same procedure for collecting cervical cell samples as conventional Pap cytology tests and is considered a safe procedure.

Effectiveness

No studies have assessed the impact of LBC with manual or automated slide reading on the incidence of invasive cervical cancer or consequent mortality rates compared to conventional cytology. The present review therefore relies on evidence about the relative accuracy of manual or automated LBC for detecting precancerous cervical lesions to draw conclusions about its relative effectiveness. This 'linked evidence' approach is justified by evidence that early detection and treatment of precancerous cervical lesions reduces the incidence of cervical cancer and consequent mortality.

Liquid-based cytology

The most recent systematic review and meta-analysis of comparative accuracy (Arbyn et al. 2008) demonstrated that LBC compared to conventional cytology has no statistically significant increase in sensitivity to detect CIN 2+ (LBC : conventional sensitivity ratio

HSIL+ threshold 1.05, 95% CI 0.95–1.16; LSIL+ 1.03, 0.96–1.11; pLSIL+ 1.03, 0.97–1.09), no statistically significant difference in specificity to detect CIN 2+ at a test threshold of HSIL or LSIL (LBC : conventional specificity ratio HSIL+ 0.99, 95%CI 0.98–1.01; LSIL+ 0.97 95%CI 0.94–1.01) and a lower specificity to detect CIN 2+ at a test threshold of pLSIL (ASCUS**+**, LBC : conventional specificity ratio 0.91, 95% CI 0.84–0.98)

The most recent HTA and Bayesian meta-analysis of comparative unsatisfactory rates (Krahn et al. 2008) found significant between study heterogeneity with

- pooled unsatisfactory rates of 2.24% (95%CI 1.20 to 3.29%) for a filtration-based LBC method and 3.04% (1.92% to 4.16%) for conventional slides, with a difference of –0.81 (–1.87 to 0.24 %) in 44 studies

- pooled unsatisfactory rates of 0.82% (95%CI 0.14 to 1.51%) for a centrifugation-based LBC method and 3.31% (0.97% to 5.67%) for conventional slides, with a difference of –2.49 (–4.43 to –0.55%) in 15 studies

- subgroup analyses by study quality, design or population were not conducted.

A recent large RCT of 45 174 women reported that LBC decreased the unsatisfactory rate (2.6% filtration-based LBC vs 4.1% conventional cytology, relative frequency 0.62, 95% CI 0.56–0.69).

A recent HTA and meta-analysis (Krahn et al. 2008) demonstrated that LBC compared to conventional cytology classified significantly more slides as LSIL and did not classify a significantly different proportion of slides as HSIL+.

Data from two systematic reviews indicated in subgroup analyses that there may be variations in accuracy and unsatisfactory rates according to proprietary name.

Automated slide reading

FocalPoint

No eligible studies of the accuracy of the FocalPoint system for reading LBC slides compared to manual reading of conventional slides were identified (primary research question). No evidence for an accuracy advantage, disadvantage or equivalence of the AutoPap system was found in two studies comparing AutoPap-assisted reading of conventional slides to manual reading of conventional slides or in a single study of highly limited applicability comparing AutoPap-assisted reading of LBC to manual reading of LBC (secondary research questions). The AutoPap-assisted reading of conventional slides reduced unsatisfactory rates (one study) compared to manual reading. However, process advantages are not considered relevant if sufficient evidence of accuracy is not available.

ThinPrep Imager

Two eligible studies of the ThinPrep Imager system addressed the primary research question. One fair-quality Australian study (Davey et al. 2007a) of the ThinPrep Imager system compared to manual reading of conventional cytology found a significant increase in the detection of CIN 2+ lesions (pHSIL threshold; additional 0.82 cases per 1000

women screened) and no significant increase in the number of false positive biopsy results (at a pHSIL threshold).

A second fair-quality Australian study (Roberts et al. 2007) of the ThinPrep Imager system, with a higher possibility of verification bias, found a non-significant increase in the detection of high-grade lesions (pHSIL threshold, 1.49 additional cases per 1000 women screened); a significant increase in the number of false positive biopsy results (pHSIL threshold, 1.93 additional cases per 1000); and a significant increase in the detection of high-grade lesions on manual LBC compared to conventional cytology (pHSIL threshold, 2.8 additional cases per 1000)

Three studies comparing ThinPrep Imager reading of LBC slides to manual reading of LBC slides (secondary research question) found no significant difference in the number of high-grade cases detected and significantly fewer false positive cases for high-grade lesions at a test threshold of pHSIL or HSIL, in two of three studies, respectively. There is a possibility of verification bias in one of these studies.

One Australian study demonstrated a significant decrease in classification of unsatisfactory slides by the ThinPrep Imager system compared to conventional cytology (1.8% vs 3.1%; $P < 0.001$).

Two Australian studies found the ThinPrep Imager compared to conventional cytology decreased slide reading time (mean difference 7.18 slides per hour [95%CI 6.17–8.20]; $P < 0.001$, Davey et al. 2007b) and significantly increased the number of slides classified with low-grade abnormalities.

It is unclear whether any increase in detection of high-grade lesions by the ThinPrep Imager system is attributable to LBC alone, to the automation-assisted reading system, or a combination of both.

Economic considerations

A modelled analysis of cervical cancer screening, diagnosis and treatment is necessary to explore the potential long-term benefits and trade-offs of using these technologies in the Australian screening program.

Published data were used to construct a model of the natural history of human papillomavirus (HPV) and cervical cancer screening in Australia. This model was used to evaluate the cost-effectiveness of LBC with manual or automation-assisted slide reading.

Modelled analysis predicts that LBC with manual reading detects an additional 314 histologically confirmed high-grade lesions and would prevent 23 cervical cancer cases and 6 deaths due to cervical cancer annually. This would be accompanied by additional investigations and treatments, including 22 763 smear tests, 6770 colpscopies, 3273 biopsies and 735 treatments for CIN 2/3 per year. The modelled analysis also predicts that automated LBC would detect an additional 1086 histologically confirmed high-grade lesions and prevent 68 cervical cancer cases and 19 deaths due to cervical cancer annually. Additional investigations and treatments would include 38 346 additional smear tests, 10 788 additional colpscopies, 5154 additional biopsies and 1751 additional treatments for CIN 2/3 per year. These predictions are made under favourable assumptions of test accuracy. Whether this outcome is considered as an improvement in overall health outcomes depends on the trade-off between the benefit from a reduction

in cancer incidence and the harms associated with increased investigations among women who are not destined to develop cancer.

Cost-effectiveness ratios were estimated from lifetime costs and LYS, both of which were discounted at 5 per cent per annum. Automated LBC (ThinPrep Imager) was associated with a cost of $194 835 per LYS. The cost associated with manual LBC varied depending on the level of reimbursement, but ranged from $126 315 per LYS ($2.40 incremental cost relative to current Medicare Benefits Schedule [MBS] reimbursement for conventional cytology) to $385 982 per LYS ($10.90 incremental cost).

The findings are sensitive to assumed relative test accuracy, differences in the unsatisfactory smear rate, assumptions about disease natural history (particularly for high-grade regression and progression) and the recommended screening interval. Favourable assumptions were made about the accuracy of the new technologies. On this basis, both technologies would result in an improvement in LYS, but this would come at a substantially higher cost, due mainly to direct cytology test costs, but also to follow-up costs for an increased number of test positives.

Net annual costs for manual LBC screening (including management and follow-up) are estimated to range from $173.4 million (when reimbursed at an incremental cost of $2.40) to $189.7 million (at $10.90). This represents an annual increase of $7.3 million to $23.6 million (or 4%–14%). Net annual costs for automated LBC are estimated as $203.5 million, which represents an annual incremental cost of $37.4 million (or 22.5%).

Cost-effectiveness ratios were calculated for both manual and automated LBC. These are high for both test technologies, and would appear to be unfavourable in the Australian context. The results presented are based on the current screening program in Australia without taking into account potential changes resulting from HPV vaccination. The findings may be different for different screening populations, for different screening programs, or once the anticipated impact of vaccination on the incidence of cervical neoplasia and precursor lesions occurs. The current Australian program of biennial screening for women aged 18–69 is more intensive than many programs internationally. As more tests are performed at the primary screening level in Australia, annual incremental costs associated with the new technologies are higher. If changes to the Australian screening program are considered in the future, and as changes due to vaccination are realised, reassessment of the cost-effectiveness of these technologies, using similar methods, would be warranted as part of any review of screening strategies and technologies.

Other relevant considerations

The collection of cervical cytology samples into an LBC medium provides the opportunity for reflex testing of a range of pathogens, including HPV, *Chlamydia trachomatis* and *Neisseria gonorrhoeae*.

There is an increasing shortage of trained cytotechnologists in Australia. Technologies which decrease cytology slide screening time and increase productivity may aid in addressing workforce shortages by decreasing staff requirements. With the recent introduction of the HPV vaccine in Australia, the expected impact is a decrease in the prevalence of HPV and pre-cancerous cytological abnormalities and also alteration of the distribution of cytological abnormalities, increasing technical difficulties for cytotechnologists manually screening slides even further.

Conclusions

Liquid-based cytology

LBC compared to conventional cytology

- is safe
- provides no statistically significant increase in sensitivity or specificity
- provides no statistically significant difference in sensitivity (HSIL, LSIL or pLSIL thresholds) or specificity (HSIL or LSIL thresholds) for the detection of CIN 2+
- reduces the specificity for the detection of CIN 2+ at a test threshold of pLSIL
- classifies more slides as positive for low-grade lesions
- reduces the rate of unsatisfactory smears
- has a high cost-effectiveness ratio which appears to be unfavourable in the current Australian setting

Automated slide reading

Automation-assisted reading of LBC slides with the ThinPrepImager system compared to manual reading of conventional cytology

- is safe
- detects at least as many CIN 2+ lesions, and may detect more
- increases the number of slides classified as having low-grade lesions on cytology
- reduces the rate of unsatisfactory slides
- reduces slide processing time
- has a high cost-effectiveness ratio which appears to be unfavourable in the current Australian setting

Introduction

The Medical Services Advisory Committee (MSAC) has reviewed the use of automation-assisted and liquid-based cytology, which is a technology for cervical cancer screening. MSAC evaluates new and existing diagnostic 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. MSAC adopts an evidence-based approach to its assessments, based on reviews of the scientific literature and other information sources, including clinical expertise.

MSAC's terms of reference and membership are at Appendix A. 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 evidence for automation-assisted and liquid-based cytology (LBC) for cervical cancer screening published since the previous MSAC assessments in 2002 and 2003 (Medical Services Advisory Committee 2002b; Medical Services Advisory Committee 2003).

Background

Cervical cancer cytology

In Australia, cervical cytology is conventionally undertaken using the Pap smear test. This involves the collection of cells from the uterine cervix by a medical practitioner using a vaginal speculum to visualise the cervix. Cells are collected from the cervix using a small cytobrush/broom or spatula. Cells are smeared onto a glass slide, fixed with a fixative solution, and then transported to the laboratory for examination under the microscope by a cytologist.

Pap tests are a safe and relatively cheap method for detecting pre-cancerous changes of the cervix and Their use in the National Cervical Cancer Screening Program has been associated with a reduction in cervical cancer incidence and mortality (Australian Government Department of Health and Ageing 2007)

Liquid-based cytology

Liquid-based cytology (LBC) uses a different method for preparing cervical cells for cytological examination than the conventional Pap test. Cells are collected from the cervix using a brush, broom or spatula in the same way they are collected for conventional Pap, but the head of the brush or spatula is rinsed into a vial of liquid to produce a cell suspension which is sent to the laboratory. In the direct-to-vial collection method, instead of smearing the cells directly onto a glass slide, All cells collected from the cervical scraping are transferred directly to the LBC preservative fluid. In the split-sample collection method, the collected cervical cells are first smeared onto the glass slide and any residual sample on the spatula is transferred to the LBC fluid. Studies comparing conventional cytology and LBC using the split-sample method may disadvantage LBC as most of the cells are transferred to the conventional slide.

At the laboratory, the cell sample is treated to remove other material, such as blood and mucus, so that a thin layer of cervical cells can be placed on a slide for microscopic examination. This method has been developed to overcome the problem of 'unsatisfactory' Pap tests, where microscopic examination of the cervical cell smear is difficult and sometimes impossible due to the presence of blood or mucus. In these situations, it is expected that LBC would be less time consuming and avoid the need for asking some women to return for a second Pap test. Other advantages of LBC are that the cell sample can be used for adjunctive testing for a range of pathogens including human papilloma virus (HPV), *Chlamydia trachomatis* and *Neisseria gonorrhoeae* and that it may be more suited to automated screening technologies. The main disadvantage of LBC is that it is more costly than Pap tests.

Automated slide reading

Automated cytology refers to the use of a computer imager to scan slides prepared using LBC or conventional techniques. The aim of automated slide reading is to reduce detection error and cytology reading time. A reduction in reading time would enable cytologists to screen more cervical cytology slides in a given time period. Automated cytology can be used to identify areas on the slide that are most likely to contain abnormal cells, deselect slides from manual reading, or as a quality control (QC)

procedure rescreen slides processed manually. Different technologies are designed specifically for use at different points in the slide screening process.

The ThinPrep system

The ThinPrep® System (Cytyc Pty Ltd) is designed for primary automated screening. The ThinPrep System incorporates the ThinPrep Imager, a computerised cytology reading device, and the Thin Prep Pap Test, a liquid-based cytology preparation system. The ThinPrep Imager identifies 22 areas on a slide that are most likely to contain abnormal cells. These areas, commonly referred to as fields of view (FOV), are then manually examined by the cytologist under high-resolution magnification. This system has been studied in a recent large prospective trial conducted in Australia (Davey et al. 2007a).

FocalPoint / AutoPap

The FocalPoint Imaging System (Becton Dickinson Pty Ltd, formerly TriPath Imaging Inc.) is also designed for primary screening of slides. The FocalPoint slide profiler was previously marketed as the AutoPap Primary Screening System. FocalPoint can be used on conventionally prepared or LBC slides. This automated cytology system ranks slides according to the likelihood of abnormality. The AutoPap system has been used to identify slides without abnormality that do not require manual review. Up to 25 per cent of slides are read by the automated system and are archived without cytologist review (Wilbur et al. 2002). The FocalPoint system is currently used to both rank slides and direct cytologists to the 10 FOV most likely to contain abnormal cells. In addition, it identifies at least 15 per cent of all slides for a QC rescreen. Earlier versions of the AutoPap system were used only for QC rescreening. Details of the processes used by the AutoPap system for primary screening or rescreening are provided in the previous MSAC review on computer-assisted image analysis (Medical Services Advisory Committee 2003). The LBC preparation system produced by this manufacturer is SurePath®, previously marketed as AutoCyte Prep.

The National Cervical Screening Program

The National Cervical Screening Program was established in Australia in 1991 to identify and treat women with precancerous cervical intraepithelial neoplasia (CIN) before it progresses to invasive cancer. Cervical cytology tests are recommended every 2 years starting at age 18 to 20 years for asymptomatic, sexually active women (or within 1 to 2 years of their becoming sexually active) and ceasing at age 70 years (Australian Government Department of Health and Ageing 2007, National Cervical Screening Program). If the cytology results are suggestive of pre-cancerous changes, women are referred to a specialist for histological diagnosis and appropriate treatment.

Classification and management of cytological and histological abnormalities

Different systems are used for classifying cytological and histological abnormalities in cervical screening. In Australia, cytological abnormalities are classified by using the Australian Modified Bethesda System (AMBS). Under this system, cytological abnormalities of squamous cells are classified as high-grade squamous intraepithelial lesions (HSIL), possible HSIL (pHSIL), low-grade squamous intraepithelial lesions (LSIL) or possible LSIL (pLSIL) (Table 1). The international literature most commonly uses the US Bethesda System, which uses a slightly different terminology. pLSIL is

equivalent to atypical squamous cells of undetermined significance (ASCUS) under the US Bethesda System (Table 1).

Women with HSIL are referred to a specialist for examination of the cervix by use of a colposcope, in a procedure called colposcopy. Abnormal lesions identified at colposcopy are biopsied and classified as CIN grades 1–3 on the basis of the histological findings (Table 2). Although it was originally believed that neoplastic cellular changes occurred along a continuum from CIN 1 to 2 to 3, CIN 1 is now regarded as a manifestation of the HPV infective process, rather than as the first step in the neoplastic process.

HPV infection of the cervix is usually asymptomatic, and most infections are transient. HPV infection may not cause any change in cell morphology or it may cause the cytopathic effect previously recognised as mild dysplasia and classified as CIN 1. Thus, CIN 1 lesions are now monitored by repeat cytology with the expectation that the cellular changes will regress when HPV infection resolves. In a small proportion of women, persistent HPV infection may occur.

Persistent infection with oncogenic genotypes precedes precancerous changes, which are classified as CIN 2 (moderate dysplasia) or CIN 3 (severe dysplasia). These lesions are treated by ablative therapy to prevent progression to invasive cancer. It is now accepted that CIN 2 or 3 can occur *de novo*, rather than as a continuum from CIN 1 lesions. A trial-based quality control assessment of community pathology biopsy diagnoses has demonstrated that the detection of CIN 2 has poor reproducibility compared to the detection of CIN 3, with 56 per cent of 523 CIN 2 cases reclassified as CIN 3 (27%) or < CIN 2 (29%) at the quality control assessment (Castle et al. 2007). The authors suggested that this evidence indicates that CIN 2 represents a mix of HPV infection and CIN 3, and that CIN 3 is the true precursor to cancer (Castle et al. 2007).

Women with a cytological finding of possible or definite LSIL are managed more conservatively, with cervical cytology repeated at 12 and 24 months and referral for colposcopy only if these lesions are persistent, because the majority represent an infective process due to HPV and will resolve spontaneously without treatment. However, around 20 per cent of LSIL cases will be confirmed as CIN 2–3 at histology if immediate colposcopy and biopsy are performed (pooled prevalence from 10 studies: 18.8% [95% CI 1.24%–25.2%]; Arbyn et al. 2006).

Table 1 Comparison of the Australian Modified Bethesda System, 2004, and the US Bethesda System, 2001.

Australian Modified Bethesda System	US Bethesda System
Squamous abnormalities	
Possible low-grade squamous intraepithelial lesion (pLSIL) Low-grade squamous intraepithelial lesion (LSIL) Possible high-grade squamous intraepithelial lesion (pHSIL) High-grade squamous intraepithelial lesion (HSIL) Squamous cell carcinoma (SCC)	Atypical squamous cells, undetermined significance (ASCUS) Low-grade squamous intraepithelial lesion Atypical squamous cells, possible high-grade lesion (ASC-H) High-grade squamous intraepithelial lesion Squamous cell carcinoma
Glandular abnormality	
Atypical endocervical cells of undetermined significance Atypical glandular cells of undetermined significance (AGUS) Possible high-grade glandular lesion Endocervical adenocarcinoma *in situ* (AIS) Adenocarcinoma	Atypical endocervical cells, undetermined significance Atypical glandular cells of undetermined significance Atypical endocervical cells, possibly neoplastic Endocervical adenocarcinoma *in situ* Adenocarcinoma

Source: Extracted from the NHMRC Screening to Prevent Cervical Cancer Guidelines (NHMRC 2005).

Table 2 Classification of histological abnormalities as grades of cervical intraepithelial neoplasia (CIN).

Grade	Definition
CIN 1	Mild dysplasia involving the basal 1/3 of the epithelium; an infective process
CIN 2	Moderate dysplasia involving the basal 2/3 of the epithelium
CIN 3	Severe dysplasia involving more than 2/3 of the cervical epithelium; also referred to as cervical cancer *in situ*

Intended purpose of LBC

This report assesses the use of LBC with or without automated slide reading as a replacement for conventional Pap cytology tests for the detection of abnormal cervical cytology.

Clinical need for LBC

Natural history of cervical cancer

Invasive cervical cancer arises most commonly in the squamous epithelium of the transformation zone of the cervix (squamous cell carcinomas), or less commonly in the glandular epithelium of the cervix (as adenocarcinoma). Persistent infection with a high-risk subtype of HPV is necessary for the initiation and progression of CIN to invasive cancer, although cervical cancer is still a rare outcome of oncogenic HPV infection (Bosch et al. 2006).

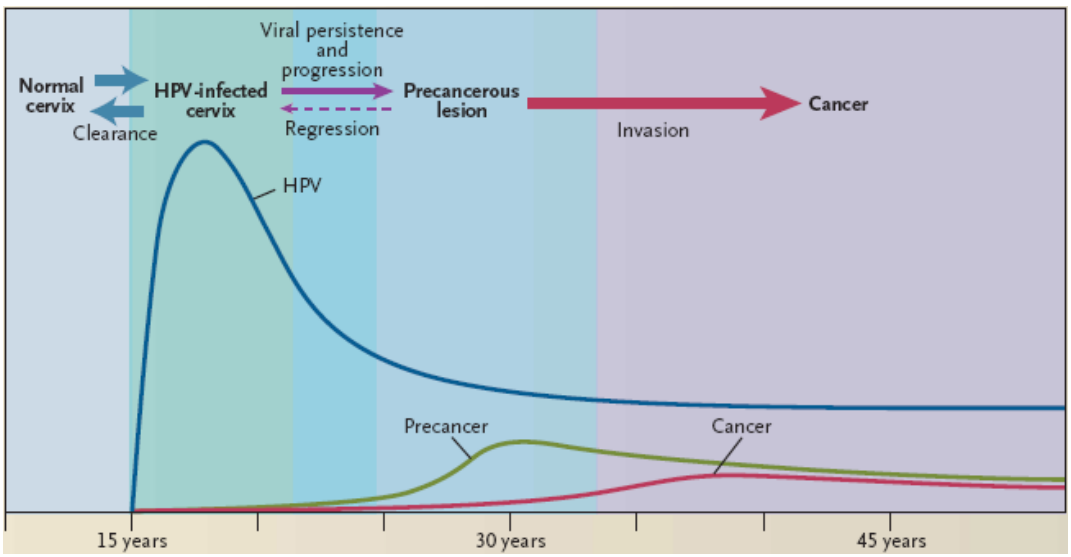
As above, manifestation of HPV infection of the cervix has the appearance of LSIL on cytological examination and CIN 1 (mild dysplasia due to HPV) on histological examination. If the HPV infection resolves, these infective intraepithelial changes regress, but if the infection persists with other mutagenic events, a neoplastic process

may be initiated with cytological findings of HSIL and histological findings of CIN 2-CIN 3. Some of these precancerous lesions may regress spontaneously; but others will progress to invasive cancer if untreated. To date there are no clear predictors of which CIN 3 lesions will progress, although various makers are currently being developed and assessed.

The prevalence of transient HPV infections is highest among young women in the first few years after the initiation of sexual activity, because genital HPV is primarily sexually acquired and readily transmissible. CIN 2 and 3 occur a few years after infection, and precancerous cervical lesions (CIN) peak a decade later (Figure 1). In a small proportion of these women, CIN 3 progresses to invasive cancer over a period of 10 to 20 years, peaking among women aged 40 to 50 years (Figure 1). The cumulative incidence of invasive cancer within 30 years of a biopsy diagnosis of CIN 3 has been estimated at 31 per cent (95% CI 23%–42%) if the lesions are untreated. This estimate is based on a retrospective analysis of cancer incidence from 143 New Zealand women diagnosed with CIN 3 between 1965 and 1974 who did not receive treatment as part of an unethical clinical study (McCredie et al. 2008). Cancer incidence was higher in the subgroup of women with persistent CIN 3 at 2 years (50%, 95% CI 37%–65%), but less than 1 per cent among the 593 women assessed as having received adequate treatment. This study was conducted before the introduction of organised cervical screening, and the average age of participants was 38 years. The results could overestimate the invasion rates associated with small CIN 3 lesions detected in younger women participating in cervical screening programs (Schiffman & Rodriguez 2008).

The success of the National Cervical Screening Program in reducing the incidence of cervical cancer has been attributed to the effectiveness of early detection and appropriate treatment of CIN (see page 9).

Figure 1 The natural history of HPV infection and cervical cancer. [a]



[a]

By permission of Massachusettes Medical Society.

Incidence of cervical cancer, mortality and survival rates

Approximately 1 in 179 Australian women will develop cervical cancer by the age of 85 years (Tracey et al. 2007). The cancer can present at any age after a woman becomes sexually active, but is extremely rare before the age of 20 years (AIHW 2008).

In 2005, there were 734 new cases of cervical cancer reported in Australia and 221 deaths (AIHW & AACR 2008). The age-standardised incidence rate was 6.9 cases per 100 000, and the age-standardised mortality rate was 2.0 deaths per 100 000 women. During the 12 month period 2006–07 there were 1846 hospitalisations with a principal diagnosis of cervical cancer.

Patient prognosis depends on the stage of disease at diagnosis. The Federation Internationale de Gynecologie et d'Obstetrique (FIGO) system is used to classify the extent of disease from Stage I (confined to the cervix) to Stage IV (extension beyond the true pelvis or clinically involving the mucosa of the bladder and/or rectum (Appendix C).

Data from the NSW Cancer Register reported the 5-year survival among women with a diagnosis of cervical cancer from 1999 to 2004 was 73 per cent (Tracey et al. 2007). Australian survival data are reported by extent of disease, not by FIGO stage. Table 3 shows 5-year and 10-year cumulative relative survival by extent of disease at diagnosis and corresponding FIGO stage.

Table 3 Cumulative relative rates of survival of invasive cervical cancer at 5 and 10 years, by extent of disease and corresponding FIGO stage, NSW, 1991–2000.

Extent of disease (FIGO stage) [a]	Cumulative relative survival 5-year	10-year
Localised (IA, IB)	87.5%	85.9%
Regional (IB or IIA & lymph node involvement)	66.7%	57.6%
Locally advanced (IIA, IIB, IIIA, IIIB, IVA)	53.6%	50.0%
Distant (IVB)	30.1%	20.5%

Source: NSW Cancer Council (unpublished data 2008).
[a] This is an approximation; the cancer register classification of extent of disease does not directly correspond to FIGO classification system.

Risk factors for cervical cancer

The addition of persistent HPV infection, other factors associated with an increased risk of cervical cancer are: high parity (large number of children), high number of sexual partners, young age at first sexual intercourse (<18 years), prior history of abnormal cytology tests or cancer of the vagina or vulva, low socio-economic status, and history of smoking (International Collaboration of Epidemiological Studies of Cervical Cancer et al. 2007). Women with an immunodeficiency disorder, for example due to HIV, are also at higher risk of cervical cancer.

High-risk populations—Indigenous women

Indigenous women are at higher risk of cervical cancer than non-Indigenous women, although there has been a substantial reduction in cervical cancer incidence and mortality in both groups following the introduction of the National Cervical Screening Program in 1991 (Cunningham et al. 2008).

High-quality national data on cancer incidence in Indigenous Australians are not available, because the Indigenous status of patients is not always correctly identified at registration. However, cancer registry information on Indigenous status from Queensland, Western Australia, South Australia and the Northern Territory (NT) is believed to be the most complete. Data from these registries indicate that the age-standardised incidence rate of cervical cancer among Indigenous women was at least 12 per 100 000 for the period 1997–2001 (Australian Bureau of Statistics [ABS] & Australian Institute of Health and Welfare [AIHW] 2005), compared with overall annual age-standardised incidence rates of between 7.4 and 9.1 per 100 000 among Australian women over the same period (AIHW 2008). In the NT, one study identified 56 cases of cervical cancer in Indigenous women from 1991 to 2005, indicating 2.9× the risk seen in non-Indigenous women (Cunningham et al. 2008).

Mortality data from Queensland, Western Australia, South Australia and the NT show that mortality rates from cervical cancer among Indigenous women aged 20–69 years in the period 2001–2004 were almost 5× the rates among non-Indigenous women (Indigenous women 9.9 per 100 000, 95% CI 6.0%–15.3%; non-Indigenous women 2.1 per 100 000, 95% CI 1.9%–2.5% (AIHW 2007b). Five-year survival rates among Indigenous women between 1991 and 2001 have been estimated at 37 per cent, compared with 79 per cent among all Australian women (Cunningham et al. 2008).

Differences in cervical cancer incidence between Indigenous and non-Indigenous women have been attributed to a lower participation rate in cervical screening programs among Indigenous women. Indigenous participation rates vary among regions; some regions report higher participation rates than the national average, but in general, Indigenous participation rates are lower than the national average (Cunningham et al. 2008). For example, participation rates of 44 per cent have been reported in Indigenous women in the NT (Cunningham et al. 2008), compared with a national average of 61 per cent (AIHW & AACR 2008). Follow-up rates among Indigenous women identified with abnormal screening tests are also lower; one study from the NT reported that 84 per cent of women with HSIL received appropriate investigation within 6 months (Cunningham et al. 2008).

Impact of the National Cervical Screening Program

In the 2-year period 2005-2006, 3305 978 women participated in the National Cervical Screening Program, of whom 98.5 per cent were aged between 20 and 69 years. These figures indicate a participation rate of 61 per cent for women in the target age range during this period (AIHW 2008). In 2006, the screening program detected 29 532 histologically verified cervical abnormalities, of which 15 118 were low-grade and 14 414 were high grade (CIN 2+), representing 0.8 per cent of the screened population (AIHW 2008).

Since its introduction, new cases of cervical cancer among women of all ages almost halved from 13.2 new cancers per 100 000 women in 1991 to 6.9 new cancers per 100 000 women in 2005. Mortality also halved from 4.0 deaths per 100 000 women in 1991 to 1.9 deaths per 100 000 women in 2006 (AIHW & AACR 2008). These improvements can be attributed to the cervical screening program as well as improvements in therapy. Substantial benefits have also been reported among Indigenous women during this period, with a 68 per cent reduction in incidence from 1991 to 2005 and 92 per cent reduction in mortality from 1991 to 2003 in the NT (Cunningham et al. 2008).

Psychological effects of abnormal cervical cytology findings

Two systematic reviews of studies investigating the impact of abnormal cytology and colposcopy have documented the negative psychological effects of receiving an abnormal cytology test, including anxiety, fears of cancer, infertility, depression, difficulties with sexual relationships and self blame (Herzog & Wright 2007; Rogstad 2002). A recent study of 3731 women aged 20–59 years who participated in the Trial of Management of Borderline and Other Low-Grade Abnormal smears (TOMBOLA) observed that 23 per cent of women with low-grade cytological abnormalities scored at levels that indicated probable clinically significant anxiety on the Hospital Anxiety and Depression Scale (Gray et al. 2006). The authors reported that these findings were similar to earlier findings among women with high-grade cytological abnormalities.

Potential utilisation

Utilisation of the cervical screening program was estimated at approximately 2.1 million smears in 2007, based on national laboratory data (Royal College of Pathologists Australasia Cytopathology Quality Assurance Programs 2008). This figure reflects a participation rate of 61 per cent of women in the target age range, based on consistent figures from 1999 to 2006 (AIHW 2008), and therefore may increase if participation increases.

Modelling predicts that under the 2005 National Health and Medical Research Council (NHMRC) guidelines, incorporating cytology and HPV DNA testing as a component of the test of cure process and cytological follow-up of low-grade abnormalities, the utilisation of LBC or automated LBC respectively would be approximately 1.9 million smears per annum (1 924 675 and 1 940 259, respectively). The number of cytology tests needed will also vary if there are further changes to the current Australian screening guidelines, or if participation patterns vary.

Current treatment of cervical cancer and CIN

Treatment of invasive cancer is determined by the stage of disease: Stage IA disease is treated by cone biopsy or simple hysterectomy, and Stage IB by radical hysterectomy and pelvic lymphadenectomy (AIHW 2007b). More advanced disease (Stage II and higher) is treated by chemo-radiation. Prognosis is less favourable for more advanced disease (see Table 3), and persistent disease will ultimately result in death.

Treatment for precancerous changes (CIN 2–3) involves local excision or ablation of the lesions with laser vaporisation, loop excision, cryosurgery, electrodiathermy or cone biopsy (NHMRC 2005). These procedures (with the exception of CO_2 laser ablation) have been associated with an increased risk of future preterm delivery and low-birth-weight infants and therefore should be recommended with appropriate caution (Kyrgiou et al. 2006). Hysterectomy for treatment of CIN may be indicated in rare cases.

Existing procedures

Cervical cytology

The Papanicolaou test

In Australia, cervical cytology is conventionally undertaken using the Pap test. This involves the collection of cells from the uterine cervix by a medical practitioner using a

vaginal speculum to visualise the cervix. Cells are collected from the cervix using a small brush or spatula. Cells are smeared onto a glass slide, fixed with a fixative solution, then transported to the laboratory for examination under the microscope by a cytologist.

Pap tests are a safe and relatively cheap method for detecting pre-cancerous changes of the cervix and their use in the National Cervical Cancer Screening Program has been associated with a reduction in cervical cancer incidence and mortality (Australian Government Department of Health and Ageing 2007. The main disadvantage of the Pap test is that it is not a highly accurate test. Pre-cancerous and cancerous changes may be missed due to sampling or laboratory error. However, due to the slow progression of disease, the consequences of these errors can be minimised if detected at repeat testing under the current guidelines for routine screening tests and follow-up of unsatisfactory smears and low-grade abnormalities. This strategy relies on patient adherence for follow-up testing and is therefore is likely to be less effective for women with low adherence.

Colposcopy

Colposcopic examination is performed after the insertion of a vaginal speculum. It allows a magnified inspection of the cervix and vagina to guide biopsy of the most abnormal areas for histological diagnosis. The procedure is usually performed by a gynaecologist. The procedure can be undertaken in about 10 minutes.

Colposcopic examination is more accurate than cervical cytology although false negatives may still occur due to failure to visualise abnormal lesions (Schiffman & Solomon 2003). The main disadvantage of colposcopy is that it is an expensive test and thus not suitable for population screening. It also causes minimal to moderate discomfort.

New and emerging cervical cancer technologies

HPV vaccination

In May 2007, the Australian Government introduced the National HPV Vaccination Program to vaccinate school girls aged 12 to 13 years against HPV. In the first 2½ years of the program, all school girls aged 12 to 18 years will be offered the vaccine, and women aged under 27 years will also be able to receive the vaccine free of charge from their GP or community immunisation clinic for 2½ from July 2007 to June 2009 (Australian Government Department of Health and Ageing 2007).

There are 2 HPV vaccines registered for use in Australia: *Gardasil*, which protects against HPV types 6, 11, 16 and 18 and is currently used in the National Vaccine Program; and *Cervarix*, which protects against HPV types 16 and 18. Persistent infection with HPV types 16 and 18 is a prerequisite for development of 7 of 10 cervical cancers, and infection with types 6 and 11 causes around 90 per cent of genital warts. The *Gardasil* vaccine is given as three doses at 0, 2 and 4–6 months. Seroconversion occurs in 99.5 per cent of women who receive the complete course. Randomised controlled trials (RCTs) have demonstrated that vaccination of HPV-naïve women prevents approximately 90 to 100 per cent of persistent infection with these virus types and of precancerous cervical lesions (Australian Government Department of Health and Ageing & National Health and Medical Research Council 2008).

The vaccine is expected to have a rapid and substantial impact on HPV and genital wart incidence and prevalence in young women (Regan et al. 2007; Smith et al. 2008). However, long-term trial results are awaited to provide evidence about duration of

protection, the implications of infection, the potential for cross-protection from non-vaccine subtypes, and the magnitude of reduction in cervical cancer incidence. The vaccine impact will also depend on the percentage uptake among young women.

Vaccination will not replace the need for routine cervical screening; however, the current screening protocol may be modified to take into account changes in population risk as a result of the vaccination program.

Hybrid Capture II (HCII) human papilloma virus (HPV) test

The HCII HPV DNA test is a nucleic acid hybridisation assay that detects HPV subtypes known to be associated with cervical cancer. It involves the collection of cells from the uterine cervix during a vaginal speculum examination by a medical practitioner, in the same way a conventional Pap cytology test is performed. The cell sample is then transported to a pathology laboratory for testing.

The test is available as a standardised test kit which includes a cervical brush, a vial with a specimen transport medium and the solution hybridization assay. It uses a combination of ribonucleic acid (RNA) probes to identify DNA from 13 high-risk HPV subtypes. HPV DNA levels of 1pg/mL are classified as positive for HPV. A positive result indicates infection with one (or more) of high-risk subtypes but cannot be used to identify which subtype is involved.

Potential impact of LBC

Potential benefits

Replacement of conventional Pap cytology tests with LBC/ automated reading of LBC cytology slides may improve the effectiveness and cost-effectiveness of cervical cancer screening through:

- earlier detection and management of women with true precancerous cervical lesions, thereby reducing cervical cancer incidence and mortality, if the test is found to be more sensitive than conventional Pap cytology

- avoidance of unnecessary follow-up cervical cytology tests and colposcopy in women without true precancerous cervical lesions, if the test is found to be more specific than conventional Pap cytology

- avoidance of unnecessary recall for additional cervical cytology tests in women, if the test is found to result in a lower rate of unsatisfactory findings

- increased efficiency of cervical cancer screening programs through increased throughput of cytology laboratories, if the test is found to reduce slide reading times. This increased efficiency would not, however, improve cost-effectiveness unless it is reflected in the reimbursement level of the test.

In addition, the use of LBC can allow the cervical cell sample to be used for adjunctive testing for a range of infectious diseases, including HPV.

Potential disadvantages

Replacement of conventional Pap cytology tests with LBC/ automated reading of LBC cytology slides may provide disadvantages to cervical cancer screening through:

- increased number of unnecessary biopsies and follow-up smears, and associated negative psychological consequences for women without true pre-cancerous cervical lesions, if the test is found to be less specific than conventional Pap cytology.

Reference standard

For studies to provide adequate information about test accuracy, a reference standard should be applied as a minimum to positive or discordant tests. Colposcopy with biopsy (threshold for positive histology CIN 2+/CIN 3+) was considered the most valid reference standard to determine the true disease status of patients with a positive test (pLSIL, dLSIL, pHSIL, HSIL or SCC). Clinical follow-up with repeat cytology at 1 year was considered the most valid reference standard for normal test results. Adjudicated cytology was considered a suboptimal reference standard for patients with either test result.

Comparator

The comparator for LBC or automation-assisted screening of cervical cytology slides is manual screening of conventional Pap smear cytology.

Methodological issues

Where comparative accuracy of two alternative replacement test strategies is required, studies can either be performed with both tests conducted in all patients, or by randomising patients to either test (Bossuyt et al. 2006). Conducting both tests in all patients (a paired design) has advantages as the patient population undergoing each test is identical. In comparing the accuracy of conventional Pap smears and LBC, the 'split-sample' technique is frequently used to perform both tests in the same patients. This approach is believed to disadvantage the technical performance of LBC as more cells are applied to the smear sample than rinsed into the LBC transport medium vial. Therefore, for the current review, randomised studies are believed to provide a more valid estimate of comparative accuracy than paired studies.

Marketing status of the technology

LBC tests with manual or automated slide reading are *in vitro* diagnostic tests that are not of human origin and are therefore exempt from the regulatory requirements of the Therapeutic Goods Act 1989. These tests comply with US FDA requirements in the United States and have received CE marking in Europe.

There are many currently marketed LBC preparation systems available. These systems use a variety of technical methods for storing and preparing the cervical cytology sample, some of which are patented. The ThinPrep Pap system (Cytyc Pty Ptd) includes a filtration method with membrane transfer to the slide. The SurePath and PrepStain system (Becton Dickinson Pty Ltd) uses a centrifugation system. The manufacturers

assert that these technical differences translate into different performance characteristics for different LBC preparation systems.

The only currently marketed fully integrated systems of LBC and automation-assisted screening are the Thin Prep® Pap Test and Imaging System (Cytyc Pty Ltd), and FocalPoint® automated screening of SurePath cytology (Becton Dickinson Pty Ltd, formerly TriPath Imaging Inc). The FocalPoint system can be used for automated screening of conventional or LBC slides. FocalPoint was previously marketed as AutoPap, an automated screening system for conventional Pap smears. Previously marketed systems used for automated screening for quality control include AutoCyte Screen and PapNet. These systems are no longer commercially available.

Current reimbursement arrangement

LBC and automated screening of cervical cytology are not listed on the Medicare Benefits Schedule (MBS).

LBC is currently provided by a number of pathology laboratories for a fee additional to the Medicare rebate, and is collected by the split-sample technique in conjunction with conventional Pap smears.

Two current MBS items relate to cervical cancer screening:

MBS 73053 (Fee = $19.60) Cytology of a smear from cervix where the smear is prepared by direct application of the specimen to a slide, excluding the use of liquid-based slide preparation techniques, and the stained smear is microscopically examined by or on behalf of a pathologist - each examination

> (a) for the detection of precancerous or cancerous changes in women with no symptoms, signs or recent history suggestive of cervical neoplasia, or
>
> (b) if a further specimen is taken due to an unsatisfactory smear taken for the purposes of paragraph (a); or
>
> (c) if there is inadequate information provided to use item 73055;

MBS 73055 (Fee $19.60) Cytology of a smear from cervix, not associated with item 73053, where the smear is prepared by direct application of the specimen to a slide, excluding the use of liquid-based slide preparation techniques, and the stained smear is microscopically examined by or on behalf of a pathologist - each test

> (a) for the management of previously detected abnormalities including precancerous or cancerous conditions; or
>
> (b) for the investigation of women with symptoms, signs or recent history suggestive of cervical neoplasia;

MSAC 2002 and 2003 reviews

A summary of the previous reviews is presented in Appendix D. MSAC assessed LBC as a replacement for Pap tests in 2002 and found insufficient evidence to conclude that it

was more effective than Pap tests (Medical Services Advisory Committee 2002b; reference 12a).

In 2003, MSAC reviewed the evidence for computer-assisted image analysis for cervical screening cytology in the context of primary screening, rescreening and triage (reference 12c). The review found that there was a lack of evidence that computer-assisted image analysis was as effective as conventional manual screening. MSAC concluded that there was insufficient evidence to draw conclusions on the appropriate use of computer-assisted image analysis, and there was no change to funding arrangements.

Approach to assessment

Research question

Evaluators from the NHMRC Clinical Trials Centre, in collaboration with researchers from Cancer Council NSW and the Screening and Test Evaluation Program at the University of Sydney, worked with members of the Advisory Panel to develop research questions to assess the value of LBC and automated systems for slide reading in cervical cancer screening. These questions were formulated *a priori* from information provided in the applications and by the Advisory Panel using flow charts to depict current guidelines for management of screen-detected cervical cancer abnormalities (**Error! Reference source not found.**, page **Error! Bookmark not defined.**).

Primary research question

What is the safety, effectiveness and cost-effectiveness of liquid-based cytology using automated image analysis systems in comparison to manual reading of conventionally prepared Pap smear cytology samples for the screening and diagnosis of cervical cancer?

Secondary research questions

What is the safety, effectiveness and cost-effectiveness of liquid-based cytology compared to conventionally prepared Pap smear cytology samples when manual reading of slides is used?

What is the safety, effectiveness and cost-effectiveness of automated image analysis systems in comparison to manual reading of conventionally prepared Pap smear cytology samples?

Additional research question

The following question was added to the assessment following the review of the results from the initial primary and secondary research questions.

What is the safety, effectiveness and cost-effectiveness of LBC using automated image analysis systems compared to manual reading of LBC?

Assessment strategy

The evaluation team conducted systematic reviews of the medical literature to address these review questions. In the absence of RCTs to directly assess the impact of these technologies on patient outcomes, this review included studies that compared LBC or automation-assisted reading of slides to conventional Pap smear cytology with manual reading and reported on one or more of the following outcomes:

- Diagnostic sensitivity and specificity, or the ratio of true positive (TP) to false positive (FP) findings, or the incremental rate of TPs, for detection of precancerous high-grade cervical lesions (CIN 2+, CIN 3+, AIS [adenocarcinoma *in situ*]) in women with a possible or definite HSIL cytology result

- Changes in management

- Patient outcomes:
  - quality of life—patient preference, satisfaction, psychological distress or anxiety
  - patient compliance
  - safety—adverse events, avoidance of unnecessary treatments
  - incidence of cervical cancer
  - overall survival
  - cervical cancer–specific mortality.

Additional outcomes extracted from studies that reported one or more of the above outcomes included unsatisfactory rates, process outcomes such as analysis time, and accuracy at a test threshold of possible or definitive LSIL or HSIL.

A comprehensive model of screening, diagnosis and treatment of cervical cancer was also developed to estimate the relative effectiveness and cost-effectiveness of incorporating these technologies into the screening program, based on the test characteristics determined by the results of this literature review (page 62).

Review of the literature

A review of existing systematic reviews and HTA reports from 2002 to August 2008 was conducted to identify the most recent and comprehensive systematic reviews published since:

- the August 2002 MSAC review (Reference 12a), 'Liquid-based cytology for cervical screening', which included a systematic review of evidence to March–April 2002

- the May 2003 MSAC review (Reference 12c), 'Computer-assisted image analysis for cervical screening', which considered all automated image analysis systems available at the time in a systematic review to September 2002.

In addition, HTA agency websites and trials registries were searched to November 2007. The databases and websites searched are listed in Appendix E.

Search strategy

Automated slide reading

A systematic review of primary studies from January 2002 to 6 February 2008 was conducted, as no recent and comprehensive English-language systematic reviews of automated slide reading systems were identified. The databases listed in Table 4 were searched with the indexing and text terms listed in Table 5.

Table 4 Electronic databases searched to identify primary studies of automated slide reading.

Database	Period covered
EMBASE.com (includes EMBASE & Medline)	2002 to 6 February 2008
PreMEDLINE	As at 6 February 2008

Table 5 Search terms for automated slide reading (EMBASE.com).

Element of clinical question	Suggested search terms
Patient (screening)	Cervical cancer screening 'papanicolaou test'/syn OR ('cancer screening'/syn OR screening: ab,ti) AND ('uterine cervix cancer'/syn OR 'uterine cervix tumor'/syn) OR 'uterine cervix cytology'/syn
Intervention/test	Automated image analysis systems ('automation'/syn OR comput* [extensive search] OR automat* [extensive search] OR 'image processing'/syn OR 'image cytometry'/syn) AND (Screening search string) OR 'thin prep' OR thin*prep : ab,ti,dn OR Focal*point: ab,ti,dn OR ('focal point' AND (Screening search string)) OR 'autopap': ab,ti,dn

Field search terms: ab = abstract, ti = title, dn = device trade name.

LBC

A systematic review of primary studies from Jan 2004 to 6 February 2008 was conducted to identify the most recent systematic reviews. The databases listed in Table 6 were searched with the indexing and text terms listed in Table 7.

Table 6 Electronic databases searched to identify primary studies of LBC.

Database	Period covered
EMBASE.com (includes EMBASE & Medline)	2004 to 6 February 2008
PreMEDLINE	As at 6 February 2008

Table 7 Search terms for LBC (EMBASE.com).

Element of clinical question	Suggested search terms
Population	Cervical cancer screening
AND	('uterine cervix tumor'/syn OR 'uterine cervix cancer'/syn) AND ('cancer screening'/syn OR screening: ab,ti)
Intervention/test	Liquid-based cytology
	'thin prep' OR thin*prep : ab,ti,dn
	OR
	liquid: ab.ti OR fluid: ab.ti
	OR
	cytorich:ab,ti,dn OR autocyte:ab,ti,dn OR ('sure path' OR sure*path : ab,ti,dn)

Field search terms: ab = abstract, ti = title, dn = device trade name.

Selection criteria

After duplicate publications were excluded, citations were appraised by two independent reviewers to determine eligibility using the criteria listed in Table 8. Discrepancies between reviewers about study eligibility were resolved by discussion.

Table 8 Inclusion/exclusion criteria for identification of relevant studies.

Characteristic	Criteria
Publication type	Clinical studies included. Non-systematic reviews, letters, editorials, animal, in-vitro, laboratory studies, conference abstracts and technical reports excluded. Systematic reviews • Excluded: systematic reviews that have been superseded Primary studies Included: primary studies published after the search period of included systematic reviews Accuracy studies excluded if: • patients were selected for inclusion in the study based on their known disease (case-referent, case-control studies) • no direct comparison of automated image analysis of liquid-based cytology vs manual reading of conventional cytology is reported • comparison of independent cohorts of women are compared to no verification of negative test results [b] Studies providing the highest level of evidence available included. Studies of lower level design excluded.
Population	Women undergoing cervical cytology for the detection of cervical cancer or precancerous lesions.
Intervention/test	Automated cervical cytology image analysis for primary screening using LBC • Excluded: systems used for rescreening or QC; systems not commercially available or not marketed in Australia Manual screening of LBC Automated image analysis for primary screening of conventional Pap smear cytology
Comparator	Manual screening of conventional Pap smear cytology Manual screening of LBC a
Outcome	Studies must report on at least one of the following outcomes: • Diagnostic accuracy for CIN 2+ / CIN 3+ / AIS of automated cytology screening vs manual cytology screening: sensitivity and specificity (and/or sufficient data for reconstruction of 2×2 table); or relative TP and FP rate rates or ratio; or incremental TP rate; using appropriate reference standard • Impact of screening results on clinical management (further investigations/treatment avoided, earlier investigation/treatment initiated)

	• Patient quality of life (eg, patient acceptability, satisfaction, psychological distress) • Patient compliance • Incidence of cervical cancer • Overall survival, cervical cancer–specific survival • Analysis time • Costs
Language	Non-English-language articles excluded.

[a] As a comparator for automated cervical cytology image analysis for primary screening using LBC.
[b] Requirement for high-quality study as per Davey et al. (2006); see appendix I.

Search results

Existing HTA reports and systematic reviews

Nine relevant HTAs or systematic reviews published between 2002 and August 2008 were identified (**Error! Reference source not found.**). No English-language HTAs or systematic reviews of automated cytology slide reading systems more recent than the 2003 MSAC review (to September 2002) were identified.

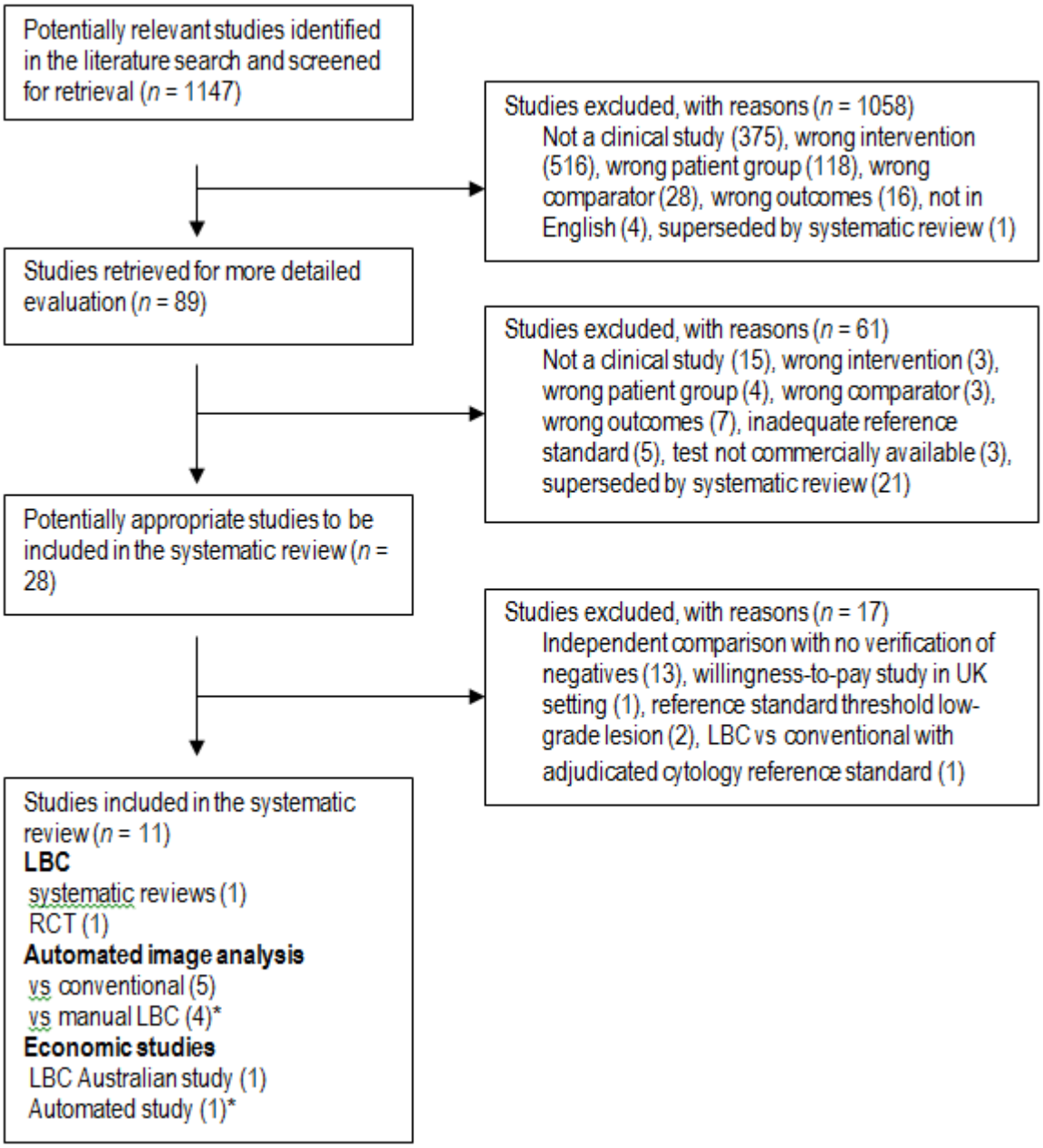
The most recent systematic review of the comparative accuracy of LBC and conventional cytology identified was Arbyn et al. (2008). This high-quality review considered evidence to May 2007. The most recent HTA report considering comparative unsatisfactory rates for LBC and conventional cytology was that by the Canadian Agency for Drugs and Technologies in Health (CADTH; Krahn et al. 2008). This report was based on evidence from a systematic literature search to June 2006.

Quorum flow chart

A total of 1147 non-duplicate citations from the searches for primary studies were screened for inclusion in the LBC or automated screening reviews.

In addition to the recent HTA report from CADTH (Krahn et al. 2008), 11 included studies were identified in the search for primary studies. The process used to identify these studies is described in Figure 2.

Figure 2 Summary of the process used to identify and select studies for the review.



* One study also provides evidence for automated vs conventional.
Adapted from Moher et al. (1999).

Appraisal of the evidence

The strength and relevance of evidence and the size of the clinical effect in individual studies were appraised, followed by the overall body of evidence for conclusions about the safety, effectiveness and cost-effectiveness of the manual and automated LBC tests.

Appraisal of the quality and applicability of individual studies

The quality and applicability of the included studies was assessed according to specified criteria according to the study design.

The quality of studies of diagnostic test accuracy was assessed using a checklist of 12 items adapted from the QUADAS (Quality Assessment of Diagnostic Accuracy Studies) tool developed by Whiting et al. (2003) (Table 9). This tool was developed by experts in the field following a systematic review of the evidence relating to sources of bias and variation relevant to studies of diagnostic test accuracy. Studies were required to meet all 12 criteria to be assessed as high quality (see details in footnote to Table 9), including a valid reference standard for at least all discordant positive slides (Davey et al. 2006). High-quality studies were required to use histology as a reference standard for comparisons of different slide preparation techniques. Adjudicated cytology reading was considered a valid reference standard for paired studies of alternate slide reading methods (Irwig et al. 2004).

Table 9 Criteria used to assess the quality of diagnostic accuracy studies—the QUADAS tool (adapted from Whiting et al. 2003).

Item	
1	Were patients prospectively recruited?
2	Were patients consecutively recruited?
3	Were selection criteria explicitly described?
4	Is the reference standard likely to correctly classify the target condition? Optimal (histology) / valid (consensus cytology) / invalid
5	Did all patients receive verification using a reference standard? All / positive / discordant participants
6	Is the time period between reference standard, comparator and index test short enough to be reasonably sure that the target condition did not change between the tests?
7	Was the test threshold specified?
8	Were test/comparator results interpreted blind to reference standard?
9	Were reference standard results interpreted blind to test/comparator results?
10	Were uninterpretable/intermediate test results reported?
11	Were withdrawals from the study explained?
12	Was sufficient data for determination of relative true and false positive rates reported?

High quality: Yes to 1, 3, 4, 5, 6, 10, 11; other items required to be either Yes or Unclear.
Low quality: No/unclear for 4 or 5; ≤ 4 yes or N/A ratings.
Other studies are assessed as fair quality.

Seven criteria were applied to assess the quality of systematic reviews (Table 10). For the criterion addressing heterogeneity, systematic reviews that did not undertake a meta-analysis were rated 'not applicable' (N/A), unless heterogeneity was specifically mentioned. Studies were required to meet al.l seven criteria to be assessed as high quality. A study with four or fewer 'yes' or 'N/A' ratings was considered to be of low quality.

Table 10 Criteria used to assess the quality of effectiveness studies (adapted from NHMRC 2000 and CRD 2001).

Study design	Quality checklist
Systematic reviews [a]	Was the research question specified?
	Was the search strategy explicit and comprehensive?
	Were the eligibility criteria explicit and appropriate?
	Was a quality assessment of included studies undertaken?
	Were the methods of the study appraisal reproducible?
	Were sources of heterogeneity explored?
	Was a summary of the main results clear and appropriate?

[a] High quality: Yes or N/A to all 7 criteria; low quality: ≤4 Yes or N/A. Other studies assessed as fair quality.

Studies of a population representing the Australian population with a mixture of patients presenting for screening or diagnosis were considered highly applicable. Studies of patients in a high-risk population with seeded cases or where different cytological preparation systems were pooled were considered of limited applicability. Studies in which all patients were undergoing cervical cytology for diagnostic purposes were considered not applicable. Although some studies have been conducted in a setting where the population is likely to be highly applicable, where this was not clearly reported the population will be rated as of limited applicability.

The grading system in Table 11 was used for summarising the quality and applicability of the studies.

Table 11 Grading system for the appraisal of studies evaluating diagnostic tests.

Validity criteria	Description	Grading System
Appropriate comparison	Did the study evaluate a direct comparison of the index test strategy vs the comparator test strategy?	C1 direct comparison CX other comparison
Applicable population	Did the study evaluate the index test in a manner similar to routine use in a population that is undergoing cervical cancer screening for mixed screening and diagnostic purposes?	P1 applicable P2 limited P3 different population
Quality of study	Was the study designed to avoid bias?	Study design: NHMRC level of evidence Study quality:
	High quality	Q1 high quality
	Fair quality	Q2 fair quality
	Poor quality	Q3 poor quality

Ranking the evidence

A structured appraisal of each study was performed to classify studies according to the type of study design (levels of evidence) (Table 12).

Table 12 NHMRC designations of levels of evidence.

Level of evidence	Study design
Studies of effectiveness	
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 pretest/post-test
Studies of test accuracy	
I	A systematic review of level II studies
II	A study of test accuracy with: an independent, blinded comparison with a valid reference standard, among consecutive patients with a defined clinical presentation
III-1	A study of test accuracy with: an independent, blinded comparison with a valid reference standard, among non-consecutive patients with a defined clinical presentation
III-2	A comparison with a reference standard that does not meet the criteria required for Level II and III-1 evidence
III-3	Diagnostic case-control study
IV	Study of diagnostic yield (no reference standard)

Modified from NHMRC (2000; 2008).

Statistical methods

The number of true positive cases detected was compared between alternative testing strategies by χ^2-test.

Estimated differences in sensitivity and specificity are calculated for studies with complete verification of positive findings comparing automated LBC with manual LBC. Whilst the total number of cases is not known, and absolute sensitivity & specificity are not known, the difference can be estimated from the known values as below:

Sensitivity difference = additional TP / (TP + known FN cases)

Specificity difference = fewer FP cases / (concordant negative + FP cases)

Methodological considerations

In studies comparing different methods of reading the same slide, the use of adjudicated cytology as a reference standard should not bias the relative accuracy results (Irwig et al. 2004), as errors in sampling or slide preparation will affect estimates of accuracy for both methods equally. Any differences in accuracy between the methods will relate only to reader interpretation of the slide.

Appraisal of the body of evidence

In addition to the appraisal of individual studies, an appraisal of the overall body of evidence about the safety, effectiveness and cost-effectiveness of the automated LBC tests was conducted using the same principles. This appraisal was based on the assessment of five criteria suggested by the NHMRC guidelines for the developers of guidelines (2008):

1. The volume of evidence—the number of studies sorted by their methodological quality and relevance to patients.

The consistency of the study results—whether the better-quality studies had results of a similar magnitude and in the same direction; ie, homogeneous or heterogeneous findings.

The potential clinical impact—appraisal of the precision, size and clinical importance or relevance of the primary outcomes used to determine the safety and effectiveness of the test.

The generalisability of the evidence to the target population.

The applicability of the evidence—integration of this evidence for conclusions about the net clinical benefit of the index test in the context of Australian clinical practice.

Data extraction and synthesis

Data were extracted by using a standardised instrument designed for this review. Data extraction was performed by one reviewer and checked by a second reviewer. Any discrepancies were resolved by discussion. The data extraction tables are provided in Appendix F. Where the publications reported percentages only, raw numbers were determined from the number of patients on which each test was performed. Where only raw numbers were reported, percentages or rates were calculated from the number of patients reported to have had the test.

Expert advice

An Advisory Panel was established to guide the health technology assessors so as to ensure that the assessment was clinically relevant and took into account consumer interests. Membership of the Advisory Panel is provided at Appendix B.

.

Results of assessment

Is it safe?

Safety

LBC with manual or automated slide reading is considered safe.

LBC with manual or automated slide reading uses the same procedure for collecting cervical cell samples as conventional Pap cytology tests and therefore does not introduce any additional risks to the patient. Collection of cervical cells is regarded as safe. Some women may experience discomfort or minor bleeding afterwards that resolves spontaneously.

Is it effective?

No studies have assessed the impact of LBC with manual or automated slide reading on incidence or mortality rates of invasive cervical cancer compared to conventional cytology. LBC and automated slide reading are tests which will be used to identify patients at the same stage of disease as the current test used in the cervical screening program (conventional Pap smears). The present review therefore relies on evidence about the relative accuracy of manual or automated LBC to detect precancerous cervical lesions to draw conclusions about its relative effectiveness. This 'linked evidence' approach is justified by existing evidence that early detection and treatment of precancerous cervical lesions leads to a reduction in the incidence and mortality of cervical cancer (AIHW 2007a; Peto et al. 2004), See 'Impact of the National Cervical Screening Program' page 9.

Liquid-based cytology

Accuracy

The most recent systematic review and meta-analysis of comparative accuracy (Arbyn et al. 2008) demonstrated that LBC compared to conventional cytology has

- no statistically significant increase in sensitivity to detect CIN 2+ (LBC:conventional sensitivity ratio HSIL threshold 1.05, 95% CI 0.95–1.16; LSIL+ 1.03, 0.96–1.11; pLSIL+ 1.03, 0.97–1.09)
- no statistically significant difference in specificity to detect CIN 2+ at a test threshold of HSIL or LSIL (LBC:conventional specificity ratio HSIL+ 0.99, 95%CI 0.98–1.01; LSIL+ 0.97 95%CI 0.94–1.01)
- lower specificity to detect CIN 2+ at a test threshold of pLSIL (ASCUS+, LBC:conventional specificity ratio 0.91, 95% CI 0.84–0.98)

Unsatisfactory rates

The most recent HTA and Bayesian meta-analysis of comparative unsatisfactory rates (Krahn et al. 2008) found significant between-study heterogeneity, with:

- pooled unsatisfactory rates of 2.24 per cent (95% CI 1.20%–3.29%) for a filtration-based LBC method and 3.04 per cent (1.92%–4.16%) for conventional slides, with a difference of –0.81% (–1.87 to 0.24 %) in 44 studies

- pooled unsatisfactory rates of 0.82 per cent (95% CI 0.14%–1.51%) for a centrifugation-based LBC method and 3.31 per cent (0.97%–5.67%) for conventional slides, with a difference of –2.49 per cent (–4.43% to –0.55%) in 15 studies

- Subgroup analyses by study quality, design or population were not conducted.

A recent large RCT (Ronco et al. 2006) of 45, 174 women reported that LBC

- decreased the unsatisfactory rate (2.6% filtration-based LBC vs 4.1% conventional, relative frequency 0.62, 95% CI 0.56–0.69).

Test yield

A recent HTA and meta-analysis (Krahn et al. 2008) demonstrated that LBC compared to conventional cytology

- classified significantly more slides as LSIL

- did not classify a significantly different proportion of slides as HSIL+

Manufacturer

Data from two systematic reviews indicated in subgroup analyses that there may be variations in accuracy and unsatisfactory rates according to proprietary name.

Included studies

The most recent and comprehensive systematic review of the comparative accuracy of LBC and conventional cytology identified was Arbyn et al. (2008). This review was considered a high-quality systematic review and provided data on the comparative accuracy of these technologies. As this review did not provide any information on unsatisfactory rates, the most recent HTA report considering this outcome was also included (Krahn et al. 2008).

The review by the CADTH (Krahn et al. 2008) was based on evidence from a systematic literature search to June 2006. The more recent review (Arbyn et al. 2008) included one study (Ronco et al. 2007a) which was published after June 2006 and conducted in a screening population. Therefore. data on unsatisfactory rates were extracted from this study and included in the current review.

Three potentially eligible primary accuracy studies published after the search period of the Arbyn et al. (2008) review were identified. Two studies (Celik et al. 2008; Lerma et al. 2007) reported the accuracy of LBC and conventional cytology for the detection of low (LSIL or CIN 1) or high-grade lesions on biopsy. These studies were excluded as the reference standard threshold was different to that specified for this review. One study by Cibas et al. (2008) compared the accuracy of the MonoPrep Pap Test with conventional cytology using a reference standard of adjudicated cytology by an independent

pathologist. This is considered a suboptimal reference standard. The study therefore provided a lower quality of evidence to that in Arbyn et al. (2008) and was excluded from review.

One primary study conducted in Scotland investigated patient's preferences and willingness-to-pay for reduced time for Pap test recalls and test results, and more frequent screening (Wordsworth et al. 2006). This study was excluded from the current review as neither the costs nor the recall rates or test waiting times in this study are applicable to the Australian setting.

Study characteristics, quality and applicability

The characteristics and quality appraisal of the included systematic reviews are summarised in Table 13. Both systematic reviews were considered high quality for accuracy outcomes, the CADTH review was considered fair quality for the outcome of unsatisfactory rates.

The systematic review by Arbyn et al. (2008) included 9 primary studies of the diagnostic accuracy of LBC compared to conventional cytology. A specified inclusion criterion was verification of disease status in all subjects by colposcopy, augmented by histology obtained from biopsy where indicated. In eight non-randomised studies all subjects received reference standard verification by colposcopy ± biopsy. Pooled relative and absolute sensitivity and specificity for three test thresholds (HSIL+, LSIL+, or pLSIL+) were determined by meta-analysis. A single RCT of 45, 174 women with more than 90 per cent follow-up of women with a positive cytology result was also included for the meta-analysis of relative test accuracy only; additional data provided by the authors of the RCT were used in the review (Ronco et al. 2007a). Seven of the nine included studies were in women undergoing follow-up of a previous cervical abnormality, but a subgroup analysis of studies in patients at higher risk compared to a screening population was presented in the review. This review pooled data from different LBC systems, but conducted sensitivity analyses by LBC proprietary name.

The HTA by CADTH (Krahn et al. 2008) included a meta-analysis of unsatisfactory rates and discordant cytological classifications around an LSIL threshold. Data were reported separately for ThinPrep (filtration-based) and SurePath (centrifugation-based) LBC systems, but no other subgroup analyses were conducted. The review included 20 studies that directly compared LBC and conventional cytology sensitivity and specificity and 66 studies reporting unsatisfactory rates. Pooled values were derived by a Bayesian meta-analysis using a random-effects model. Most outcomes demonstrated significant between study heterogeneity. The authors noted that the rates of unsatisfactory slides reported in different studies varied with both reporting terminology and screening practices.

Table 13 Characteristics and appraisal of systematic reviews and HTAs of LBC.

Author, year Country	Objective & methods	Included studies Outcomes	Quality assessment of review
Arbyn et al. 2008 Belgium	Objective: To compare accuracy of conventional Pap and LBC cervical samples Literature review: • PubMed (MEDLINE & PreMEDLINE) & EMBASE Jan 1991 to May 2007 • Table of contents of 5 gynecologic & 4 cytopathology journals • Reference lists Inclusion/exclusion criteria: *Study design:* Comparative studies (concomitant split-sample or direct-to-vial, or 2-cohort design) *Reference standard:* All subjects receive reference standard verification by colposcopy ± biopsy for CIN 2+, if indicated; OR RCTs with ≥ 90% complete follow-up confirmation of positive women *Intervention:* LBC *Comparator:* Conventional Pap *Language:* No language restrictions	9 primary studies • 7 studies with concomitant testing (4 split-sample) • 1 two-cohort study • 1 RCT • 6 ThinPrep, 1 AutoCyte, 1 DNA Citoliq, 1 CellSlide • 2 studies screening population Outcomes • Accuracy for CIN 2+ (absolute & relative sensitivity & specificity)	Quality: High Explicit review questions: yes Explicit & appropriate eligibility criteria: yes Explicit & comprehensive search strategy: yes Quality of included studies appraised: yes Methods of study appraisal reproducible: yes Heterogeneity between studies assessed: yes Summary of main results clear and appropriate: yes Applicability: high Most studies in women with previous abnormalities due to criteria for all subjects receiving high-quality reference standard, but subgroup analysis conducted. Pooled data on different LBC systems, but subgroup analysis conducted.
Krahn et al. 2008 CADTH Canada	Objective: To assess the effectiveness and cost-effectiveness of LBC with and without HPV compared to conventional cytology Literature review: • BIOSIS Previews, CancerLit, EMBASE, MEDLINE, Cochrane Library Nov 2002 to June 2006 Inclusion/exclusion criteria: *Study design:* Systematic reviews, HTAs or primary studies with direct comparison *Intervention:* LBC *Comparator:* Conventional Pap	21 secondary studies • 19 HTAs/systematic reviews/economic evaluations identified in update 108 primary studies • 44 identified in update • 20 direct comparison studies • 49 LBC studies • 47 split-sample • 31 two-cohort studies • 66 studies with unsatisfactory rate	Quality: high (accuracy), fair (unsatisfactory rates) Explicit review questions: yes Explicit & appropriate eligibility criteria: yes Explicit & comprehensive search strategy: yes Quality of included studies appraised: yes Methods of study appraisal reproducible: yes Heterogeneity between studies assessed: accuracy studies yes, unsatisfactory rates no Summary of main results clear and appropriate: yes Applicability: limited (unsatisfactory rates) Non-comparative studies included. LBC manufacturer data reported separately, however no subgroup analyses by study design/ setting/ quality conducted.

Test accuracy

The pooled absolute and relative sensitivities and specificities and the diagnostic odds ratios (DORs) of LBC versus conventional cytology from the Arbyn et al. (2008) systematic review are summarised in Table 14. As the test positivity threshold was lowered from HSIL+ to ASCUS+, the sensitivity for the detection of CIN 2+ increased and the specificity decreased for both test preparation methods.

The meta-analysis demonstrated a modest non-statistically significant increase in sensitivity to detect CIN 2+ for LBC over that of conventional cytology when all studies were pooled, using three different test threshold levels (Table 14 and Figure 3; LBC:conventional sensitivity ratio HSIL+ 1.05, 95% confidence interval [CI] 0.95–1.16; LSIL+ 1.03, 0.96–1.11; pLSIL+ 1.03, 0.97–1.09). The same result was found when considering only the Italian RCT (Table 14; LBC:conventional sensitivity ratio HSIL+ 1.07, 95%CI 0.71–1.26; LSIL+ 1.03, 0.74–1.43).

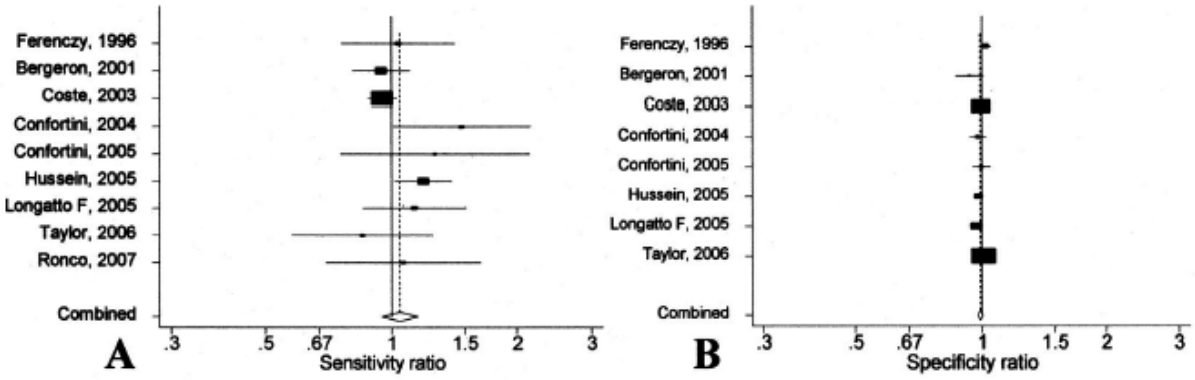


Figure 3 Relative sensitivity (A) and specificity (B) at test threshold of HSIL+ from Arbyn et al. (2008).

The specificity of LBC and conventional cytology were not significantly different at HSIL+ or LSIL+ test thresholds (Table 14; HSIL+ 0.99, 95%CI 0.98–1.01; LSIL+ 0.97 95%CI 0.94–1.01). The specificity of LBC was significantly lower than conventional cytology at a test threshold of pLSIL (ASCUS+, LBC:conventional specificity ratio 0.91, 95% CI 0.84–0.98).

The authors reported that the contrast in DOR between conventional cytology and LBC was not influenced by the number of quality issues of the Standards for Reporting of Diagnostic Accuracy (STARD) guidelines that were appropriately addressed in the individual study reports (HSIL+ $P = 0.84$, LSIL+ $P = 0.37$, ASCUS+ $P = 0.65$).

There was no significant heterogeneity between studies in the relative sensitivity of LBC and conventional cytology at a test threshold of HSIL+. There was significant heterogeneity in the relative sensitivity and specificity estimates across the different studies at all other test thresholds. Subgroup analyses of the relative accuracy estimates of LBC versus conventional cytology for CIN 2+ (test threshold HSIL+) are summarised in Table 15. None of the factors analysed (screening vs high-risk setting, paired vs independent patient cohorts, split-sample vs direct-to-vial studies, completeness of biopsy as reference standard, or proprietary name) contributed to heterogeneity in sensitivity.

Table 14 Estimates of the relative accuracy of LBC versus conventional cytology for CIN 2+ from Arbyn et al. (2008).

Study *N*	Test threshold	LBC Absolute accuracy			Conventional Absolute accuracy			Relative accuracy LBC : conventional	
		Sensitivity % (95% CI)	Specificity % (95% CI)	DOR (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Pooled data 7 studies absolute data 9 studies relative Se data 8 studies relative Sp data 7 studies Se pLSIL 6 studies Sp pLSIL	HSIL+	57 (46–67)	97 (94–99)	43 (21–89)	55 (46–65)	97 (96–98)	36 (21–62)	1.05 (0.95–1.16)	0.99 (0.98–1.01)
	LSIL+	79 (70–86)	79 (70–86)	14.1 (8–26)	76 (67–83)	81 (72–88)	13 (7–25)	1.03 (0.96–1.11)	0.97 (0.94–1.01)
	pLSIL+ [a]	90 (83–95)	65 (50–77)	17 (9–34)	88 (80–93)	71 (58–82)	19 (10–35)	1.03 (0.97–1.09)	0.91 (0.84–0.98)
Italian RCT (Ronco)	HSIL+	–	–	–	–	–	–	1.07 (0.71–1.26)	–
	LSIL+	–	–	–	–	–	–	1.03 (0.74–1.43)	–

[a] ASCUS+ category. Se = sensitivity, Sp = specificity, DOR = diagnostic odds ratio.

Table 15 Subgroup analysis of relative accuracy estimates of LBC versus conventional cytology for CIN 2+ from Arbyn et al. (2008).

Parameter	Subgroup	Pooled accuracy ratio LBC : conventional			
		Sensitivity (95% CI)	Trials *N*	Specificity (95% CI)	Trials *N*
Setting	Screening setting	1.03 (0.76–1.40)	2	1.00 (0.99–1.00)	1
	High risk / follow-up population	1.05 (0.94–1.17)	8	0.99 (0.97–1.01)	8
Study design*	Paired study design	1.07 (0.95–1.20)	7	0.99 (0.97–1.00)	7
	Independent study design	0.94 (0.71–1.26)	2	1.01 (1.00–1.02)	1
Reference standard	Colposcopy + biopsy if indicated	1.07 (0.60–1.64)	1		0
	Complete colposcopy, histology if indicated	1.10 (0.94–1.28)	6	0.99 (0.98–1.01)	6
	Complete histology	0.96 (0.84–1.11)	2	0.98 (0.89–1.08)	2
LBC type*	ThinPrep	1.07 (0.92–1.23)	6	1.00 (0.99–1.01)	5
	AutoCyte	0.95 (0.81–1.11)	1	0.94 (0.87–1.01)	1
	CellSlide	1.27 (0.75–2.15)	1	1.00 (0.95–1.05)	1
	DNA Citoliq	1.14 (0.85–0.51)	1	0.97 (0.95–0.99)	1

* Significant between-study heterogeneity for relative specificity only.
NB: Test threshold HSIL+.

Figure 3 displays the Forest plot for the pooled relative accuracy measures. Two individual studies showed a significantly higher sensitivity for LBC at a HSIL or LSIL threshold (Confortini et al. 2004, Hussein et al. 2005); both of these studies investigated the ThinPrep LBC system. A subgroup analysis by proprietary name did not demonstrate a higher sensitivity for LBC over conventional cytology in the ThinPrep studies (Table 15). This analysis included a single study of an earlier generation of the ThinPrep system (ThinPrep beta, Ferenczy et al. 1996); this study did not demonstrate the lowest relative sensitivity of all included ThinPrep studies.

Study design and proprietary name contributed to study heterogeneity for specificity at a HSIL+ threshold; there was a higher specificity in studies using independent samples than in paired studies, and a lower specificity for the DNA Citoliq system (Table 15). This heterogeneity was due to a single study in both cases. At a test threshold of LSIL+ no significant differences in subgroup analyses were found. At a test threshold of ASCUS+, the DNA Citoliq system had a significantly higher relative sensitivity (1.25, 95% CI 1.11–1.42) and lower relative specificity (0.83, 95% CI 0.79–0.87). The authors also report that 'Summary receiver operating characteristic regression, using ThinPrep as reference, identified a lower DOR for AutoCyte at cutoff HSIL+', however data were not reported.

The inclusion criteria for studies in this systematic review were reference standard validation by colposcopy for all subjects plus biopsy where indicated as a minimum, thus the reported absolute sensitivity and specificity estimates can be considered valid. Subgroup analysis indicated the completeness of histological follow-up did not contribute to study heterogeneity. The major limitation of this review (as a consequence of these inclusion criteria) is that many included studies were conducted in patients at higher than average risk. However, a subgroup analysis demonstrated that studies conducted in screening populations (two studies including the Italian RCT) gave similar estimates of relative sensitivity and specificity compared to studies conducted in high risk or follow-up populations (seven studies) and to the overall pooled values (Table 15). The population setting did not contribute to study heterogeneity, but only one study provided an estimate of relative specificity in a screening population. The clinical question for the

present review is concerned with the accuracy of cervical cytology in a mixed screening and diagnostic population.

Unsatisfactory rate

A systematic review and HTA report by CADTH (Krahn et al. 2008) included data on unsatisfactory slide rates from 44 studies of ThinPrep. Overall, 0.95 per cent (6674/704 813) of LBC slides and 1.04 per cent (13 664/1 316 318) of conventional slides were unsatisfactory. A Bayesian meta-analysis gave pooled unsatisfactory estimates of 2.24 per cent (95% CI 1.20%–3.29%) for LBC and 3.04 per cent (1.92%–4.16%) for conventional slides, with a difference of –0.81 per cent (–1.87% to 0.24%). statistically significant heterogeneity was observed between studies for this outcome.

Fifteen studies provided data on unsatisfactory slide rates for SurePath LBC. Overall, 0.42 per cent (2539/597 565) of LBC slides and 1.39 per cent (9598/692 406) of conventional slides were unsatisfactory. A Bayesian meta-analysis gave pooled unsatisfactory estimates of 0.82 per cent (95% CI 0.14%–1.51%) for LBC and 3.31 per cent (0.97%–5.67%) for conventional slides, with a difference of –2.49 per cent (–4.43% to –0.55%). There was also significant between-study heterogeneity in these studies.

No analyses were conducted to investigate whether results varied by study quality, design or population. A Forest plot indicated that the difference in unsatisfactory rates between the two techniques was close to zero in larger studies. The authors concluded that, on average, LBC may have a lower unsatisfactory rate, but the estimate from different studies varied.

One study identified in the more recent systematic review of accuracy data by Arbyn et al. (2008) was an RCT of 22 708 women receiving LBC (ThinPrep) and 22 466 women receiving conventional Pap testing (Ronco et al. 2007a). This study reported unsatisfactory rates from a greater number of slides than the earlier publication of preliminary findings (Ronco et al. 2006) that was included in the CADTH HTA report. In this study all women with positive cytology were referred for colposcopy. The colposcopists were not blinded to the test result, but the biopsy results from women with a confirmed CIN 2+ lesion were reviewed blind to the cytology test and result. A reference standard of colposcopy ± biopsy was available for 91 per cent of women receiving conventional cytology and 93 per cent receiving LBC, with a cytology result of pLSIL (ASCUS) considered as the test positivity threshold. This study is not of ideal quality for determining test accuracy. However, reference standard verification is not required for measurement of unsatisfactory rates.

This RCT reported that fewer women in the LBC arm of the study had an unsatisfactory result (2.57% LBC versus 4.11% conventional). The relative frequency for unsatisfactory slides was lower with LBC (0.62, 95% CI 0.56–0.69). The authors reported that this difference was due to a decrease in the proportion of women with unsatisfactory slides due to obscuring inflammation (0.44% LBC, 2.15% conventional; relative frequency 0.21, 95% CI 0.17–0.25). The proportion of women with unsatisfactory slides due to other reasons did not differ between cytology preparation methods. The reduction in unsatisfactory slides was larger for women aged 25–34 than women aged 35–60 (relative frequency 0.53, 95% CI 0.44–0.63 and 0.67, 95% CI 0.59–0.76, respectively).

Test yield

The HTA by CADTH (Krahn et al. 2008) also provided some information on comparative test yields of LBC and conventional cytology, as categories of discordant results. The authors reported that in 33 split-sample studies of ThinPrep in 113 286 women, 3.1 per cent (95% CI 2.0%–4.1%) of slides were classified as LSIL+ by LBC but as negative or ASCUS by conventional cytology and 2.2 per cent (95% CI 1.3%–3.1%) were classified as LSIL+ by conventional cytology but as negative or ASCUS by LBC. The pooled values for SurePath, derived from 13 studies (24 633 women), were 4.1 per cent (1.3%–6.8%) LSIL+ by LBC but <LSIL by conventional cytology and 3.1 per cent (0.4%–5.71%) LSIL+ by conventional cytology, but < LSIL by LBC. These values were derived by random-effects Bayesian meta-analysis. There was significant heterogeneity in disease prevalence across the studies.

The difference in total slide classifications between LBC and conventional cytology are reported for two-cohort studies. ThinPrep classified 1.3 per cent (95% CI 0.02%–2.6%) more slides as LSIL+ than conventional cytology did in 22 studies, and SurePath classified 0.66 per cent more (95% CI 0.23%–1.09%) as LSIL+ in 10 studies. There was no significant difference in the classification of slides as HSIL+ (ThinPrep – conventional: –0.25%, 95% CI –1.6% to +1.1%, 12 studies; SurePath – conventional: –0.11%, 95% CI –0.72% to +0.51%, 10 studies). There was significant between-study heterogeneity in all outcomes except SurePath classification of LSIL+.

Automated slide reading

FocalPoint

No eligible studies of the accuracy of the FocalPoint system for reading LBC slides compared to manual reading of conventional slides were identified.

No evidence for an accuracy advantage, disadvantage or equivalence of the AutoPap system was found in

- two studies comparing AutoPap-assisted reading of conventional slides to manual reading of conventional slides
- a single study of highly limited applicability comparing the AutoPap-assisted reading of LBC to manual reading of LBC.

The AutoPap-assisted reading of conventional slides reduced unsatisfactory rates (one study) compared to manual reading. However, process advantages are not considered relevant if sufficient evidence of accuracy is not available.

ThinPrep Imager

Accuracy

One fair-quality Australian study (Davey et al. 2007a) of the ThinPrep Imager system compared to manual reading of conventional cytology found

- a significant increase in the detection of CIN 2+ lesions (pHSIL threshold; additional 0.82 cases per 1000 women screened)

- no significant increase in the number of false positive biopsy results (pHSIL threshold)

A second fair-quality Australian study (Roberts et al. 2007) of the ThinPrep Imager system, with a higher possibility of verification bias, found

- a non-significant increase in the detection of high-grade lesions (pHSIL threshold, 1.49 additional cases per 1000 women screened)
- a significant increase in the number of false positive biopsy results (pHSIL threshold, 1.93 additional cases per 1000)
- a significant increase in the detection of high-grade lesions on manual LBC compared to conventional cytology (pHSIL threshold, 2.8 additional cases per 1000)

Three studies comparing ThinPrep Imager reading of LBC slides to manual reading of LBC slides found

- no significant difference in the number of high-grade cases detected
- significantly fewer false positive cases for high-grade lesions at a test threshold of pHSIL or HSIL, in two of three studies, respectively. There is a possibility of verification bias in one of these studies.

Unsatisfactory rates

One Australian study (Davey et al. 2007a) demonstrated a significant decrease in classification of unsatisfactory slides by the ThinPrep Imager system compared to conventional cytology (1.8% vs 3.1%; $P < 0.001$).

Process outcomes

Two Australian studies found the ThinPrep Imager compared to conventional cytology

- decreased slide reading time (mean difference 7.18 slides per hour [95% CI 6.17–8.20]; $P < 0.001$, Davey et al. 2007b)
- significantly increased the number of slides classified with low-grade abnormalities

Comparison to conventional cytology

Included studies

The search for studies published since the MSAC 2003 review identified five primary studies of automation-assisted slide reading systems that met the inclusion criteria for review.

An excluded discussion paper by Wilbur and Norton (2002) cites evidence comparing AutoPap-assisted reading of conventional slides to conventional cytology from a trial published in 1999. This trial was considered in the previous MSAC (2003) review. Nevertheless, it is noteworthy that this study would not provide high-quality evidence for the primary research question because it does not refer to use of the current FocalPoint

system for SurePath slides and includes the use of consensus cytology as a reference standard.

Study characteristics, quality and applicability

Two studies comparing the accuracy of ThinPrep Imager reading of ThinPrep LBC slides to manual reading of conventional Pap smear slides were included (Davey et al. 2007a; Roberts et al. 2007). These studies address the primary question for the present review. An additional study provided information on process outcomes of ThinPrep Imager reading of ThinPrep LBC slides compared to manual reading of conventional cytology slides, over the same time period and for the same cytologists as one of these accuracy studies (Davey et al. 2007b). Two studies comparing AutoPap reading of conventional cytology slides to manual reading of conventional slides (Confortini et al. 2003; Stevens et al. 2004) relevant to a secondary research question were also included for review. The characteristics and quality of these five studies are summarised in Table 16.

The four studies of accuracy were fair-quality, paired diagnostic accuracy studies. Two studies were conducted in applicable populations (Confortini et al. 2003; Davey et al. 2007a); in two studies the applicability of the population was unclear (Roberts et al. 2007; Stevens et al. 2004). In three studies using a reference standard of histology, this was not applied to all of the patients with discordant positive results (Confortini et al. 2003; Davey et al. 2007a; Roberts et al. 2007). In the fourth study a suboptimal reference standard of adjudicated cytology was used (Stevens et al. 2004). The two studies of the ThinPrep Imager system (Davey et al. 2007a; Roberts et al. 2007) were conducted in an Australian setting.

The study of ThinPrep Imager reading by Davey et al. (2007a) was conducted in a mixed screening and diagnostic population. The study reported results according to three different reference standards of histology obtained up to 6 months following the test. These were a) the histological result from the Pap test register, b) blinded re-reading of the histology slides, and c) the more severe of these two results. No reference standard was available for 30 per cent of discordant samples. However, the proportion of discordant cases with histology results available did not differ between the two tests (χ^2 = 0.50, 1 df, *P* = 0.48). The authors also reported no significant association between which test gave the higher result and the odds of verification among discordant test pairs using logistic regression ($\chi 2$ = 0.34, 1df, P = 0.56). There was no evidence that the odds of verification varied across categories of discordant tests either ($\chi 2$ = 3.7, 10df, P = 0.96). A similar analysis for glandular classifications was not reported. The study reported the number of additional cases detected and biopsies conducted using a CIN 1 test threshold. A complete contingency table for results of both tests and the Pap test register biopsy results as the reference standard was also provided as an appendix. The reviewers calculated accuracy outcomes at different test thresholds from the data in this appendix. The authors state that at the time the study was conducted, guidelines recommended referral of women with CIN 1 or higher lesions to colposcopy.

A second publication by Davey et al. (2007b) compared slide reading times for ThinPrep Imager-reading LBC slides versus conventional cytology. This retrospective study considered process outcomes over the time period at the beginning of the data collection for the accuracy study (August 2004 to February 2005, the accuracy study was conducted over August 2004 to June 2005). These data included the same cytologists and slides as in the accuracy study (personal correspondence, Dr Elizabeth Davey, 14th April 2008). The

comparisons of slide reading times for the two approaches are based on paired data for 20 cytologists. In addition, data on conventional cytology reading times for 21 cytologists not trained to use the ThinPrep Imager system were included for comparison. Data from 1645 slide reading sessions were available in total. Sessions where session values were more than 3 standard deviations from the mean for a particular reader, or where non-conventional cytology slide reading took place in conventional cytology sessions, were excluded (2.3% of sessions). The final analysis was based on 581 conventional cytology sessions and 379 ThinPrep Imager sessions (paired data), plus an additional 685 conventional cytology sessions from ThinPrep Imager untrained cytologists. Subgroup analyses investigating the role of experience and cytologist reading speeds were also undertaken.

The study of ThinPrep Imager reading by Roberts et al. (2007) included 103 seeded cases in addition to routinely received samples, but true and false positive results are also reported excluding the seeded cases. *Only data from routine samples are presented in this assessment report*. The population included was not clearly described, but most likely includes a mixed screening and diagnostic population. The authors reported sensitivity of ThinPrepImager reading of LBC slides, manual reading of LBC slides and manual reading of conventional slides for the detection of high-grade lesions, however these data are considered invalid as there is incomplete verification and the total number of cases is not known (ie no verification of negative cases and degree of verification of discordant positives not reported). The reported statistical analysis of these accuracy measures is therefore also considered invalid. In lieu, the reviewers have compared the number of cases detected between the arms of the study by χ^2-test (although a McNemar's test is ideal, this was not possible here due to the lack of information on concordant test results; a χ^2-test will have less power to detect a difference). The authors define histological high-grade disease as HSIL, AIS or carcinoma, however this does not clearly relate to histological categories. Slides that could not be read by the imager (3.7% of slides) were excluded from the analysis. Slides suspected of being unsatisfactory underwent full manual screening and were included in the accuracy data for the study as negative test results, however this approach has no impact on the number of true and false positive findings. The total number of slides categorised as unsatisfactory with each screening method was not reported. A reference standard of histology was used for slides with a pHSIL or higher cytology result. No histology was available for 9.0% of positive ThinPrep Imager read slides, 8.3% of positive LBC slides read manually and 10.8% of positive conventional slides (χ^2-test $P = 0.834$). It is not reported what proportion of these unconfirmed positive results were discordant, thus there may be some verification bias in the data reported. It is also unclear whether or not the referring practitioners were aware of the type of cytology that led to the most serious cytological finding in the final report.

The study of AutoPap-assisted reading of conventional slides by Confortini et al. (Confortini et al. 2003) also used histology as a reference standard, but the number of slides without reference standard verification was not reported. Test results of ASCUS favouring squamous intraepithelial lesion (SIL) cytology or more severe prompted a colposcopy. Test results of ASCUS favouring a reactive process did not refer to colposcopy (instead patients were recommended for a repeat smear after 6 months). Slides that could not be read by the AutoPap system but were classified as 'process review' (12.8%) were included in the study as the manual reading result in the AutoPap and manual arms. This reflects standard practice in the use of automation-assisted reading. This differs from the design of the other studies where slides not processed were

excluded from the study. The study provided contingency tables of test results for patients with histologically confirmed CIN 2+ lesions, thus the true positive rates for different test thresholds could be calculated. Test findings for glandular lesions were only reported as a combined category of ASCUS-favouring SIL lesions or atypical glandular cells of undetermined significance (AGUS), so glandular test results could not be separated from squamous findings. The study also reported estimated reading time saved in slides allocated by the AutoPap system as 'not for review' (NFR). As the current standard use of the FocalPoint system does not involve allocating slides as NFR (see **Error! Reference source not found.**), these data are not relevant to current use of the FocalPoint system.

One additional Australian study of AutoPap-assisted reading of conventional slides in comparison to manual reading of conventional slides was identified (Stevens et al. 2004). This study reported the detection of high-grade cytology by each method with a reference standard of adjudicated cytology for discordant results. This study excluded from the analysis 986 slides designated as 'not for review' by the AutoPap system, potentially introducing bias into the study findings. The reporting of results in the study is unclear. It is uncertain whether data in contingency tables is for concordance between screening approaches or accuracy data. Despite these shortcomings, the study is included in the present review as it was conducted in an Australian setting.

Table 16 Characteristics and quality of studies reporting on the relative accuracy of automated slide reading and conventional Pap smear cytology with manual slide reading.

Author, year Setting	Population (*N*, inclusion criteria)	Test comparison/s	Study design	Quality & applicability
Davey et al. (2007a) Australia One pathology laboratory Aug 2004 – June 2005	*N* = 55164 Inclusion criteria • Any age • Electing LBC sample • Screening and diagnostic population	Index test ThinPrep LBC with ThinPrep Imager reading Comparator test Conventional cytology with manual reading	Study design: Diagnostic accuracy study with split-sample pairs Test threshold: Test: CIN 1+, Reference standard: CIN 2+ Contingency data provided enabling re-calculation of accuracy data for alternative test thresholds. Reference standard: Histology (within 6 months) for discordant cytology, using 1) register results; 2) blinded re-reading of histology slides; and 3) more severe result from 1) or 2). No reference standard for 254/844 (30%) discordant samples. Logistic modelling indicated no association between test and whether verified (χ^2 = 0.50, 1df, *P* = 0.48). Outcomes • Accuracy for squamous lesions: incremental TP, TP:FP, accuracy for glandular lesions • Unsatisfactory rate • Test yield	C1 P1 Q2 Accuracy study NHMRC level III-2 Quality: fair Prospective, consecutive, reference standard valid but not applied to all discordant participants Applicable Applicable population (mixed screening & diagnostic), index test & comparator
Davey et al. (2007b) Aug 2004 – Feb 2005	Slide reading sessions TPI trained readers: *N* = 581 CC, 379 TPI Non-TPI readers: *N* = 685 CC Exclusion criteria • CC sessions: any including non-CC • Sessions > 3SDs from mean for reader • 38/1645 (2.3%) excluded		Study design: Retrospective, cross-sectional study. Substudy of diagnostic accuracy study. Outcomes • Screening time: min/slide, slides/h Subgroups • Trained on TPI vs not trained on TPI • Years experience • Speed reading conventional cytology	

Author, year Setting	Population (*N*, inclusion criteria)	Test comparison/s	Study design	Quality & applicability
Roberts et al. 2007 Australia One pathology laboratory Feb 2005 – April 2005	*N* = 11 416 + 103 seeded LBC cases Inclusion criteria • Routinely received samples ((*N* = 11 416) • Cases HSIL, AIS or carcinoma (*N* = 103), these cases excluded from data reported here Exclusion criteria • Slides not read by imager (3.7%)	Index test ThinPrep LBC with ThinPrep Imager reading Comparator test • Conventional cytology with manual reading • Manual reading of LBC samples	Study design: Three-arm diagnostic accuracy study with split-sample pairs and seeded cases. Data reported separately for routine and seeded cases. Test threshold: Test: HSIL/pHSIL (including possible glandular high grade), Reference standard: high grade Reference standard: Histopathology (within 9 months) for high-grade and possible high-grade lesions, no reference standard for 12/134 TPI (9.0%), 15/180TPM (8.3%), 10/93 conventional (10.8%) slides (χ^2-test *P* = 0.834) None for low-grade positive, negatives. Outcomes • Accuracy for squamous lesions: sensitivity[a], incremental TP (with and without seeded cases) • Test yield • Screening time	C1 P2 Q2 Accuracy study NHMRC level III-2 Quality: fair Prospective, not consecutive, reference standard valid but not applied to all discordant participants. Excluded slides not read by imager. Applicability unclear Applicable index test & comparator, population applicability unclear
Confortini et al. 2003 Italy April 2000 – NR	*N* = 14 145 Inclusion criteria • Smear samples from Florence screening program	Index test AutoPap 300QC—assisted reading of conventional smears Comparator test Manual reading of conventional smears	Study design: Diagnostic accuracy study Test threshold: Test: HSIL; pLSIL (ASCUS-R), Reference standard: CIN 2+ Reference standard: Histology (biopsy or loop excision); ASCUS favouring reactive process: repeat smear 6mo then colposcopy if necessary; ASCUS favouring SIL +: colposcopy; HSIL, colposcopy neg: repeat cytology 6mo. Proportion reference standard unavailable NR Outcomes • Accuracy: TP, not FP • Unsatisfactory • Test yield • Costs & cost-effectiveness (cost per CIN 2+ detected)	C1 P1 Q2 Accuracy study NHMRC level III-2 Quality: fair Prospective, consecutive, reference standard valid but not applied to all discordant participants Applicable Applicable population, index test & comparator

Author, year Setting	Population (*N*, inclusion criteria)	Test comparison/s	Study design	Quality & applicability
Stevens et al. 2004 Australia Jan 2000 – Feb 2000	*N* = 5583 Inclusion criteria • Consecutive 6000 slides Exclusion criteria • Broken, cracked, 2 coverslips, vaginal smears, LBC (422, 7.0%) • AutoPap NFR (986, 16.4%)	Index test AutoPap location guided screening (LGS)-assisted screening of conventional smears, including archiving NFR Comparator test Manual reading of conventional smears	Study design: Diagnostic accuracy study with split-sample pairs and seeded LBC cases Test threshold: Low-grade abnormality (LGEA+) Reference standard: Concordant positives and negatives: nil. Discordant positives: adjudicated cytology, majority agreement of 3 experienced cytology professionals Outcomes • Accuracy: cases detected, not clearly reported	C1 P2 Q2 Accuracy study NHMRC level III-2 Quality: fair Retrospective, not consecutive (NFR excluded), reference standard valid but suboptimal Limited applicability Applicable population & comparator Index test limited applicability (NFR slides)

Abbreviations: CC = conventional cytology, LBC = liquid-based cytology, NFR = not for review, NR = not reported, TPI = ThinPrep Imager, TPM = ThinPrep manual.

[a] Reported sensitivity invalid as not based on full reference standard.

Test accuracy

Table 17 summarises the relative accuracy data from the included studies comparing automation-assisted slide reading to manual reading of conventional slides. As none of the studies included verification of negative results, nor fully verified positive results, a valid estimate of absolute test sensitivity and specificity cannot be determined. In lieu of this, Table 17 presents the number of additional or absolute true and false positive findings detected by each screening method.

Data from a recent large Australian split-sample study of ThinPrepImager reading of ThinPrep slides indicated that this screening method detected an additional 0.82 true cases of CIN 2 or higher lesions per 1000 women screened than manual reading of conventional slides, when **pHSIL** was used as the test threshold (Davey et al. 2007a). This was a significantly greater number of cases detected than by manual reading (P = 0.003 χ^2-test). This increased detection rate was associated with an additional 0.74 biopsies per 1000 women (P = 0.06, χ^2-test). Fewer colposcopies and biopsies per 1000 women were conducted that had a negative finding (0.07 fewer per 1000 women, P = 0.80), considering this threshold for referral to colposcopy.

When **CIN 1** (LSIL excluding HPV effect) was used as the test threshold, and the three alternative histological reference standard approaches were considered (see Table 16), ThinPrep Imager reading of ThinPrep slides detected 1.29 to 1.92 additional true cases of CIN 2 or greater lesions per 1000 women screened than manual reading of conventional slides ($P < 0.001$ χ^2-test for all 3 cases) (Davey et al. 2007a). This was associated with 2.28 to 3.08 additional biopsies per 1000 women, significantly more than in the conventional arm ($P < 0.001$, χ^2-test). These data represent the actual number of cases detected and biopsies performed in a standard clinical setting, based on recommendations for referral for biopsy for CIN 1+ lesions and incomplete follow-up. The authors state that this test threshold for referral for biopsy was recommended practice at the time, but current guidelines recommend referral for biopsy for pHSIL and higher lesions.

Data are also presented in Table 17 for a test threshold of LSIL (atypia with HPV effect or higher). Although a cytological threshold of CIN 1+ was recommended for referral for biopsy at the time, 15 per cent (99/656) of women with discordant lesions indicative of atypia with HPV effect were also biopsied. At this test threshold, with the Pap test register result as the reference standard, the ThinPrep system detected an additional 1.40 true cases of CIN 2+, with 3.77 additional biopsies per 1000 women.

Another Australian split-sample study of TPI reading of ThinPrep slides showed a similar trend, with 1.58 more cases of high-grade histology per 1000 women detected by the ThinPrep system than by CC, at a threshold of pHSIL/HSIL (Roberts et al. 2007) (Table 17). Where cytology findings of possible high-grade glandular lesions are also included as a positive cytological result, the ThinPrep system detected 1.49 more cases per 1000 than CC. But the number of cases detected did not differ statistically between these two screening approaches (χ^2-test). The number of biopsies with negative results for high-grade disease was significantly greater with TPI than with CC (1.93 additional cases, $P < 0.01$). The number of cases detected at a pLSIL test threshold is also presented in Table 17, but during the time period when this study was conducted, the test threshold for referral for biopsy was pHSIL and higher lesions.

There is inconsistency in the number of additional false positive findings due to ThinPrep Imager reading between the Davey et al. (2007a) and Roberts et al. (2007) studies at a pHSIL threshold. The reason for this is not apparent, however differences in terminology and classification may be contributing factors. In the Davey et al. (2007a) study, a test threshold of 'inconclusive high-grade disease' was used to detect histological CIN 2+. In the Roberts et al. (2007) study, a test threshold of 'possible high-grade disease' was used for a reference standard threshold of 'high-grade disease' defined as HSIL, AIS or carcinoma.

Interpretation of the findings of the Roberts et al. (2007) study must also be considered in light of the three-way paired comparison of these tests to manual reading of ThinPrep slides as discussed below ('Direct comparison of LBC, automated LBC reading and conventional Pap cytology', page 58).

Two studies did not find an accuracy advantage of using AutoPap-assisted reading of *conventional* slides over manual reading of conventional slides (Confortini et al. 2003, Stevens et al. 2004; secondary research question). The data from Confortini et al. (2003) include manual reading of 3 slides that were not read by the AutoPap system and were processed manually (Table 17). An Australian study (Stevens et al. 2004) stated that AutoPap reading of conventional slides detected four cases of high-grade or suspected high-grade (pHSIL+) disease that were classified as normal by conventional cytology, and five cases were detected by conventional reading that were classified as normal by AutoPap. The contingency data for this study were not clearly reported and it is assumed that these were true positive cases. These data excluded 18 smears diagnosed as low-grade positive by conventional reading but high-grade positive by AutoPap-assisted reading. This study also excludes data from slides designated as 'not for review' by the AutoPap system. These studies did not report power calculations, but it is unlikely that they were powered to demonstrate true equivalence of the two approaches.

Table 17 Estimates of the relative accuracy of automation-assisted versus conventional reading of cervical cytology for CIN 2+.

Study *N*	Test	Test threshold	Automation-assisted reading, additional positives per 1000 screened				Positives						TP cases detected *P* (χ^2-test)
							Automation-assisted reading			Conventional cytology			
			TP	FP	TP:FP	Extra bx	TP	FP	TP:FP	TP	FP	TP:FP	
Automation-assisted reading of LBC													
Davey et al. 2007[f]								*discordant*			*discordant*		
55 164	TPI	pHSIL+, register ref std	0.82	–0.07	1:–0.09	0.74	141	121	1:0.86	96	125	1:1.30	0.003
		LSIL+ [b], register ref std	1.40	2.37	1:1.70	3.77	132	297	1:2.25	55	166	1:3.02	
		CIN 1+[a], register ref std	1.29	1.79	1:1.39	3.08	133	247	1:1.85	62	148	1:2.37	< 0.001
		CIN 1+[a], most severe histology as ref std	1.92	1.16	1:0.60	3.08	196	184	1:0.94	90	120	1:1.33	
		CIN 1+[a], blinded re-reading as ref std	1.54	0.74	1:0.48	2.28	153	127	1:0.83	68	86	1:1.26	
Roberts et al. [e] 2007								*total*			*total*		
11 416	TPI	pHSIL/ HSIL+, incl glandular	1.49	1.93	1:1.29	3.42	80	42	1:0.53	63	20	1:0.32	0.15
		pHSIL/ HSIL+, excl glandular	1.58)	–	–	–	78	–	–	60	–	–	0.12
		pLSIL/ LSIL+, incl glandular	1.58	–	–	–	99	–	–	81	–	–	0.18
		pLSIL/ LSIL+, excl glandular	1.66	–	–	–	97	–	–	78	–	–	
Automation-assisted reading of conventional cytology													
Confortini et al. 2003													
	AutoPap	HSIL+, incl gland	–0.21	–	–	–0.2%[d] colposcopy	13	–	–	16	–	–	0.58
14 145		pLSIL+ [c], incl gland	–0.07	–	–	–	30	–	–	31	–	–	0.90

Abbreviations: bx = biopsy, FP = false positive, LBC = liquid-based cytology, ref std = reference standard, TP = true positive, TPI = ThinPrep Imager.

[a] LSIL excluding HPV effect. [b] Atypia with HPV effect+. NB CIN 1+ lesions referred for colposcopy. [c] ASCUS favouring reactive process. [d] $P = 0.07$. [e] Routine cases only. [f] Considering biopsy confirmed +ve findings only.

Table 18 summarises the accuracy of automation-assisted and conventional cytology findings of a glandular abnormality to detect histological adenocarcinoma *in situ* or adenocarcinoma. The data from Davey et al. (2007a) come from all cases with a histological reference standard (41% of slides with glandular abnormalities) and indicates the accuracy for detection of adenocarcinoma *in situ* or adenocarcinoma. When a test threshold of a possible high-grade lesion is used, the ThinPrep Imager missed one case, however this case would have been referred to biopsy due to diagnosis of CIN 3 for squamous abnormalities. Thus no confirmed cases would have been missed by either the ThinPrep Imager or conventional cytology when considering both glandular and squamous classifications together.

In the study by Roberts et al. (2007), classification of glandular abnormalities by the ThinPrep system detected 2 cases of adenocarcinoma-in-situ compared with 3 cases detected on conventional cytology. The small number of cases makes the relevance of these data uncertain.

Table 18 Estimates of the relative accuracy of automation-assisted versus conventional reading of glandular cervical cytology for AIS+.

Study *N*	Test	Test threshold	Positives						χ^2-test, TP cases detected *P*
			Automation-assisted reading			Manual reading conventional slides			
			TP	FP	TP:FP	TP	FP	TP:FP	
Automation-assisted reading of LBC									
Davey et al. 2007 35 599[a]	TPI	Atypia	6	8	1:1.3	6	16	1:2.7	1.0
		Possible high grade	5	3	1:0.6	6	9	1:1.5	0.76
Roberts et al. 2007 11 416	TPI	Atypia	2	–	–	3	–	–	0.65
		Possible high grade	2	–	–	3	–	–	0.65

Abbreviations: FP = false positive, TP = true positive. TPI = ThinPrep Imager

[a] Satisfactory slides with endocervical component.

Process outcomes

The ThinPrep Imager system classified more slides as abnormal than conventional cytology (pLSIL+ for TPI vs conventional from Davey et al. 2007a, 7.4% vs 6.0%; Roberts et al. 2007 7.9% vs 4.2%) (also see Table 20).

In the Davey et al. (2007a) study, the classification of slides between the cytology categories significantly differed between ThinPrep Imager and conventional cytology ($P < 0.001$, χ^2-test, Table 19 and Table 20). A significantly higher proportion of slides were classified as atypia, atypia with HPV effect, CIN 1, CIN 2 and CIN 3+ by the ThinPrep Imager reading compared to manual reading of conventional cytology (Table 20). A lower proportion were classified as pHSIL by the ThinPrep Imager (inconclusive, high-grade histology to be excluded, OR = 0.78, 95% CI 0.67 - 0.92, $P = 0.0028$).

In the Roberts et al. (2007) study the ThinPrep Imager classified a significantly greater proportion of slides as pLSIL, LSIL and HSIL than conventional cytology (Table 19).

The study by Davey et al. (2007a) demonstrated that the use of the ThinPrep Imager system was associated with a lower proportion of unsatisfactory slides (982 vs 1704,

1.8% vs 3.1%; $P < 0.001$ χ^2-test; Table 19), as expected from the Italian RCT using LBC, as discussed above (see 'Unsatisfactory rate' page 32).

Confortini et al. (2003) reported that the use of AutoPap reading of conventional slides also reduced unsatisfactory rates (Table 19). However, process advantages are not relevant if sufficient evidence of accuracy is not available (Table 17). Process time was also reduced, as only 83 per cent of slides underwent manual review; the study used the AutoPap system to designate slides as "not for review" (NFR). As the current use of the FocalPoint system does not involve allocating slides as NFR (see **Error! Reference source not found.**), these data are not relevant.

Data reported on slide reading times from included accuracy studies are summarised in Table 21. The study by Davey et al. (2007b) found that the number of slides read per hour was significantly increased with TPI-assisted reading (mean within-reader difference 7.18 [95% CI 6.17–8.20] slides per hour; $P < 0.001$, paired t-test). There was no difference in the time taken to read conventional slides between cytologists trained to use the ThinPrep Imager system and those who only read conventional slides (TPI trained, $n = 20$: 10.61 min, 95% CI 9.73–11.49 min; TPI-untrained, $n = 21$: 10.61 min, 95% CI 10.03–11.18 min). This study also reported that slower readers had a greater increase in slide reading time with TPI than fast readers ($P < 0.001$, linear regression, Figure 4). Experience did not significantly affect ThinPrep Imager reading times ($P = 0.41$) or within-reader differences ($P = 0.16$, Table 21).

The study by Roberts et al. (2007) reported that the ThinPrep system reduced mean slide reading times by 54 per cent (3.4 min for ThinPrep Imager reading, 7.4 min for manual reading of conventional slides). However, it did not report the degree of variability, completeness of data or any statistical analysis. This did not include time spent checking previously screened slides.

Exclusion of slides that were unsuitable for reading by the automated system from both of these analyses may bias these estimates slightly.

Table 19 Slide classifications of automation-assisted reading of LBC versus manual reading of conventional cytology slides.

Study	*N*	Not read (%)	Unsatisfactory (%)		Automation-assisted yield (%)							Conventional cytology yield (%)						
			Auto	Conv	pLSIL	CIN 1	LSIL	pHSIL	HSIL	AIS/SCC	Gland	pLSIL	CIN 1	LSIL	pHSIL	HSIL	AIS/SCC	Gland
ThinPrep Imager reading of ThinPrep LBC slides																		
Davey et al. 2007	55 164	–	1.8	3.1**	3.1[a*]	1.4[#]	3.07[b]	0.42[c**]	*CIN 2* 0.40[#]	≥ *CIN 3* 0.49[*]	0.04	2.8[a]	1.0	1.97[b]	0.54[c]	*CIN 2* 0.28	≥ *CIN 3* 0.43	0.09
Roberts et al. 2007	11 416	3.7	–	–	3.8[#]	–	2.9[#]	0.43	0.74*	–	–	1.8	–	1.6	0.33	0.48	–	–
AutoPap-assisted reading of conventional cytology																		
Confortini et al. 2003	14 145		0.88	1.3**	1.82[d]	–	0.2	–	0.13	0.007	–	2.99[d]	–	0.41	–	0.19	0.007	–

Abbreviations: AIS = adenocarcinoma *in situ*; Conv = conventional; Gland = glandular; SCC = squamous cell carcinoma.
[a] Atypia. [b] CIN 1 or atypia with HPV. [c] Inconclusive; high grade to be excluded. [d] ASCUS favouring a reactive process / ASCUS favouring squamous epithelial lesion or more severe / AGUS.
**P* < 0.05, ** *P* < 0.001, #*P* < 0.001 vs CC.

Table 20 Yield of automation-assisted reading of LBC versus manual reading of conventional cytology (Davey et al. 2007a).

Cytological classification	**Automation-assisted reading**	**Manual reading conventional slides**	OR	(95% CI)	*P*
Atypia	3.1	2.8	1.08	(1.01–1.15)	0.018
Atypia with HPV	1.6	1.0	1.65	(1.51–1.80)	< 0.001
CIN 1	1.4	1.0	1.47	(1.35–1.60)	< 0.001
Inconclusive; high grade to be excluded	0.4	0.5	0.78	(0.67–0.92)	0.0028
CIN 2	0.4	0.3	1.45	(1.20–1.75)	< 0.001
CIN 3+	0.5	0.4	1.15	(1.02–1.30)	0.019
Glandular	0.04	0.09			

Abbreviations: OR = odds ratio.

Table 21 Reading times of automation-assisted versus conventional reading of cervical cytology.

Slide reading	Study	**Automation-assisted reading LBC slides**	**Manual reading conventional slides** Mean (95% CI)	Difference	*P*
ThinPrep Imager reading of ThinPrep LBC slides					
Mean reading time min (95% CI)	Roberts et al. 2007	3.42 (NR)	7.40 (NR)		
	Davey (2007b), all readers	4.71 (4.38–5.04)	10.61 (9.73–11.49)	5.90 (5.04–6.75)	< 0.001
	1–10 y exp (*n* = 5)	4.8 (4.0–5.5)	11.6 (9.0–14.2)	6.8 (4.0–9.7)	
	11–20 y exp (*n* = 7)	4.3 (3.7–5.0)	10.5 (8.3–12.6)	6.1 (4.2–8.1)	
	21–30 y exp (*n* = 6)	5.0 (4.2–5.7)	10.1 (8.7–11.6)	5.2 (4.0–6.4)	
	31–40 y exp (*n* = 2)	5.2 (2.6–7.7)	10.0 (5.6–14.4)	4.9 (–2.1 to 11.8)	
Mean slides / h	Davey (2007b)	13.31 (12.25–14.36)	6.12 (5.76–6.49)	7.18 (6.17–8.20)	< 0.001

Abbreviations: CI = confidence interval, LBC = liquid-based cytology, min = minutes, NR = not reported.

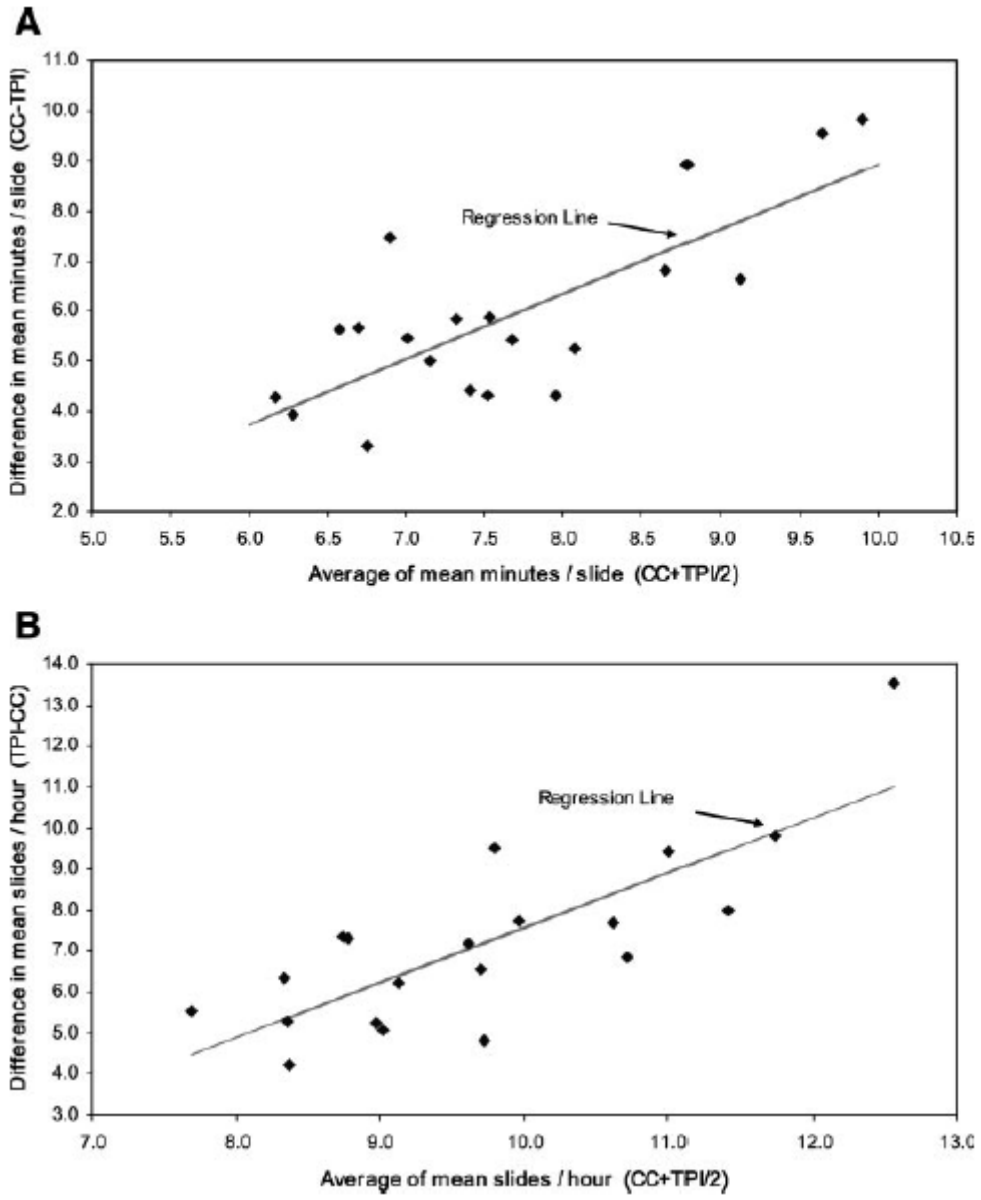


Figure 4 Relationship between cytologist reading speed and difference between conventional cytology and TPI-assisted LBC slide reading time.

Reproduced from Davey et al. (2007b).

Comparison to manual LBC slide reading

Included studies

The search for studies published since the MSAC 2003 review identified four primary studies of automated versus manual reading of LBC slides that met the inclusion criteria for review. Three studies investigated the TPI system (Biscotti et al. 2005; Bolger et al. 2006; Roberts et al. 2007) and one investigated the AutoPap system (Wilbur et al. 2002). One of the studies of TPI, by Roberts et al. (2007), provided a three-way comparison of automated and manual reading of LBC with manual reading of conventional slides. Bolger et al. (2006) reports two TPI studies. The first study, of 6000 slides, reports the accuracy of the TPI review of the 22 FOVs compared only with full manual screening. The second study, also of 6000 slides, reports the accuracy of TPI-assisted screening where full manual screening was conducted when any abnormality was detected by the cytologist in the 22 FOVs. As only the second of these studies reflects the use of the TPI in standard practice, only this study is included here.

Studies comparing manual with automation-assisted reading of LBC in independent (unpaired) samples were excluded unless all or a random sample of negative slides were verified against a reference standard (requirement for high-quality study in the systematic review by Davey et al. 2006; see Appendix I). These studies would provide a lower quality of evidence than that provided in the four included studies. Thirteen studies comparing independent (usually historical) samples were identified; none verified a sample of negative results.

Study characteristics, quality and applicability

In studies comparing different methods of reading the same slide, the use of adjudicated cytology as a reference standard should not bias the relative accuracy results (Irwig et al. 2004), as errors in sampling or slide preparation will affect estimates of accuracy in both methods equally. Any differences in accuracy between the methods will relate only to reader interpretation of the slide.

The characteristics, quality and applicability of the included studies are summarised in Table 22. All studies were considered of fair quality and limited applicability. Detailed discussion of the study by Roberts et al. (2007) is also provided above in Table 16 and the associated text.

All four studies excluded slides not analysed by the ThinPrep Imager from the analysis. In the ThinPrep Imager studies Roberts et al. excluded 3.7 per cent of slides, Bolger et al. 3 per cent, and Biscotti et al. 7.1 per cent for this reason. Wilbur et al. (2002) excluded 1.7 per cent of slides not processed by AutoPap. Therefore any difference in accuracy reported only applies to the proportion of slides that can be read by the automated-slide reading system. In all studies slide markings made were removed between the alternate arms of the study.

All three studies of ThinPrep Imager versus manual reading of ThinPrep slides were paired diagnostic accuracy studies. Biscotti et al. (2005) used a reference standard of adjudicated cytology for all positive slides, all discordant slides, and a random 5 per cent of negative slides. Bolger et al. (2006) used a reference standard of adjudicated cytology for all positive slides. Neither reported blinding of adjudication to test results. This may introduce verification bias. Roberts et al. (2007) used a reference standard of histology

for slides with a pHSIL or higher cytology result. There is incomplete verification in the study by Roberts et al. (2007): no histology was available for 9.0 per cent of positive TPI-read slides or 8.3 per cent of LBC slides read manually ($P > 0.05$, χ^2-test). The degree of verification of discordant positive results by each test was not reported.

The Bolger et al. study includes no seeded cases. The Biscotti et al. study includes seeded cases, although the total number of high-grade cases according to the reference standard is similar to that in Bolger et al. (2006). The data from the Roberts et al. study reported here exclude seeded cases, but there are fewer known high-grade cases on account of incomplete verification.

Biscotti et al. (2005) reported the sensitivity and specificity of the alternative screening strategies with different test thresholds for the equivalent reference standard threshold. The reported accuracy values for HSIL+ are sensitivities and specificities at detecting adjudicated HSIL+; those for LSIL+ were for adjudicated LSIL+; and those for ASCUS+ were for adjudicated ASCUS+. Only data for the accuracy of HSIL+ are included in this review. True and false positive data for lower test thresholds are calculated from the contingency table to reflect detection of adjudicated HSIL+.

Bolger et al. (2006) report the sensitivity and specificity of the different screening strategies. However, neither the test nor reference standard thresholds used for these values are reported, nor are the assumptions made about non-confirmed negative slides, nor can the reported values be reproduced. Therefore the reliability of these measures cannot be confirmed and these data are not considered in this review (values are provided in Appendix F for completeness). Contingency tables of test and reference standard results are provided, thus true and false positive findings and the estimated difference in sensitivity & specificity for alternative screening strategies for different thresholds have been calculated and are considered in this review.

Wilbur et al. (2002) described the use of the AutoPap system versus manual reading of AutoCyte Prep LBC slides in a study considered to be of fair quality. The AutoPap system was used to designate up to 25 per cent of slides as 'no further review' which then did not undergo any manual screening. The use of the AutoPap system in this study differs to from that specified for the Focal Point system in the clinical flow chart (**Error! Reference source not found.**), and as advised by the manufacturer. In the slides designated for review, the cytologist reviewed 10 fields of view indicated as most likely to show an abnormality. Where the cytologist identified a potential abnormality, these slides proceeded to full manual review. The 15 per cent of slides that proceeded to review and were determined by the cytologist to be within normal limits with the highest AutoPap ranking underwent QC full manual rescreening. In the manual screening comparison arm of the study, a random 10 per cent of slides also underwent QC rescreening. The spectrum of disease in the Wilbur et al. study has limited applicability to a screening population, as 19.7 per cent (251/1275) of slides were seeded, giving a prevalence of verified high-grade disease of 9.7 per cent, including 6.8 per cent carcinoma cases. This severely limits the applicability of the reported accuracy data to a routine screening population. This study was considered of highly limited applicability owing to the limitations of both the population and the index test. In this study the test and reference standard threshold for accuracy data are not clearly reported for sensitivity and specificity pairs. True positive cases for corresponding test and reference standard thresholds are reported. False positive cases are not clearly defined by consistent reference standard thresholds. Total cytological classifications for the separate screening strategies are not reported.

Table 22 Characteristics and quality of studies reporting on the relative accuracy of manual and automated reading of LBC slides.

Author, year Setting	Population (*N*, inclusion criteria)	Test comparison/s	Study design	Quality & applicability
Roberts et al. 2007 Australia One pathology laboratory Feb 2005 – April 2005	*N* = 11 416 + 103 seeded LBC cases Inclusion criteria • Routine samples • Cases HSIL, AIS or carcinoma (*N* = 103), these cases excluded from data reported here Exclusion criteria • Slides not read by TPI (3.7%)	Index test ThinPrep LBC with ThinPrep Imager reading Comparator test • Conventional cytology with manual reading • Manual reading of LBC (TP) samples	Study design: Three-arm diagnostic accuracy study with split-sample pairs and seeded cases. Data reported separately for routine and seeded cases. Test threshold: Test: HSIL/pHSIL (including possible glandular high grade). Reference standard: high grade Reference standard: Histopathology (within 9 months) for high-grade and possible high-grade lesions, no reference standard for 12/134 TPI (9.0%), 15/180 TPM (8.3%), 10/93 conventional (10.8%) slides None for low-grade positive, negatives Outcomes • Accuracy for squamous lesions: sensitivity [a], incremental TP (with and without seeded cases) • Test yield • Screening time	C1 P2 Q2 Accuracy study NHMRC level III-2 Quality: fair Prospective, not consecutive, reference standard valid but not applied to all discordant participants. Excluded slides not read by imager Applicability unclear Applicable index test & comparator, population applicability unclear (likely applicable, data for routine cases reported separately)
Bolger et al. 2006 Ireland One clinical site April 2004 – March 2005, second half of cases	*N* = 6000 (second study) Inclusion criteria • routine samples • likely screening and diagnostic population (unclear) Exclusion criteria slides not read by TPI (approx 3%)	Index test ThinPrep LBC with ThinPrep Imager reading Comparator test • Manual reading of LBC (TP) samples	Study design: Paired diagnostic accuracy study Test threshold: Test: NR for sensitivity & specificity Reference standard: NR for sensitivity & specificity Contingency data provided Reference standard: Positive and discordant cases: adjudicated cytology (one senior cytologist + one cytopathologist) Negative cases: nil Outcomes • Accuracy (reference standard threshold NR): sensitivity & specificity [b]; categorical contingency data enabling relative TP and FP rates to be calculated for different thresholds • Test yield • Unsatisfactory rates	C1 P2 Q2 Accuracy study NHMRC level III-2 Quality: fair Prospective, not consecutive, reference standard valid (suboptimal but acceptable for this comparison) applied to all positive and discordant participants. Excluded slides not read by imager Applicability unclear Applicable index test & comparator, population applicability unclear

Biscotti et al. 2005 US 4 clinical sites Dec 2000 – July 2001	*N* = 10 742 Inclusion criteria • Routine screening & diagnostic population (10 359) • Seeding with HSIL cases (NR, likely 383, 3.6%) Exclusion criteria • Slides not read by TPI (7.1%) • No consensus diagnosis on adjudicated review (6/361 concordant slides)	Index test ThinPrep LBC with ThinPrep Imager reading Comparator test • Manual reading of ThinPrep LBC slides (2000 or 3000 processor)	Study design: Masked, paired diagnostic accuracy study with seeded cases. Test threshold: Test: HPV or reactive cell changes, epithelial abnormalities or malignant cells Reference standard: ASCUS+. Contingency data provided. Reference standard: Positive: Cytological adjudication (3 independent pathologists) for all concordant ASCUS+, and discordant ≥ 1 level Negative: Cytological adjudication for 5% concordant negative Adequacy: Independent cytologist for all concordant unsatisfactory, discordant adequacy, random 5% of concordant satisfactory or 'satisfactory but limited by' Outcomes • Accuracy: sensitivity & specificity; categorical contingency data for relative TP and FP rates • Productivity: average slides screened per day and time expended • Test yield • Unsatisfactory rates	C1 P2 Q2 Accuracy study NHMRC level III-2 Quality: fair Prospective, not consecutive, reference standard valid (suboptimal but acceptable for this comparison) applied to all positive, discordant and 5% negative participants. Excluded slides not read by imager Applicability limited Applicable index test & comparator, population applicability limited owing to seeded cases
Wilbur et al. 2002 US 1 cytopathology laboratory Recruitment period NR	*N* = 1275 Inclusion criteria • 1049 retrospective, consecutive routine samples + 251 seeded cases • Likely screening and diagnostic pop (unclear) Exclusion criteria • Slides not read by AutoPap (1.7%) • No AutoPap review (*n* = 2) • Slide read twice (*n* = 1)	Index test AutoPap reading of AutoCyte PREP slides (SlideWizard 2), including designation as no further review, review slides indicate 10 FOVs. Comparator test • Manual reading of AutoCyte PREP slides	Study design: Paired sample, masked diagnostic accuracy study, with retrospectively selected samples Test threshold: Test: NR. Reference standard: NR (likely ASCUS+) Reference standard: Discordant and positive: Cytological adjudication (1 cytopathologist) Negative cases: Nil Adequacy discrepancy: Cytological adjudication (1 senior cytotechnologist) Outcomes • Accuracy (reference standard threshold NR): sensitivity & specificity [b]; categorical contingency data enabling relative TP and FP rates to be calculated for different test thresholds • Proportion slides designated 'no further review'	C1 P2 Q2 Accuracy study NHMRC level III-2 Quality: fair Prospective, not consecutive, reference standard valid (suboptimal but acceptable for this comparison) applied to discordant positive participants. Excluded slides not read by AutoPap Applicability: limited Index test limited applicability Applicable comparator Population applicability limited owing to many seeded cases

[a] Reported sensitivity invalid as not based on full reference standard. [b] Test & reference standard threshold not reported.
Abbreviations:, LBC = liquid-based cytology, NR = not reported, TP = ThinPrep, TPI = ThinPrep Imager, TPM = ThinPrep manual.

Test accuracy

Table 23 summarises the true and false positive rates of Thin-Prep Imager-assisted reading of ThinPrep slides compared to manual reading of ThinPrep slides.

The data from Roberts et al. (2007) and Bolger et al. (2006) do not include seeded cases. The data from Biscotti et al. (2005) includes seeded HSIL cases, although the known prevalence is not different to that in Bolger et al. (2006). The data from Roberts et al. (2007) has incomplete verification, therefore the data from the Bolger et al. (2006) study is considered the most reliable.

There were no statistically significant differences in the number of true positive cases detected by the ThinPrep Imager system and manual reading of LBC slides in any of the three studies (Table 23). There was not a consistent trend across studies, for true positive cases, either. In two of three studies the ThinPrep Imager detected slightly fewer true positive cases than ThinPrep slide with manual reading; In one study with seeded cases the ThinPrep Imager detected slightly more true positives (Bolger et al. 2006).

In the studies by Roberts et al. (2007) and Biscotti et al. (2005), the ThinPrep Imager detected significantly fewer false positive cases than manual reading of ThinPrep, for a pHSIL and HSIL threshold, respectively (Table 23). Although Biscotti et al. (2005) includes seeded cases, these are all HSIL cases thus this cannot affect the false positive rates in the study.

There was significant heterogeneity between studies for the difference in true positive and false positive findings between ThinPrep Imager and manual reading of TP slides ($P < 0.001$).

Bolger et al. (2006) reported the accuracy of test classification to detect adjudicated glandular lesions. There were 2 glandular lesions according to adjudicated review. These were classified as glandular by both automation-assisted reading and manual reading.

One included study of AutoPap-assisted reading of LBC slides by Wilbur et al. (2002) includes a high proportion of seeded cases, Therefore the spectrum of disease in this study is likely to differ greatly from that in a screening population. In addition, the use of the AutoPap system does not reflect current recommendations of use of the FocalPoint system. The data are considered of highly limited applicability. Whilst AutoPap detected an additional 30 true positive cases for adjudicated HSIL, with a test threshold of HSIL, compared to manual screening (AutoPap 103 TPs, manual 73 TPs; $P = 0.02$), the false positive rates for the same reference standard and test threshold were not clearly reported. Given these major limitations, the data are of limited value for informing the relative accuracy of AutoPap versus manual reading of AutoCyte LBC slides.

Table 24 displays the estimated differences in sensitivity and specificity of ThinPrep Imager reading of ThinPrep slides compared to manual reading of ThinPrep slides. These measures cannot be calculated from the study by Roberts et al. (2007) as not all discordant positive slides were verified. The differences for the HSIL threshold from Biscotti et al. (2005) are as reported in the study. All values from Bolger et al. (2006) are calculated from the contingency table provided in the study, as described in 'Statistical methods', page 23.

The estimated differences in sensitivity and specificity determined from Bolger et al. (2006) were not consistent with the differences reported in Biscotti et al. (2005). Statistical comparisons were performed and reported for the number of true and false positive cases detected (Table 23).

Table 23 Estimates of the relative true & false positive rates of Thin Prep Imager–assisted versus manual reading of LBC for CIN 2+.

Study *N*	Min prev %[c]	Test threshold	Automation-assisted reading, additional positives per 1000 screened			Positives						χ^2-test, cases *P*	
						Automation-assisted reading			Manual reading				
			TP	FP	TP:FP	TP	FP	TP:FP	TP	FP	TP:FP	TP	FP
Without seeded cases													
Roberts et al. 2007	0.95	pHSIL/ HSIL, incl glandular	–1.31	–2.45	–1 : –1.87	80	42	1:0.53	95	70	1:0.74	0.26	< 0.001
11 416		pHSIL/ HSIL, excl glandular	–1.14	_	_	78	_	_	91	_	_	0.32	_
Bolger, 2006	1.4	HSIL [a]	–0.83	–0.33	–1 : –0.4	69	5	1:0.07	74	7	1:0.09	0.67	0.56
6000		CIN 1	–0.17	–0.83	–1 : –5.0	80	370	1:4.63	81	375	1:4.63	0.94	0.85
		pLSIL [b]	0	–0.83	–	82	706	1:8.61	82	711	1:8.67	1.0	0.89
Including seeded cases													
Biscotti, 2005	1.3[d]	HSIL, excl AGUS	0.74	–1.68	1 : –2.3	111	32	1:0.29	103	50	1:0.49	0.58	*P* < 0.01
10 742[e]		LSIL, excl AGUS	0.37	–1.12	1 : –3.0	128	253	1:1.98	124	265		0.80	0.59
		ASCUS, excl AGUS	0.47	2.51	1 : 5.4	137	618	1:4.51	132	591		0.76	0.42
		AGUS alone	0	0	–	2	0	–	2	0	–	1.0	–

Abbreviations: bx = biopsy, FP = false positive, min = minimum, prev = prevalence, ref std = reference standard, TP = true positive, TPI = ThinPrep Imager.

[a] CIN 2. [b] Borderline nuclear abnormality. [c] Reference standard positives. [d] Includes seeded cases. [e] Total slides unclear: Methods refers to 10 742, Results to 10 359.

Biscotti et al. (2005) reported sensitivity and specificity for HSIL+ test diagnoses to detect adjudicated cytology of HSIL+. In this study, ThinPrep Imager reading of LBC had a sensitivity of 79.9 per cent (72.2%–86.2%) and a specificity of 99.6 per cent (99.5%–99.7%) compared to manual reading of LBC slides with a sensitivity of 74.1 per cent (66.0%–81.2%) and a specificity of 99.4 per cent (99.2%–99.6%). This equates to a 5.8 per cent (–1.1% to +12.6%) higher sensitivity and 0.2 per cent (0.06%–0.4%) higher specificity of the ThinPrep Imager, however these differences were not statistically different. The study also reported the sensitivity and specificity of LSIL+ diagnoses for detecting LSIL+ adjudicated cytology and of ASCUS+ diagnoses for detecting ASCUS+ adjudicated cytology (see Appendix F). The current review, however, is concerned with the accuracy of different slide reading methods at detecting HSIL+ at different test thresholds, and thus the differences in sensitivity and specificity for this reference standard threshold have been calculated from the contingency table provided in the publication (Table 24).

Table 24 Estimates of the relative accuracy of manual versus ThinPrep Imager-assisted reading of LBC for HSIL/CIN 2+.

Study *N*	Test threshold	Accuracy difference: automated – manual			
		Sensitivity		Specificity	
		%	(95% CI)	%	(95% CI)
Biscotti, 2005	HSIL [c]	5.8	(–1.1, 12.6)	0.2	(0.06, 0.4)
10 742	LSIL [c]	2.9		0.13	
	ASCUS [c]	3.6		–0.29	
Bolger, 2006	HSIL [a]	–6.0		0.03	
6000	CIN 1	–1.2		0.09	
	pLSIL [b]	0		0.09	

[a] CIN 2+. [b] Borderline nuclear abnormality. [c] Excluding 10 adjudicated AGUS cases .
Abbreviations: CI = confidence interval.

Process outcomes

The cytological classification of slides by automation-assisted reading and manual reading is summarised in Table 25. In the study by Roberts et al. (2007), significantly fewer slides were classified as pLSIL and pHSIL than by manual reading of LBC slides. In the other studies, LBC slide classification was similar between reading methods.

In the study by Wilbur et al. (2002) of the AutoPap screening of AutoCytePrep slides, 1.7 per cent of cases could not be processed by the automated screening system. The AutoPap system designated 17.1 per cent of slides as 'no further review' which were classified as within normal limits without any cytologist screening. Unsatisfactory and total cytological classifications were not reported according to screening method used.

Table 25 Process outcomes of automation-assisted versus manual reading of cervical LBC samples.

Study	*N*	TPI not processed (%)	Unsat (%)		Reading time (min)		Automation-assisted reading yield (%)							Manual reading yield (%)						
			Auto	Man	Auto	Man	pLSIL	CIN 1	LSIL	pHSIL	HSIL	AIS/SCC	Gland	pLSIL	CIN 1	LSIL	pHSIL	HSIL	AIS/SCC	Gland
TPI-assisted versus manual reading of TP slides																				
Roberts et al. 2007	11 416	3.7	–	–	3.42	4.71	3.8**	–	2.9	0.43**	0.74	–	–	4.5	–	2.7	0.72	0.86	–	–
Bolger et al. 2006	6000	~3	1.05	1.07	–	–	5.7[a]	6.3			1.23	0	0.03	5.6[a]	6.3			1.35	0	0.03
Biscotti et al. 2005	10 359	6.8[c]	0.27[c]	0.61[c]	3.2	6.7	3.90[b]		2.47		1.44	0.04	0.10	3.49[b]		2.45		1.54	0.06	0.12
AutoPap-assisted versus manual reading of AutoCytePrep slides																				
Wilbur et al. 2002	1275	1.7	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

Abbreviations: Gland = glandular, Unsat = unsatisfactory.
[a] Borderline nuclear abnormality. [b] ASCUS. [c] Assuming total of 10 742.
***P* < 0.01 vs manual reading.

Direct comparison of LBC, automated LBC reading and conventional Pap cytology

One study (Roberts et al. 2007) provided a direct comparison of all three alternative screening technologies in 11, 416 cervical samples. Roberts et al. (2007) reported the relative performance of ThinPrep cytology with manual reading, ThinPrep Imager reading of ThinPrep slides and manual reading of conventional cytology slides. The characteristics of this study are summarised in Table 16. The study is a split-sample paired diagnostic accuracy study in which 89.0 per cent of positive ThinPrep Imager read slides, 91.7 per cent of LBC slides read manually and 89.2 per cent of conventional slides with a pHSIL/HSIL cytology result had a histological reference standard. The data from this study comparing LBC and conventional cytology did not meet the inclusion criteria for the systematic review by Arbyn et al. (2008) and are therefore not included in the data presented on the relative performance of LBC and conventional cytology above (from page 25). The results from this study are summarised in Table 27.

This head-to-head comparison of the three alternatives demonstrated that manual reading of ThinPrep slides detected 2.8 more cases of high-grade histology per 1000 women screened than conventional cytology, a significant increase in the detection rate (95/11 416 vs 63/11 416 cases, $P = 0.01$, χ^2-test). Automated reading of LBC slides detected 1.5 additional cases per 1000 women screened, but this difference was not statistically significant.

These results are difficult to interpret in the context of the evidence presented above that LBC was not more accurate than conventional cytology. There was no significant difference in the relative sensitivity of LBC over that of conventional cytology when all studies were pooled in Arbyn et al.'s (2008) meta-analysis, or when only the Italian RCT was considered, at up to three different threshold levels for test positivity. Studies comparing automated LBC with manual LBC did not provide clear evidence of a difference between these approaches either (see Table 23).

The evidence presented in the section comparing manual LBC to conventional cytology (from page 25) comes from a systematic review including non-randomised studies in which all patients received appropriate verification of positive cytology results, and a large randomised trial in a screening setting of 45 174 women. In the RCT, a reference standard of colposcopy ± biopsy was available for 91 per cent of women receiving CC and 93 per cent receiving LBC, and a cytology result of pLSIL (ASCUS) was used as the test positivity threshold. In both arms of the RCT the median age of the women was 41 years, and 49 per cent had had a cervical cytology test in the previous 4 years. This RCT had over 80 per cent power to detect a 50 per cent increase in sensitivity (relative detection of 1.5), indicating that it was not adequately powered to demonstrate significant differences in smaller increases in sensitivity.

In the Roberts et al. (2007) study, high-grade lesions were detected in 0.55 per cent (63/11 416) of slides in the conventional arm. In the Italian RCT, 0.37 per cent (84/22466) of women in the conventional arm were diagnosed with CIN 2+ histology.

In the three-armed study by Roberts et al. (2007) the number of false positive findings per true case detected was significantly higher in manual reading of LBC than conventional cytology (1:0.74 vs 1:0.32, $P = 0.005$, χ^2-test) but did not significantly differ

between automated reading of LBC and conventional cytology (1:0.53 vs 1:0.32, P = 0.11, χ^2-test).

Correspondence with an author of the Davey et al. (2007a) study confirmed that there were no data available for a three-way comparison.

The three-way direct comparison for process outcomes reported in Roberts et al. is presented in Table 26. These data have been presented separately in the relevant sections above. Manual reading of ThinPrep slides classified more slides as possible high-grade lesions than manual reading of conventional slides (P = 0.023, χ^2-test).

Table 26 Process outcomes of LBC, automation-assisted reading of LBC and manual reading of conventional cervical cytology (Roberts et al. 2007).

Outcome		ThinPrep Imager	Manual reading ThinPrep slides	Manual reading conventional slides
Yield	pLSIL	3.8%#	4.5%	1.8%
	LSIL	2.9%#	2.7%	1.6%
	pHSIL	0.4%	0.7%*	0.3%
	HSIL	0.74%	0.86%	0.48%
Mean reading time		3.42 min	4.71 min	7.40 min

* P = 0.023 vs conventional slides; #P < 0.001 vs conventional slides.

Ongoing trials

One ongoing trial of automated LBC slide reading, HPV testing and conventional cytology is being conducted in Finland (ISRCTN23885553). This trial has already reported relative rates of detection of cervical pathology, most recently among 777 144 women (Nieminen et al. 2007). Results of interval cervical cancer incidence are expected to be reported over the period 2007–2015 (Anttila et al. 2006). Results from this trial are not included in the current review, as the technology used (PapNet) is no longer commercially available.

The UK National Institute for Health Research is currently conducting an RCT of manual LBC screening with automated primary screening technologies (MAVARIC; ISRCTN66377374). This trial will compare the diagnostic accuracy, unsatisfactory rates and productivity of manual screening and FocalPoint, Pathlore or Cytyc Imager (ThinPrep) reading of LBC, with and without HPV triage of low-grade positive cytology results. The trial will also include a cost-effectiveness analysis. The trial began in 2005 and results are expected to be published in mid 2010.

Table 27 Direct comparison of the relative accuracy of LBC, automation-assisted reading of LBC and manual reading of conventional cervical cytology for CIN 2+ in *N* = 11 416 routine samples (Roberts et al. 2007).

Test threshold	Automation-assisted reading LBC, additional positives per 1000 screened			Manual reading LBC, additional positives per 1000 screened			Total positives								
							Automation-assisted reading LBC			Manual reading LBC slides			Manual reading conventional slides		
	TP	FP	TP:FP	TP	FP	TP:FP	TP	FP	TP:FP	TP	FP	TP:FP	TP	FP	TP:FP
pHSIL/ HSIL, including glandular	1.49	1.93	1:1.29	2.8	4.38	1:1.56	80	42	1:0.53	95*	70	1:0.74	63	20	1:0.32
pHSIL/ HSIL, excluding glandular	1.58	_	_	2.7	_	_	78		_	91*	_	_	60	_	_

Abbreviations: FP = false positive, TP = true positive.

* $P < 0.05$ χ^2-test for (i) comparison of automated LBC vs manual LBC vs manual Pap test; and (ii) comparison of manual LBC vs conventional cytology.

Conclusions

A modelled analysis of cervical cancer screening, diagnosis and treatment in the Australian setting is necessary to explore the potential long-term benefits and trade-offs of these technologies in the Australian setting.

Manual LBC

High-quality systematic reviews and a large randomised trial have indicated that liquid-based cytology compared to conventional cytology

- provides no statistically significant increase in sensitivity or specificity
- provides no statistically significant difference in sensitivity (HSIL, LSIL or pLSIL thresholds) or specificity (HSIL or LSIL thresholds) for the detection of CIN 2+
- reduces the specificity for the detection of CIN 2+ at a test threshold of pLSIL
- classifies more slides as positive for low-grade lesions
- reduces the rate of unsatisfactory smears

Automated LBC

There is no evidence of an advantage, disadvantage or equivalence of the accuracy of the Focal Point system compared to conventional cytology.

The ThinPrep Imager system compared to manual reading of conventional cytology

- significantly decreases slide reading time
- reduces the rate of unsatisfactory smears
- detects at least as many CIN 2+ lesions as conventional cytology, and may detect more
- classifies more slides as positive for low-grade lesions

It is unclear whether any increase in detection of high-grade lesions with the ThinPrep Imager system is attributable to LBC alone, to the automation-assisted reading system, or a combination of both.

What are the economic considerations?

Background

Economic evaluation of new health care technologies is particularly important when the new technology offers health benefits at additional cost. It is clear that there will always be a limit to the additional cost which would be paid for a given health gain. Economic evaluation is generally aimed at determining whether such incremental costs represent value for money.

The usual process for an economic evaluation is first to consider the additional benefits of the new technology relative to the comparator (the incremental effectiveness), and then to determine cost differences between the two (the incremental costs). Effectiveness is measured in clinically appropriate natural units or by a multidimensional measure such as quality-adjusted life years (QALYs). When both costs and effects are known, then an incremental cost-effectiveness ratio (ICER) can be determined as:

$$\text{ICER} = \frac{Cost_{NEW} - Cost_{COMPARATOR}}{Effectiveness_{NEW} - Effectiveness_{COMPARATOR}}$$

Existing literature

An overview of existing economic evaluation studies of LBC and automation-assisted LBC was undertaken, in addition to a primary modelled economic analysis of the cost-effectiveness of these technologies in the Australian setting.

One primary economic study of manual LBC identified in the systematic literature review was considered applicable to the research question (Neville & Quinn 2005). In addition, a summary of the appraisal of economic studies of LBC included in the most recent HTA report (Krahn et al. 2008) has been provided.

A single study of automated LBC included in the systematic review of comparative accuracy data also provided a costing study.

LBC—Australian study

Many modelled health economic analyses of the cost-effectiveness of LBC have been conducted and published since the 2002 MSAC review. Only one of these applied to Australia (Neville & Quinn 2005). This study applied alternate sources of data measuring the accuracy of ThinPrep to a model from the MSAC (2002b) review. An MSAC critique of the data used has been published (Blamey et al. 2006). A Canadian HTA report rated the quality of this study as low (3/7; Krahn et al. 2008). In summary, the data used in this analysis was not more recent that that included in the MSAC (2002) analysis, and was considered less robust.

LBC—systematic review

The HTA by CADTH (Krahn et al. 2008) included the most recent systematic literature review of economic studies reporting the cost-effectiveness of LBC versus CC in developed screening programs. Nine studies published to June 2006 were identified. The results of these studies, as described by Krahn et al. (2008), are provided below.

One of the studies was the Australian publication discussed above (Neville & Quinn 2005). Five studies published over the period 1999 to 2002 were conducted in the United States, two 2004 studies were conducted in the United Kingdom, and one 2005 study came from Alberta, Canada. The results of these studies will not be applicable to the Australian setting, as costs and screening programs (eg, screening intervals and ages, participation rates and baseline unsatisfactory rates) differ between settings.

Two studies assumed equivalent sensitivity, and the other seven assumed a marginal sensitivity gain of between 3 and 25 per cent for using LBC over conventional cytology. Two industry-funded studies using a sensitivity gain for LBC of >20 per cent demonstrated a favourable cost-effectiveness of LBC (Hutchinson et al. 2000; Montz et al. 2001). A high-quality study by the US Agency for Healthcare Policy Research (McCrory et al. 1999; Myers et al. 2000) did not provide a recommendation owing to a lack of data. Of the remaining five high-quality studies (rated 5.5/7 to 6/7), two were conducted in the United States, two in the United Kingdom and one in Canada. All reported cost-effectiveness of LBC at 1-year (Canadian study) to 5-year (3 other studies) intervals.

The Canadian analysis (Lier & Jacobs 2005) used an unsatisfactory rate of 0.4 per cent for LBC and 1.0 per cent for conventional cytology; with LBC sensitivity 8.4 per cent greater than conventional, equivalent specificity and a marginal cost of LBC of CAN$3.61. The use of LBC at a 1-year screening interval was cost-effective at CAN$20 000 per life year gained. This is most likely driven by the favourable test accuracy characteristics assumed (in particular, equivalent specificity), which are not supported by the findings of the current systematic review (see 'Liquid-based cytology', page 25).

One UK study (Sherlaw-Johnson & Philips 2004) assumed equivalent sensitivity and a conventional cytology unsatisfactory rate of 10 per cent, with an LBC unsatisfactory rate of 5 per cent and an LBC cost of £0.76 lower than CC. It found that strategies of screening 3-yearly and 5-yearly dominated and had a reduction in costs. However, neither the marginal cost nor the baseline unsatisfactory rate reflects that in the Australian setting (see 'Discussion', page 100). The other (Karnon et al. 2004) was an HTA based on results of the UK pilot study, which found unsatisfactory rates of 9.1 per cent for conventional cytology and 1.6 per cent for LBC. This analysis was based on a marginal sensitivity for LBC of 3 to 5 per cent, a marginal cost of between £1.50 and £3.92, and screening intervals of 2 to 5 years. It also found LBC to be cost-effective, more so at a 3-yearly interval than 5-yearly. This analysis also differs greatly from the Australian setting owing to the large reduction in unsatisfactory rates seen with the use of LBC.

The two US studies found LBC to be cost-effective at a 3-yearly screening interval. They used an LBC gain in sensitivity of 15 per cent (Brown & Garber 1999) and 9 per cent (Maxwell et al. 2002) and a marginal LBC cost of $US9.75 and $US5.00, respectively. The unsatisfactory rates were not reported. Maxwell et al. analysed 1-, 2- and 3-yearly intervals; Brown and Garber analysed 1-, 2-, 3- and 4-yearly intervals. Brown and Garber found that LBC primary screening was not cost-effective, but that 3- or 4-yearly intervals were more cost-effective than annual screening. Maxwell et al. found LBC to be cost-effective at a 3-yearly interval, and that the ICER rose sharply at more frequent screening intervals, particularly with higher test sensitivities. The latter study found that test cost and increased likelihood of detecting transient LSIL are key determinants, particularly if an increase in sensitivity is accompanied by a decrease in specificity.

Krahn et al. (2008) commented that in general studies indicated shorter screening intervals were generally less cost-effective than longer screening intervals. Marginal sensitivity gains with LBC were partly responsible for an increase in life expectancy at longer screening intervals.

Sensitivity analyses in the studies indicated that LBC was more cost-effective when the sensitivity of LBC and conventional cytology was low. The results were sensitive to changes in screening compliance and LBC cost.

Since the publication of this HTA, a study of the resource use in detecting CIN 2+ by LBC compared to conventional cytology in the Italian RCT has been published (Giorgi-Rossi et al. 2007). This cost study is based on data from 28 000 women aged 35 to 60 years, enrolled in the Italian RCT of conventional cytology versus LBC (ThinPrep) plus HPV testing. In this study, Women with ASCUS+ cytology were referred for colposcopy. The unsatisfactory rates were 3.7 per cent for conventional Pap and 2.4 per cent for LBC, and the relative sensitivity of LBC was 1.03. This study found that 'the unit cost of LBC used alone should be less than that of a conventional Pap . . . to result in the same overall cost per CIN 2+ detected as screening by CC'. In Australia, the unit cost of colposcopy + biopsy is approximately 15× that of conventional Pap. According to the results of the study, this would mean that the cost of LBC alone would need to be <0.9× the cost of conventional Pap (<$17.64) to give the same cost per CIN 2+ detected (ie, equivalent cost-effectiveness).

LBC—primary cost-effectiveness analysis in Ontario, Canada

The CADTH report (Krahn et al. 2008) also included a primary modelled cost-effectiveness analysis of LBC compared to conventional cytology screening of women aged 18 to 70 years. This model assumed the use of LBC was associated with a 6.4 per cent gain in sensitivity and a 4.0 per cent loss of specificity, based on a meta-analysis of 20 head-to-head studies of varied quality. Unsatisfactory rates of 0.24 per cent for LBC and 0.58 per cent for CC were assumed on the basis of an Ontario study. Biennial screening with LBC compared to conventional cytology cost approximately $CAN31 000 per life year saved (LYS) or $CAN29 000 per QALY gained. Sensitivity analyses indicated that if LBC was associated with a 7.7 per cent increase in sensitivity and a 4.4 per cent loss of specificity (based on a meta-analysis of high-quality studies; no requirement of histology as reference standard), the cost-effectiveness of LBC would be $17 000 per LYS. If LBC was associated with a 1.1 per cent gain in sensitivity and 0.6 per cent loss of specificity (based on 13 studies with histology as a reference standard), the cost-effectiveness of LBC would be $298 000 per LYS. Unsatisfactory rates were not tested in a sensitivity analysis. Annual LBC screening had an incremental cost of CAN$147 000 per LYS, or CAN$149 000 per QALY, compared to LBC screening every 2 years. Thus, while the main analysis found LBC to be cost-effective compared to conventional cytology, this was highly variable and dependent upon the accuracy measures assumed. Also, the impact of unsatisfactory rates was not explored.

Automated slide reading

A study included in the systematic review of effectiveness (Confortini et al. 2003) reported the cost per CIN 2+ case detected using manual vs AutoPap-assisted reading of conventional slides. It did not provide evidence of an accuracy advantage, disadvantage or equivalence. Therefore, the effectiveness evidence for this system is not considered adequate to form the basis of an informative economic analysis. In addition, the costing study considered the use of the AutoPap system in designating slides as NFR. As the

current standard use of the FocalPoint system does not involve allocating slides as NFR (see **Error! Reference source not found.**), these findings are not considered relevant to this assessment.

No other economic studies of automation-assisted slide reading more recent than the previous MSAC (2003) review were identified. A UK HTA report was published in 2005 (Willis et al. 2005), but it included a systematic review and economic analysis based on evidence to the end of 2000 only. It concluded that there was insufficient evidence to recommend implementation of automated image analysis systems at the time.

Economic evaluation

Modelled evaluation

Published data were used to construct a model of HPV natural history and cervical cancer screening in Australia. This model was used to evaluate the cost-effectiveness ratios of manual and automation-assisted reading of LBC.

Both manual and automated LBC were predicted to reduce cancer cases and deaths:

- Manual LBC could avert 23 cancer cases and 6 cancer deaths per annum.
- Automated LBC could avert 68 cancer cases and 19 cancer deaths per annum, although as favourable assumptions were made about test characteristics, these could be overestimates.

Both manual and automated LBC were predicted to increase investigations and treatments:

- Manual LBC could result in an additional 22 763 smears, 6770 colposcopies, 3273 biopsies, and 735 treatments for CIN 2/3 per annum.
- Automated LBC could result in an additional 38 346 smears, 10 788 colposcopies, 5154 biopsies, and 1751 treatments for CIN 2/3 per annum.

Incremental cost-effectiveness rations were calculated for both manual and automated LBC compared to current practice. At the currently requested level of reimbursement, these are high for both test technologies.

- Manual LBC was associated with a cost of $126 315 to $385 982 per LYS, depending on the level of reimbursement for LBC.
- Automated LBC was associated with a cost of $194 835 per LYS.

Both manual and automated LBC are estimated to increase total screening and treatment costs:

- Manual LBC is estimated to increase screening costs by $7.3 million to $23.6 million per annum, depending on the level of reimbursement for manual LBC.
- Automated LBC is estimated to increase annual screening costs by $37.4 million.

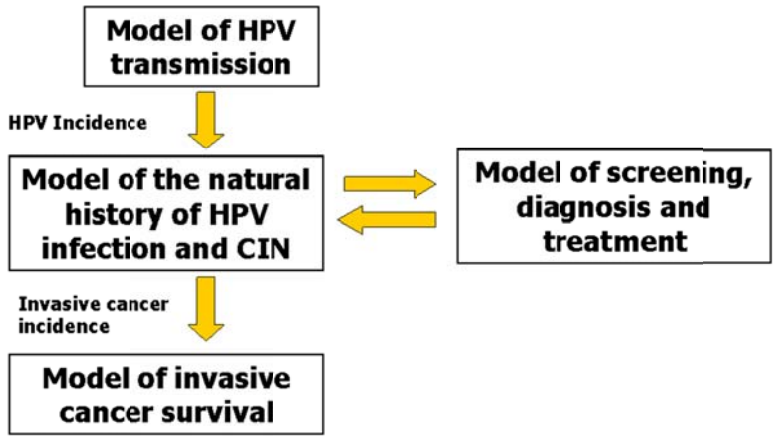
Introduction

The new cytology technologies under assessment have higher test costs, but also have some improvements in test performance. If either were used to replace conventional

cytology, the additional program costs would need to be recovered in full or in part via a reduction in the rate of unsatisfactory smears or an improvement in test performance.

It is complex to predict the potential savings from test characteristics alone, and for this reason a more detailed economic evaluation was conducted using a set of linked models to simulate (i) sexual behaviour and HPV transmission in Australia; (ii) the natural history of CIN and invasive cervical cancer; and (iii) screening, diagnosis and treatment according to practice in Australia (Figure 5). These models were based on previously published work (Canfell et al. 2004; Smith et al. 2008) and on a screening model developed for the NZ National Screening Unit (Canfell et al. 2008).

Figure 5 Linked models used to perform economic evaluation.



These models required parameterisation with population, screening, treatment and cost data specific to Australia. Various tests involved in the screening and treatment pathways were also characterised. There may be some variation in these across different settings, so where possible, local data were preferred. Data are also required to model the underlying processes involved in the transitions between health states of HPV infection, cervical precancer (CIN) and cervical cancer. Data sources are described in the following section, Summary of data sources for the baseline evaluation model.

The models were used to estimate the incremental cost-effectiveness ratios (ICERs) of (i) automated reading and (ii) manual reading of LBC slides in comparison to current practice of manual reading of conventionally prepared Pap smear cytology samples. The ICERs were based on the lifetime costs and effects of each strategy. These lifetime outcomes were calculated with a cohort model which ran from age 10 until age 84. Life years were the primary outcome measure, but health care resource usage was also predicted. All other screening practices, such as the time between screening tests and the management of abnormal cytology, reflected current practice, taking into account compliance. The analysis used a health services perspective. Future costs and outcomes were discounted at 5 per cent. One-way sensitivity analyses were performed on those parameters for which there was substantial uncertainty, or which could have had a significant influence on the results.

Summary of data sources for the baseline evaluation model

HPV incidence

Age-specific HPV incidence was obtained from a dynamic model of HPV transmission in Australia, which used survey data on sexual behaviour in the community to predict HPV incidence by single years of age. The dynamic transmission model and its parameterisation and calibration are fully described in (Smith et al. 2008). For the current evaluation, we assumed that HPV incidence reflected an unvaccinated population.

CIN natural history parameters

The parameters used for the natural history model were derived from a review of the international literature in 2004 (Canfell et al. 2004) and updated in 2006–07 in the course of a project for the NZ National Screening Unit (Canfell et al. 2008).

Invasive cancer parameters

To model progression of undiagnosed invasive cervical cancer, we adapted previous estimates of the rates of progression between undiagnosed FIGO disease states (Myers et al. 2000). We also calculated rates of developing symptomatic and therefore diagnosed disease for each possible disease extent from Myers et al. (2000).

Cancer survival parameters were based on cumulative relative survival ratios for invasive cervical cancer by extent of disease from the NSW Central Cancer Registry (NSW CCR) for women diagnosed with cervical cancer between 1991 and 2000, and followed up to 2001 (personal communication, Dianne O'Connell, Cancer Council NSW).

A more detailed description of parameter derivations is presented in Appendix J Detailed model data sources.

Hysterectomy rate

Data on the annual age-specific probability of having a hysterectomy were derived from the 2001 and 2005 National Health Surveys (ABS 2002; AIHW 2005).

Mortality rate

To model deaths from causes other than cervical cancer, we used data for all-cause mortality by 5-year age groups (ABS 2007; AIHW 2007a) and subtracted the cervical cancer mortality rate (AIHW 2008).

Compliance with screening and management recommendations

The model incorporated information on compliance with screening and management recommendations obtained via analysis of data from the Victorian Cervical Cytology Register (VCCR). Data on cytology and histology for women from Victoria were extracted from the VCCR for the period 1995–2007. Details are provided in Appendix J Detailed model data sources.

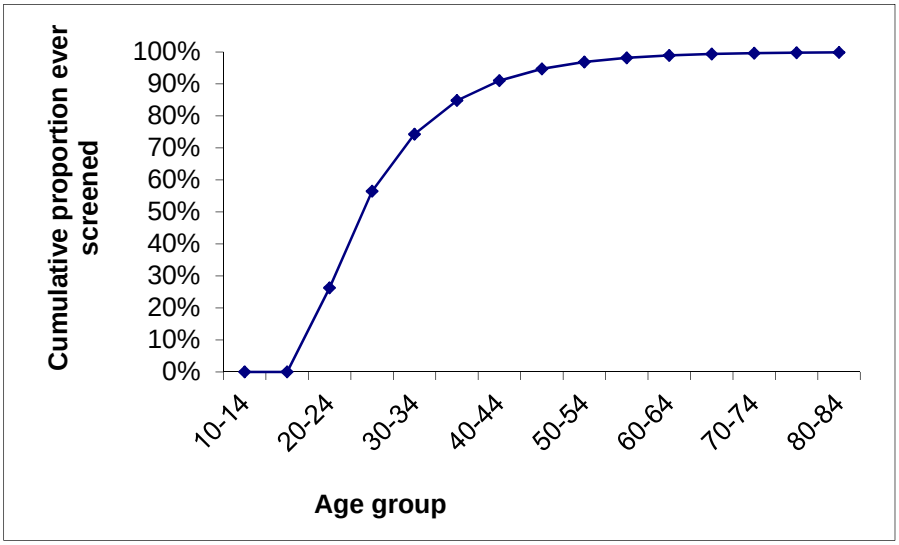
Age at screening initiation

We did not use direct registry data on age of first attending for cervical screening, as it is influenced by the fact that the organised screening program was not in place when some older cohorts initiated screening. In particular, registry data on women who apparently first attend screening after the age of around 35 years will reflect cohort effects in women with different exposure to the organised program during their lifetime. Therefore we used the following approach.

The proportion of women who initiate screening under the age of 20 was based on information from the VCCR. From age 20, the additional proportion initiating screening each year was chosen to be consistent with the age-standardised rate of women aged 20 or more who are never screened (ABS 2002; ABS 2006; AIHW 2008), and with 2-year participation in 20- to 24-year-olds (AIHW 2007b).

The modelled cumulative proportion of women who have attended screening at least once is shown in Figure 6. Additional details are provided in Appendix J Detailed model data sources.

Figure 6 Cumulative proportion of women ever screened by age.



Test characteristics — cytology

The model required a matrix of values that specify the relationship between each possible underlying natural history health state at the time of testing and each possible test result. We derived a series of test probability matrices for each of conventional, liquid-based, and automated cytology, based on the available evidence.

First, we derived a test probability matrix for conventional cytology, using published data to estimate how true underlying health states are distributed within each cytology result category, and data from registries to determine Australian reporting rates for each cytology test result. Our baseline test characteristics were chosen on the basis of best and most complete verification of results, use of Australian data where possible, and consistency of relative test performance with published data.

The accuracy of the conventional cytology implied by the selected test probability matrix is detailed in Table 28. The calibrated accuracy of manual and automated LBC relative to conventional cytology is shown in Table 29 (see also Table 49 and Table 50 in 'Calibration of the screening model', page 88). A range of values were derived for sensitivity analysis.

Table 28 Implied sensitivity and specificity of conventional cytology for CIN 2+, based on the baseline test probability matrix.

Test threshold	Sensitivity %	Specificity %
Possible LSIL	88.2	95.4
Definite LSIL	72.2	97.6

Table 29 Implied sensitivity and specificity of manual and automated LBC for CIN 2+, relative to conventional cytology (test threshold of possible pLSIL), based on the calibrated test probability matrices for manually read and automated LBC.

Cytology test technology	**Sensitivity relative to conventional**	Specificity relative to conventional
LBC (manual reading)	1.03	0.99
Automated LBC	1.11	0.99

Test probability matrices for manual and automated reading of LBC slides were derived from the baseline conventional cytology matrix. We adjusted the distribution of cytology test results to reflect the different distributions of cytology test results of manual and automated reading of LBC slides, while keeping constant the prevalence of disease implied by the matrix. Call rates for LBC and automated LBC relative to conventional cytology were based on data from Davey et al. 2007a (automated LBC), and from the review of data described in an earlier section (manual LBC; see 'Test yield', page 33). Where assumptions were required, we chose values which favoured the new technologies over conventional cytology. Estimates of additional true and false positive results detected by the new technologies were derived from the test probability matrices, and matrices were then calibrated to reproduce the reported rates of additional true and false positive results for the new technologies (see Table 17 and Table 23 under the earlier section 'Results of assessment', on pages 43 and 55). These rates are shown in Table 50 (page 88).

The derivation of the test probability matrices are described in detail in 'Appendix K Derivation of test probability matrices for cytology'.

Cytology unsatisfactory rates

We used the available data on cytology unsatisfactory rates in Australia to derive baseline values and ranges for sensitivity analysis, which are summarised in Table 30. Further details are provided in Appendix J Detailed model data sources'.

Table 30 Model values for rates of unsatisfactory cytology, by test technology.

Test type	Unsatisfactory rate used in model		Reference
	Baseline (%)	Range for sensitivity analysis (%) [a]	
Conventional cytology	2.2	0.5–5.0	Davey et al. 2007a; data from NSW Pap Test Registry & VCCR
Manual LBC	1.8	0.5–2.57	Davey et al. 2007a; pilot study in DHM [b]; Ronco et al. 2007; Bolger et al. 2006; CADTH (Krahn et al. 2008)
Automated LBC	1.8	0.5–2.0	Davey et al. 2007a; pilot study in DHM [b]; Bolger et al. 2006

[a] Within these ranges, the sensitivity analysis of the relative performance of the test technologies was constrained such that the unsatisfactory rate of conventional cytology was always ≥ the unsatisfactory rate of manual LBC or automated LBC.

[b] Personal communication, Dr. Annabelle Farnsworth, Douglass Hanly Moir Laboratories, Sydney.

Test characteristics — colposcopy

We derived a test probability matrix for colposcopy based on data on over 21 000 colposcopies supplied by the Royal Women's Hospital in Victoria (personal communication, Dr. Jeffrey Tan, Royal Women's Hospital, Melbourne). The test probability matrix specifies the relationship between each possible underlying natural history health state at the time of testing and the probability that a biopsy would be taken. These data were also used to derive an age-specific probability of unsatisfactory colposcopy. Details are provided in Appendix J Detailed model data sources.

Test characteristics—Hybrid Capture II (HCII)

The test characteristics of HCII were assessed in order to simulate current practice in Australia, where it is used as a test of cure following treatment for CIN 2 and 3. Values used in the model are presented in Table 31. Details are provided in Appendix J Detailed model data sources.

Table 31 Modelled test characteristics of Hybrid Capture II HPV test (based on international data).

Model health state	Gold standard used	Hybrid Capture II positivity rate	
		Test of cure	
		Baseline (%)	Range (%)
Normal	PCR –ve, normal cytology	1.4	1.4–4.2
HPV (no CIN)	PCR +ve, normal cytology	49.7	49.7–92.5
CIN 1	Histology (or cytology if no histology)	84.2	69.4–98.9
CIN 2	Histology	94.4	90.9–97.9
CIN 3+	Histology	94.4	90.9–97.9

Abbreviations: PCR = polymerase chain reaction; CIN = cervical intraepithelial neoplasia.

Treatment failure and post-treatment recurrence

The treatment failure rate within the first year and annual recurrence rates thereafter were obtained from our own review and meta-analysis of data from relevant studies (Baldauf et al. 1998; Bigrigg et al. 1994; Cecchini et al. 2002; Flannelly et al. 1997; Oyesanya et al. 1993; Paraskevaidis et al. 2000; Powell 1996; Wright et al. 1992). Details are provided in Appendix J Detailed model data sources.

Costs

Costs included in the modelled economic evaluation included those related to screening, diagnosis, treatment and follow-up of cervical abnormalities, and to the treatment of cervical cancer, based on current Australian clinical practice.

Unit costs for modelled resources were sourced from:

- the Applicant's requested MBS fees
- Medicare Benefits Schedule Online (August 2008) (outpatient medical services)
- National Hospital Cost Data Collection Round 11 (2006–07, public) (inpatient hospital services)
- Pharmaceutical Benefits Schedule (PBS) Online (August 2008) where applicable.

Data on the usage of resource items were based on PBS and MBS use statistics for 2007 (Medicare Australia, 2008). Literature sources and advisory panel expert opinion were used, as necessary, to supplement data on local clinical practice. The distribution of cancer stage (FIGO) at diagnosis was based on data provided by the Queensland Gynaecological Cancer Centre and the Royal Women's Hospital, Melbourne.

Table 32 summarises of the cost items included in the modelled evaluation. Subsequent sections describe in more detail the calculation of the summary cost items.

Table 32 Cost items used in the modelled economic evaluation.

Cost item	Calculated unit costs	Table
Cost of conventional Pap test	$58.50	Table 35
Cost of repeat conventional Pap test after unsatisfactory screening test	$65.96	Table 35
Cost of LBC (manual reading, $10.90 incremental cost)	$69.40	Table 35
Cost of repeat LBC after unsatisfactory screening test (manual reading, $10.90 incremental cost)	$76.86	Table 35
Cost of LBC (manual reading, $2.40 incremental cost)	$60.90	Table 35
Cost of repeat LBC after unsatisfactory screening test (manual reading, $2.40 incremental cost)	$68.36	Table 35
Cost of LBC automated (TPI)	$74.90	Table 35
Cost of repeat LBC automated after unsatisfactory screening test (TPI)	$82.36	Table 35
Cost of results/ referral consultation following abnormal cytology result	$38.11	Table 36
Cost of colposcopy, no biopsy	$134.90	Table 37
Cost of colposcopy, with biopsy	$284.78	Table 38
Cost of cytology performed at colposcopy (conventional)	$23.60	Table 39
Cost of cytology performed at colposcopy (LBC manual, $10.90 increm cost)	$32.84	Table 39
Cost of cytology performed at colposcopy (LBC manual, $2.40 increm cost)	$25.64	Table 39
Cost of cytology performed at colposcopy (LBC automated)	$37.50	Table 39
Cost of treating CIN 2/3	$1112.26	Table 40 & Table 41
Cost of follow-up for treated CIN 2/3 (Conventional cytology)	$378.04	Table 42
Cost of follow-up for treated CIN 2/3 (LBC manual, $10.90 incremental cost)	$397.72	Table 42
Cost of follow-up for treated CIN 2/3 (LBC manual, $2.40 incremental cost)	$372.63	Table 42
Cost of follow-up for treated CIN 2/3 (LBC automated)	$413.96	Table 42
Cancer work-up—proportion weighted		
Localised	$1815.28	Table 43 & Table 47
Locally advanced / regional	$1994.50	Table 43 & Table 47
Distant	$1965.24	Table 43 & Table 47
Cancer treatment—proportion weighted		
Localised	$9907.88	Table 43 & Table 45
Locally advanced / regional	$15434.18	Table 43 & Table 46
Distant	$8957.25	Table 43 & Table 46
Cost of terminal care (assumption)	$8957.25	Assumption

Cost of Pap test

The cost of screening tests includes the average costs of a medical consultation, the MBS fee for reading of cytology, and the patient episode initiation fee, applicable to all medical consultations where a Pap test is taken.

The weighted-average cost of a medical surgery consultation in 2008, based on use data from Medicare Australia in 2007, is $38.11 (Table 33).

Table 33 Average costs of medical consultation (Medicare Australia 2007).

Components	MBS item	Unit cost	Proportion of consultations	Weighted cost (schedule fee)
GP consultation—limited management	3	$15.00	0.015	$0.23
GP consultation, < 20 min	23	$32.80	0.735	$24.09
GP consultation, 20–40 min	36	$62.30	0.108	$6.75
GP consultation, > 40 min	44	$91.70	0.010	$0.92
Non-GP consultation, < 5 min	52	$11.00	0.000	$0.00
Non-GP consultation, 5–25 min	53	$21.00	0.032	$0.67
Non-GP consultation, > 45 min	54	$38.00	0.006	$0.23
Non-GP consultation, 25–45 min	57	$61.00	0.001	$0.08
Specialist consultation, initial	104	$77.25	0.040	$3.12
Specialist consultation, subsequent	105	$38.80	0.052	$2.01
Weighted-average cost per consultation			1.000	$38.11

An additional adjustment is made to this average cost per consultation to take account of the fact that a Pap test may be the secondary reason for a medical consultation. A weighting factor to account for a 'single service' has been calculated on the basis of AIHW general practice activity data for 2006–07 (Britt, 2008) (Table 34).

Table 34 Problems managed per GP encounter—single-service weighting (AIHW (2008)—General practice activity in Australia 2006–07 (Table 7.1)).

No. problems managed at visit	No. encounters	Proportion of encounters	Single service weighting	Weighted-average single service adjustment
1	59635	0.650	1	0.650
2	22073	0.240	0.5	0.120
3	7835	0.085	0.33	0.028
4	2262	0.025	0.25	0.006
Total	91805	1		0.804

The costs of a Pap test using conventional cytology, LBC with manual reading and LBC with automated reading are presented below. It is assumed that medical consultation costs for screening tests are adjusted using the single service adjustment factor above; if repeat cytology is required because of an unsatisfactory smear, it is assumed that this is the only reason for the consultation.

Table 35 Cost of screening tests (CC, LBC with manual reading and automated LBC).

Components	MBS item	Unit cost (100% schedule)	Weighting for single service	Cost (initial screen)	Cost (repeat cytology for unsat)
Cost of conventional Pap test					
Weighted-average cost of a consultation	3, 23, 36, 44, 52, 53, 54, 57, 104, 105	$38.11	0.804	$30.65	$38.11
Pap test	73053, 73057	$19.60	–	$19.60	$19.60
Initiation of patient episode	73922	$8.25	–	$8.25	$8.25
Total cost				$58.50	$65.96
Cost of LBC (manual reading—$10.90 incremental cost)					
Weighted-average cost of a consultation	3, 23, 36, 44, 52, 53, 54, 57, 104, 105	$38.11	0.804	$30.65	$38.11
LBC (manual—higher incremental cost) [a]	n/a	$30.50	–	$30.50	$30.50
Initiation of patient episode	73922	$8.25	–	$8.25	$8.25
Total cost				$69.40	$76.86
Cost of LBC (manual reading—$2.40 incremental cost)					
Weighted-average cost of a consultation	3, 23, 36, 44, 52, 53, 54, 57, 104, 105	$38.11	0.804	$30.65	$38.11
LBC (manual—lower incremental cost) [b]	n/a	$22.00	–	$22.00	$22.00
Initiation of patient episode	73922	$8.25	–	$8.25	$8.25
Total cost				$60.90	$68.36
Cost of LBC (automated reading)					
Weighted-average cost of a consultation	3, 23, 36, 44, 52, 53, 54, 57, 104, 105	$38.11	0.804	$30.65	$38.11
LBC automated [c]	n/a	$36.00	–	$36.00	$36.00
Initiation of patient episode	73922	$8.25	–	$8.25	$8.25
Total cost				$74.90	$82.36

[a] Applicant-requested MBS fee (ThinPrep sponsor, 2002).
[b] Applicant-requested MBS fee (SurePath sponsor, 2007).
[c] Applicant-requested MBS fee (ThinPrep Imager sponsor, 2007).

Cost of results / referral consultation following abnormal cytology

We assumed that any abnormal cytology test result would prompt a second medical consultation for the purpose of communicating the test results and proposed management to the patient. When immediate referral for colposcopy is indicated, this visit would also involve providing a referral to a specialist. We assumed that this is the only reason for the consultation, and therefore the single-service adjustment factor was not used.

Table 36 Cost of results / referral consultation.

Components	MBS item	Unit cost (100% schedule)
Cost of results / referral consultation		
Weighted-average cost of a consultation	3, 23, 36, 44, 52, 53, 54, 57, 104, 105	$38.11
Total cost		

Cost of colposcopy without biopsy

The cost of a colposcopy without biopsy is comprised of specialist consultation and a colposcopy procedure. It is assumed that all colposcopies are performed in specialist's rooms without anaesthesia.

Table 37 Cost of colposcopy, no biopsy.

Components	MBS item	Unit cost (100% schedule)	Weighting for single service	Cost (initial screen)
Specialist consultation	104	$77.25	1.000	$77.25
Colposcopy	35614	$57.65	1.000	$57.65
Total cost				$134.90

Cost of colposcopy with biopsy

The cost of a colposcopy with biopsy is composed of specialist consultation, a biopsy procedure and a colposcopy procedure, discounted because the two procedures are performed at the same time (Multiple services rule, MBS Schedule note T8.2 Multiple Operations Rule). It is assumed that all colposcopies are performed in specialist's rooms without anaesthesia.

Table 38 Cost of colposcopy, with biopsy.

Components	MBS item	Unit cost (100% schedule)	Proportion of patients	Cost
Specialist consultation	104	$77.25	1.000	$77.25
Colposcopy [a]	35614	$28.83	1.000	$28.83
Biopsy	35608	$57.75	1.000	$57.75
Histopathology [b]	72823	$97.95	0.667	$65.37
	72824	$142.30	0.333	$47.34
Initiation of patient episode	73926	$8.25	1.000	$8.25
Total cost				$284.78

[a] MBS Multiple Operation Rule, Note T8.2.

[b] 2007 MBS use data (470 818 services for 723823; 234681 services for 72824).

Cost of cytology performed at colposcopy

We assumed that at 84.75 per cent of colposcopies, a cytology sample will be taken (based on analysis of data provided by Dr Jeffrey Tan, Royal Women's Hospital, Melbourne). Thus, each colposcopy will have a weighted cost for this associated cytology, the exact value of which depends on the type of cytology being assessed.

Table 39 Cost of cytology done at colposcopy.

Components	MBS item	Unit cost (100% schedule)	Proportion of patients	Cost
Cost of conventional Pap test				
Pap test	73053, 73055, 73057	$19.60	0.8475	$$16.61
Initiation of patient episode	73922	$8.25	0.8475	$6.99
Total cost				$23.60
Cost of LBC (manual reading—higher incremental cost)				
LBC (Manual)	n/a	$30.50	0.8475	$25.85
Initiation of patient episode	73922	$8.25	0.8475	$6.99
Total cost				$32.84
Cost of LBC (manual reading—lower incremental cost)				
LBC (manual)	n/a	$22.00	0.8475	$18.65
Initiation of patient episode	73922	$8.25	0.8475	$6.99
Total cost				$25.64
Cost of LBC (automated reading)				
LBC automated	n/a	$36.00	0.8475	$30.51
Initiation of patient episode	73922	$8.25	0.8475	$6.99
Total cost				$37.50

Cost of treating CIN 2/3

The cost of treating CIN 2/3 is composed of the cost of ablation treatments (including laser therapy [no hospitalisation] and diathermy [hospitalisation]) and excision treatments (loop excision [no hospitalisation] and cone biopsy [hospitalisation]) weighted by the estimated proportion of patients receiving these services. We further assumed that 5 per cent of patients with CIN 2/3 will be treated with hysterectomy (based on analysis of data provided by Dr Jeffrey Tan, Royal Women's Hospital, Melbourne).

Table 40 Cost of treating CIN 2/3 - Summary.

Components	Proportion of patients	Cost	Weighted cost
Ablation treatment	0.15	$1233.50	$185.03
Excision treatment	0.80	$704.28	$563.42
Hysterectomy (non-cancer)	0.05	$7276.25	$363.81
Total cost			$1112.26

Table 41 Cost of treating CIN 2/3 - Details.

Components	MBS / DRG item	Unit cost (100% schedule)	Proportion of patients	Cost
Ablation therapy (laser + diathermy) (15% total CIN 2/3 treatments)			0.15	$1233.50
Laser therapy			*0.5*	*$355.75*
Specialist consult	104	$77.25	1	$77.25
Colposcopy [a]	35614	$28.83	1.000	$28.83
Laser therapy 1 site	35539	$246.50	0.925 [a]	$227.92
Laser therapy 2 sites	35542	$288.60	0.075 [a]	$21.75
Diathermy			*0.5*	*$2111.25*
Specialist consult	104	$77.25	1	$77.25
Diathermy, public hospital	DRG N09Z	$2034.00	1	$2034.00
Excision therapy (80% total CIN 2/3 treatments)			0.8	$704.28
LEEP no hospitalisation			*0.855* [a]	*$465.94*
Specialist consult	104	$77.25	1	$77.25
Cervix, large loop excision of transformation zone together with colposcopy	35647	$183.90	0.944 [a]	$173.66
Cervix, large loop excision diathermy, in conjunction with ablative treatment of additional areas of intraepithelial change	35648	$287.85	0.056 [a]	$16.02
Histopathology	72830 (level 5 complexity)	$190.75	1	$190.75
Initiation of patient episode for histopathology	73926	$8.25	1	$8.25
Cone biopsy			*0.145* [a]	*$2111.25*
Specialist consult	104	$77.25	1	$77.25
Hospitalisation, public	DRG N09Z	$2034.00	1	$2034.00
Hysterectomy (non-cancer) (5% total CIN 2/3 treatments)			0.05	$7276.25
Hysterectomy			*1.000*	*$7276.25*
Specialist consult	104	$77.25	1	$77.25
Hysterectomy for non-malignancy	DRG N04Z	$7199.00	1	$7199.00
Total weighted-average cost of treatment for CIN 2/3				$1112.26

[a] MBS use data 2007. DRG = diagnosis-related group. LEEP = loop electrosurgical excision procedure.

Cost of CIN 2/3 follow-up

The cost of CIN 2/3 follow-up is based on the management pathway in the 2005 NHMRC guidelines, and is made up of medical consultations at 6, 12 and 24 months, Pap tests at 6, 12, and 24 months, colposcopy at 6 months, and HPV-DNA tests at 12 and 24 months. Costs have been calculated separately for each screening method.

Table 42 Cost of CIN 2/3 follow-up (conventional cytology, LBC with manual reading and automated LBC) (based on 2005 NHMRC guidelines).

Components	MBS / DRG item	Prop. patients	Cost conventional	Cost LBC manual (higher)	Cost LBC manual (lower)	Cost LBC automated
4–6 months after treatment						
Specialist consult	105	1	$38.80	$38.80	$38.80	$38.80
Initiation of patient episode	73922	1	$8.25	$8.25	$8.25	$8.25
Pap test	73053, 73057 [a]	1	$19.60	$30.50	$22.00	$36.00
Colposcopy	35614	1	$57.65	$57.65	$57.65	$57.65
12 months after treatment						
Weighted-average cost of a consultation	3, 23, 36, 44, 52, 53, 54, 57, 104, 105	1	$38.11	$38.11	$38.11	$38.11
Initiation of patient episode	73922	1	$8.25	$8.25	$8.25	$8.25
Pap test	73053, 73057 [a]	1	$19.60	$30.50	$22.00	$36.00
HPV test	69418	1	$64.00	$57.60 [b]	$57.60 [b]	$57.60 [b]
24 months after treatment (discounted @ 5% pa)						
Weighted-average cost of a consultation	3, 23, 36, 44, 52, 53, 54, 57, 104, 105	1	$36.30	$36.30	$36.30	$36.30
Initiation of patient episode	73922	1	$7.86	$7.86	$7.86	$7.86
Pap test	73053, 73057 [a]	1	$18.67	$29.05	$20.95	$34.29
HPV test	69418	1	$60.95	$54.86 [b]	$54.86 [b]	$54.86 [b]
Total follow-up costs CIN 2/3			$378.04	$397.72	$372.63	$413.96

[a] MBS item number for CC only. DRG = diagnosis-related group.
[b] Assume a 10% lower cost for HPV test from LBC as sample collection will not need to be done separately.

Cost of cancer treatment and work-up by stage and disease extent

There are no direct data on the stage-specific treatment or work-up costs for cervical cancer. Therefore, Advisory Panel expert opinion has been used to provide an estimate of likely treatment practices by disease stage. In the absence of data on the distribution of treatment costs over time, all cancer treatment costs are applied as one-off costs in the year of diagnosis, and a cost of cancer-related death is applied in the year a patient dies from cervical cancer.

Distribution of cancer stage (FIGO) at diagnosis was based on data provided by the Queensland Gynaecological Cancer Centre and the Royal Women's Hospital, Melbourne. As the economic model is structured by disease extent (localised, locally advanced, regional, distant), FIGO stages were grouped to broadly represent these disease extent categories.

Table 43 Summary treatment and work-up costs by FIGO stage and disease extent.

Disease extent	% of all cervical cancer cases [a]	% cases in disease extent category	Stage-specific work-up costs	Proportion weighted work-up costs	Stage-specific treatment costs	Proportion weighted treatment costs
Localised	54.4%			$1815.28		$9907.88
1a	1.1%	2.0%	$1815.28	$37.16	$5489.25	$112.38
1a1	16.5%	30.4%		$551.73		$1668.39
1a2	4.5%	8.2%		$148.65	$10190.61	$834.51
1b	6.9%	12.8%		$231.55	$12283.30	$1566.85
1b1	25.3%	46.6%		$846.18		$5725.76
Locally advanced / regional	43.7%			$1994.50		$15434.18
1b2	9.4%	21.6%	$2261.83	$487.85	$15348.80	$3310.53
2a	4.2%	9.6%	$1917.26	$184.21	$15218.35	$1462.16
2b	14.7%	33.7%		$646.61	$15479.25	$5220.45
3a	2.9%	6.7%		$127.82		$1031.95
3b	9.4%	21.6%		$413.53		$3338.66
3c	0.3%	0.8%		$15.04		$121.41
4a	2.7%	6.1%	$1965.24	$119.46	$15613.10	$949.03
Distant	2.0%			$1965.24		$8957.25
4b	2.0%	100%	$1965.24	$1965.24	$8957.25	$8957.25

[a] Based on data from Queensland Gynaecological Cancer Centre and the Royal Women's Hospital, Melbourne.

Table 44 Summary treatment and work-up costs by FIGO stage.

Treatments [a]	Unit cost	Source										
							% of patients					
Treatment option			Ia1	Ia2	Ib1	Ib2	IIa	IIb	IIIa	IIIb	IVa	IVb
Surgery alone [b]												
conisation	$2 111.25	Table 45	25.0%									
hysterectomy	$6 615.25	Table 45	75.0%	50.0%								
radical hysterectomy	$12 870.25	Table 45		50.0%	85.0%	5.0%	10.0%					
exenteration	$18 156.25	Table 45									5.0%	
Radiotherapy alone	$8 957.25	Table 46			15.0%							100.0%
Adjuvant radiotherapy	$8 957.25	Table 46		5.0%								
Chemo-radiation	$15 479.25	Table 46				95.0%	90.0%	100.0%	100.0%	100.0%	95.0%	
Stage-specific treatment costs			$5 489.25	$10 190.61	$12 283.30	$15 348.80	$15 218.35	$15 479.25	$15 479.25	$15 479.25	$15 613.10	$8 957.25

Work-up [c]			Ia1	Ia2	Ib1	Ib2	IIa	IIb	IIIa	IIIb	IVa	IVb
Colposcopy	$335.08	Table 47	100.0%	100.0%	100.0%	100.0%	90.0%	90.0%	90.0%	90.0%	90.0%	90.0%
Chest x-ray	$47.15	MBS 58503	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
CT scan	$480.05	MBS 56507	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
PET scan	$953.00	MBS 61529	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
MRI	$446.55	MBS 63470 / 63473	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Bone scan	$479.80	MBS 61421	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	10.0%	10.0%
Cystoscopy	$150.55	MBS 36812	0.0%	0.0%	0.0%	0.0%	90.0%	90.0%	90.0%	90.0%	90.0%	90.0%
Stage-specific work-up costs			$1 815.28	$1 815.28	$1 815.28	$2 261.83	$1 917.26	$1 917.26	$1 917.26	$1 917.26	$1 965.24	$1 965.24

[a] Treatment proportions from Prof. Ian Hammond (via L. Farrell).
[b] Stage Ia1 surgery = 25% conisation (DRG N09Z) 75% hysterectomy (DRG N03A/B) (Neville Hacker, personal communication).
[c] Work-up proportions from N. Hacker, personal communication.

Cost of surgery

Surgery options include conisation, simple hysterectomy, radical hysterectomy and exenteration. As current diagnosis-related groups (DRGs) do not distinguish between simple and radical hysterectomy, we have assumed that a simple hysterectomy has a cost equivalent to a hysterectomy without complications, and a radical hysterectomy has a cost equivalent to a hysterectomy with complications.

Table 45 Cost of surgical managements.

Components	DRG item	Unit cost	Proportion of patients	Cost
Cost of conisation				
Specialist consult	104	$77.25	1	$77.25
Conisation procedure, public	DRG N09Z	$2034.00	1	$2034.00
Total				$2111.25
Cost of simple hysterectomy (assume as for hysterectomy –CC)				
Specialist consult	104	$77.25	1	$77.25
Uterine, adnexa procedure for non-ovarian/adnexal malignancy –CC	DRG N03B	$6538.00	1.00	$6538.00
Total				$6615.25
Cost of radical hysterectomy including nodes (assume as for hysterectomy +CC)				
Specialist consult	104	$77.25	1	$77.25
Uterine, adnexa procedure for non-ovarian/adnexal malignancy +CC	DRG N03A	$12793.00	1.00	$12793.00
Total				$12870.25
Cost of exenteration				
Specialist consult	104	$77.25	1	$77.25
Pelvic evisceration & radical vulvectomy	DRG N01Z	$18079.00	1.00	$18079.00
Total				$18156.25

DRG = diagnosis-related group.

Cost of non-surgical management

Non-surgical management options include primary radiotherapy, adjuvant radiotherapy and chemo-radiotherapy. We have assumed that adjuvant radiotherapy has the same cost as primary radiotherapy.

Table 46 Cost of non-surgical managements.

Components	DRG item	Unit cost	Units	Cost
Cost of radiation				
Specialist consult	104	$77.25	1	$77.25
Inpatient radiation	DRG R64Z	$740.00	2	$1480.00
Outpatient radiation [a]	DRG R64Z	$370.00	20	$7400.00
Total				$8957.25
Cost of chemo-radiation				
Specialist consult	104	$77.25	1	$77.25
Inpatient radiation	DRG R64Z	$740.00	2	$1480.00
Outpatient radiation [a]	DRG R64Z	$370.00	20	$7400.00
Outpatient chemotherapy [b]	DRG R63Z	$1087.00	6	$6522.00
Total				$15479.25

[a] Assumed to be 50% cost of inpatient radiation.

[b] Assumed to be same as cost of inpatient chemotherapy (length of stay = 1 day).

DRG = diagnosis-related group.

Cost of work-up

Table 47 Unit costs of work-up procedures.

Components	MBS/DRG item	Unit cost	Proportion of patients	Cost
Colposcopy + biopsy				
Colposcopy [a]	35614	$28.83	1.000	$28.83
Biopsy	35608	$57.75	1.000	$57.75
Histopathology [b]	72823	$97.95	0.667	$97.95
	72824	$142.30	0.333	$142.30
Initiation of patient episode	73926	$8.25	1.000	$8.25
Total cost colposcopy				$335.08
Chest x-ray	58503	$47.15	1	$47.15
CT scan	56507	$480.05	1	$480.05
PET scan (assume as for non-small-cell lung cancer)	61529	$953.00	1	$953.00
MRI**	63470	$403.20	0.806	$446.55
	63473	$627.20	0.194	
Bone scan	61421	$479.80	1	$479.80
Cystoscopy	36812	$150.55	1	$150.55

[a] MBS multiple operation rule Note T8.2.
[b] 2007 MBS use data.
DRG = diagnosis-related group.

Management

Management of cytology, colposcopy and histology results was modelled to reflect current Australian recommendations (NHMRC 2005). In some instances the guidelines provide for the discretion of the clinician, and thus do not specify management at the level of detail required by the model. In these areas we assumed pathways of management practice based on input from the Advisory Panel—details of these assumptions are provided in Appendix J Detailed model data sources, in the section 'Management assumptions made for modelling in cases where guidelines do not specify outcomes', on page 148.

Baseline model calibration and validation

Calibration of the natural history model

As previously described (Canfell et al. 2008), the natural history model was calibrated to two sources of data:

1) The predicted age-specific prevalence of oncogenic HPV infection in Australia was calibrated to a number of published data sources from the international literature. Figure 7 (adapted from Smith et al. 2008) compares the model results with those obtained from an international meta-analysis of the cross-sectional prevalence of HPV (Burchell et al. 2006). The predicted and actual values showed close agreement in women aged less than 50 years. The modelled prevalence is also consistent with Australian data on HPV prevalence in a population of 805 non-Indigenous, cytologically normal women attending for routine screening who were recruited to the Women, Human Papillomavirus Prevalence, Indigenous, Non-Indigenous, Urban, Rural Study (WHINURS) in Australia (pers. comm., Prof. Suzanne Garland).

2) The age-specific incidence of cervical cancer in 25 developing countries without significant levels of cervical screening was calculated from data published in Parkin et al. (2002) (Figure 8). Modelled and actual values showed close agreement in women aged less than 50 years. Over the age of 50 years, the model predicts a plateau in age-specific cancer incidence; the decrease observed in the data from developing countries is likely to reflect cohort effects in women over 65 years of age (owing to a lesser risk of exposure to HPV infection as younger women). The predicted cumulative lifetime risk of cervical cancer was 2.6 per cent in women aged 10 to 85 years.

Figure 7 Predicted age-specific prevalence of oncogenic HPV infection, compared to a 2006 meta-analysis of international data (with 95% CI) (data from Smith et al. 2008).

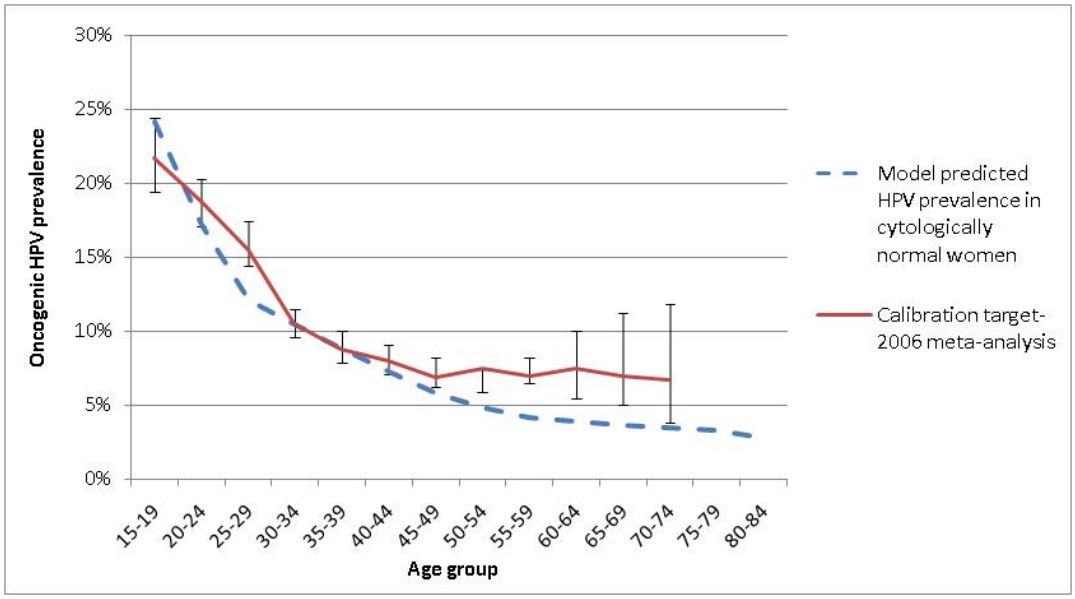
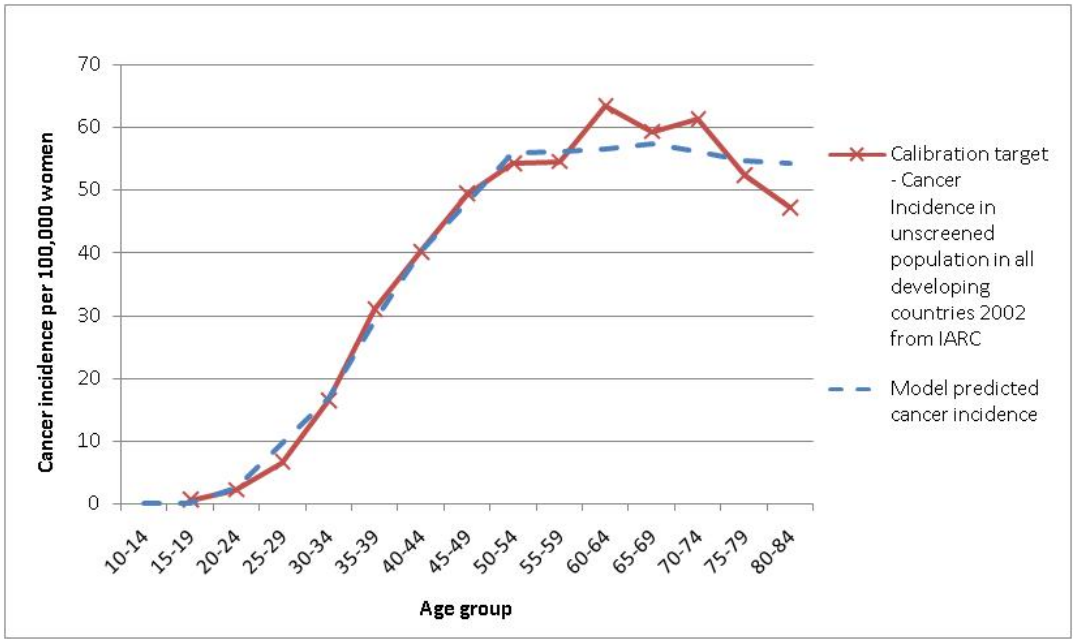


Figure 8 Predicted age-specific incidence of cervical cancer in an unscreened population, compared to data from 25 developing countries (data from Parkin et al. 2002).



Calibration of the screening model

Following calibration of the natural history model, the complete model of screening, diagnosis and management in Australia was implemented. The full model was considerably complex and incorporated data on age-specific screening initiation and screening and management recommendation compliance in Australian women (as informed by an analysis of VCCR data), and incorporated the estimates of test characteristics of conventional cytology and colposcopy.

The output of the full screening model was compared with:

(i) the age-specific and age-standardised incidence of cervical cancer in Australia over the period 2002–2004 (AIHW 2008)

(ii) the age-specific and age-standardised mortality due to cervical cancer in Australia over the period 2002–2004 (AIHW 2008)

(iii) the age-specific and age-standardised rate of histologically confirmed high-grade disease during 2006

(iv) the age-specific and age-standardised rate of histologically confirmed low-grade disease during 2006.

The results are shown in Figure 9, Figure 10, Figure 11, Figure 12, and Table 48.

Modelled and actual values of age-specific cancer incidence and mortality showed close agreement. The model predicts lower than observed rates of cervical cancer mortality among women over 70 years of age, potentially reflecting a cohort effect in the survival probabilities used to parameterise the model.

Figure 9 Predicted age-specific incidence of cervical cancer in Australia, compared with cancer registry data from 2002–2004.

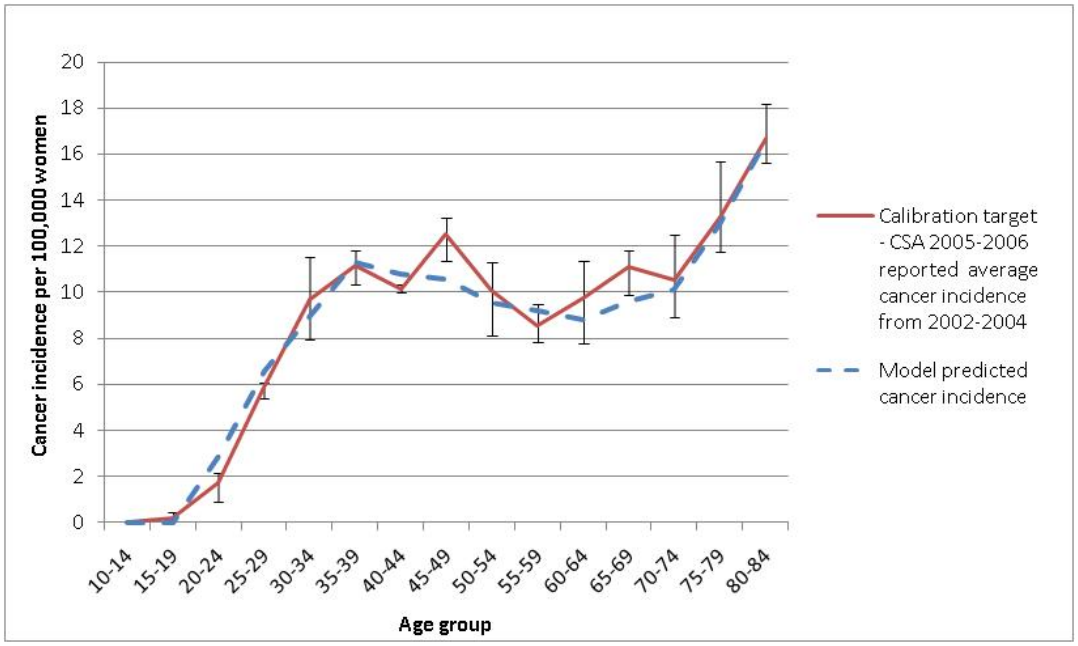


Figure 10 Predicted age-specific mortality in Australia, compared with cancer registry data from 2003 to 2005.

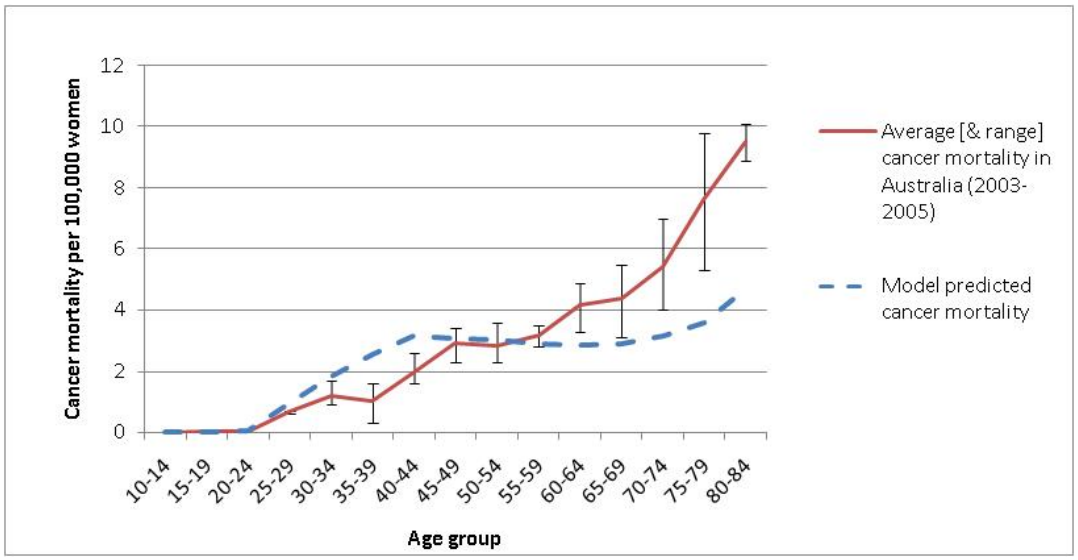


Figure 11 Predicted age-specific rate of histologically confirmed high grades detected, compared with registry data (2006).

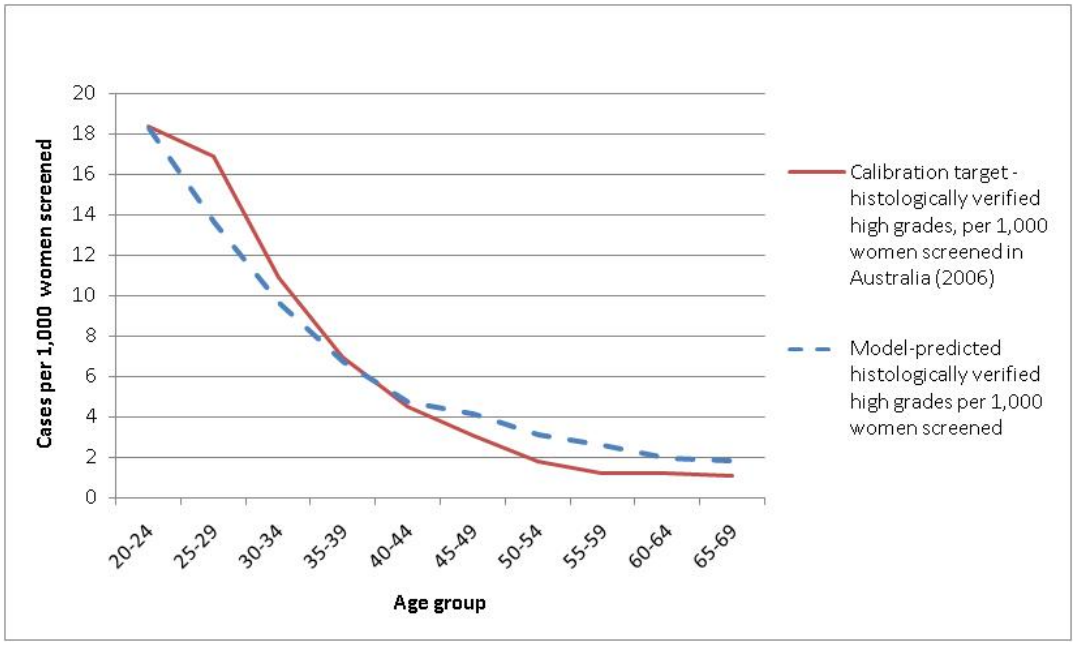
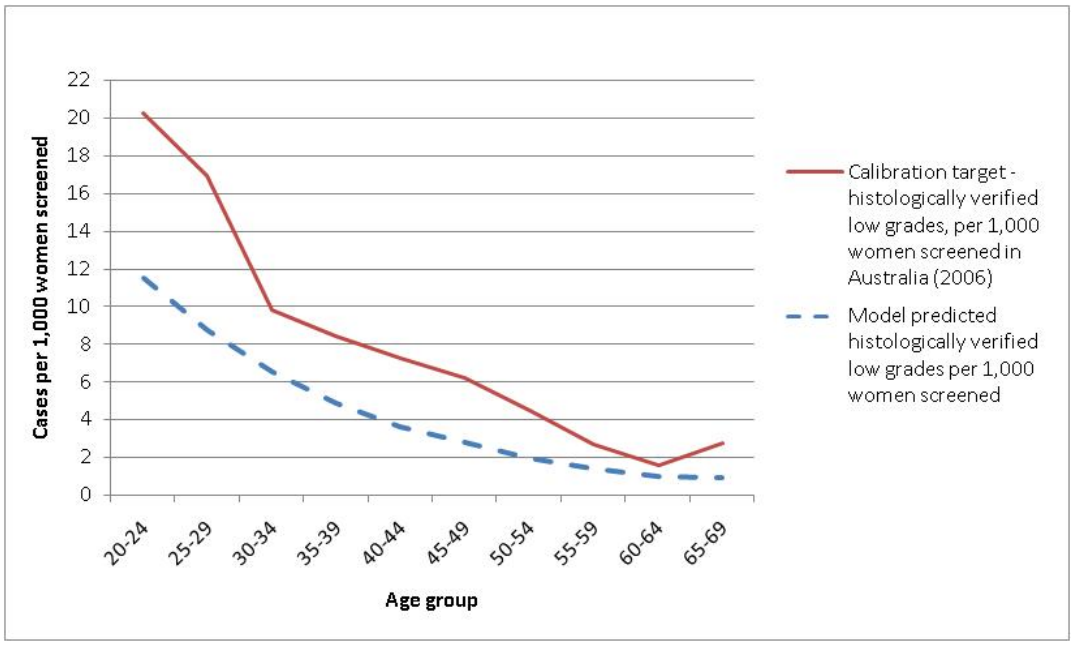


Figure 12 Predicted age-specific rate of histologically confirmed low grades detected, compared with registry data (2006).



Modelled and actual values of histologically confirmed high-grade lesions showed close agreement (Figure 11). The model predicts lower than observed rates of histologically confirmed low-grade lesions (Figure 12). This discrepancy is due largely to the model's simulating the current NHMRC management guidelines, which came into effect from July 2006, whereas the calibration data mostly reflect the management practice under the previous NHMRC management guidelines (NHMRC 1994). We confirmed this by performing a simplified test to determine whether the difference in management was sufficient to explain the difference between actual and predicted rates of histologically verified low-grade disease. The test involved simulating the referral of increasing proportions of women with cytological dLSIL (but not pLSIL) for immediate colposcopy. As the number of immediate referrals increased, the age-specific rates more closely resembled the calibration target.

Table 48 Model prediction versus actual data for detection of histological high disease, and cervical cancer incidence and mortality.*

	Model predicted	Australian data †
Detection of histologically confirmed high-grade disease per 1000 women screened, age 20–69 years	7.3	7.4 (range 7.3–7.5) average 2004–2006
Cumulative lifetime risk of cervical cancer	0.64% (to age 85)	0.66%
Invasive cancer incidence per 100 000 women, all ages	6.6.	6.9 (range 6.8–7.0) average 2002–2004
Invasive cancer incidence per 100 000 women, age 20–69 years	10.4	9.0 (range 8.9–9.1) average 2002–2004
Invasive cancer mortality per 100 000 women, all ages	1.7	2.0 (range 1.9–2.2) average 2003–2005
Invasive cancer mortality per 100 000 women, age 20–69 years	2.3	2.0 (range 1.8–2.2) average 2003–2005

* All values except detected histologically confirmed high-grade disease and cumulative lifetime risk are age-standardised to the Australian 2001 population. † Source (AIHW 2008).

Additional cytology calibration

As described above in 'Test characteristics — cytology' and in detail in Appendix K Derivation of test probability matrices for cytology, the test characteristics for conventional cytology were derived from several sources. They were also calibrated in conjunction with the test characteristics for HCII in the context of HPV triage in order to predict distributions of triage test results by age as reported in the literature (refer to *MSAC reference 39: HPV triage*). For this aspect of the calibration, the targets used were sensitivity and specificity of HPV triage for detection of CIN 2+, and HPV test positivity rates, in women with pLSIL and dLSIL cytology results. For consistency, we used the same calibrated test characteristics for conventional cytology and for HCII as reported in MSAC reference 39. The relevant calibration results are given in Appendix K Derivation of test probability matrices for cytology.

Test characteristics for the new cytology test technologies were adjusted so that their accuracies relative to conventional cytology were within the ranges reported in the literature. The most reliable data for manually read LBC came from a meta-analysis (Arbyn et al. 2008) which reported relative sensitivity and specificity compared to conventional cytology as ratios. There was no equivalent information on the ratios of sensitivity and specificity for automated LBC and conventional cytology. The relative sensitivity measured in terms of additional true positives and false positives per 1000 women screened was used as the calibration target for automated LBC. The sources of these estimates are described in the literature review section of this report ('test accuracy') and summarised in Table 17 (page 43).

It was not possible to precisely replicate the values in Table 17 in the context of other constraints on the test probability matrices. We therefore selected a set of values which were more favourable to the newer technologies. The results are shown in Table 49.

Table 49 Predicted and reported sensitivity and specificity of manual LBC, relative to conventional cytology, for CIN 2+.

	Sensitivity relative to conventional		Specificity relative to conventional	
Test threshold	Predicted by model	Reported [Arbyn et al.] (95% CI)	Predicted by model	Reported [Arbyn et al.] (95% CI)
pLSIL	1.03	1.03 (0.97–1.09)	0.990	0.91 (0.84–0.98)
dLSIL	1.03	1.03 (0.96–1.11)	0.991	0.97 (0.94–1.01)

The relative sensitivity of manual LBC compared to conventional cytology used in the model is consistent with that found by Arbyn et al. The relative specificity used in the model is greater than the point estimate in Arbyn et al., and is at the upper end of the 95% CI. As increased specificity will tend to improve cost-effectiveness, this is a favourable assumption.

Table 50 Predicted and reported additional TPs and FPs from automated LBC (test threshold of possible HSIL, excluding glandular where possible).

Comparator cytology test technology	Additional cases per 1000 women screened			
	True positives		False positives	
	Predicted by model	Reported [a]	Predicted by model	Reported [a]
Conventional cytology	1.57	0.82 [b] to 1.58	–0.07	–0.07 to 1.93

[a] Based on Table 17. [b] Significant at the 0.05 level.

The model-predicted additional true positives due to automated LBC are at the upper end of the range in published Australian data, and exceed the statistically significant estimate of 0.82 (Table 50). This is a favourable assumption, indicating an increase in sensitivity, which is likely to improve effectiveness.

The model-predicted additional false positives due to automated LBC are at the lower end of the reported range, indicating that the model assumes an improved specificity for CIN 2+ at the test threshold of pHSIL. This is also a favourable assumption, as a reduction in false positives will tend to increase cost-effectiveness.

Calibration of the total number of smears

The model was used to calculate the predicted total number of cytological smears performed annually in Australia, and the results were compared to the total number of 2.1 million smears observed in 2007 (Royal College of Pathologists Australasia Cytopathology Quality Assurance Programs 2008). To reproduce the total observed number of smears nationally, additional data were used to inform the model on multiple early rescreening tests in a small proportion of women, practices not following the guidelines, or management practices which are outside the scope of the guidelines (such as some aspects of colposcopy, biopsy, and treatment management practices). Using data from Victoria (personal communication, Dr Jeffrey Tan, Royal Women's Hospital,

Melbourne, Victoria; and data from the Victorian Cervical Cytology Register 1995–2007), the following assumptions were made:

- 85 per cent of women have cytology at time of colposcopy (Victorian colposcopy data).
- 19 per cent of women have cytology at time of treatment (Victorian colposcopy data).
- 20 per cent of women have an additional cytology test associated with post-treatment follow-up (assumption).
- 3 per cent of women given a recommendation to return for their next smear in 2 years (ie, routine screening) have 2 additional smears within 15 months of their index smear (guided by Victorian Cervical Cytology Registry 2006).
- 25 per cent of women given a recommendation to return for their next smear in 12 months have 2 additional smears within 15 months of their index smear (guided by VCCR statistical report 2006).

Results of the economic evaluation

The model was used to calculate predicted discounted lifetime costs and effects if the new test technologies were introduced in Australia. The cost-effectiveness ratio (calculated as a dollar amount per LYS) was then calculated for each new technology compared with current practice. The ICERs for the new technologies versus the next most cost-effective alternative were also calculated using standard methods. A discount rate of 5 per cent was used for costs and effects.

The cost-effectiveness ratios for each technology relative to current practice are given in Table 51. For LBC with manual reading, we performed two separate evaluations of the cost-effectiveness at differing levels of reimbursement for the test: $22.00 ($2.40 incremental cost) and $30.50 ($10.90 incremental cost).

The main findings of the economic evaluation were that cost-effectiveness ratios are high for both test technologies at the currently requested level of reimbursement. As described under 'Additional cytology calibration', these findings were made in the context of favourable assumptions with respect to test accuracy (see Table 49 and Table 50).

Table 51 Baseline results of economic evaluation.

Scenario	Cost-effectiveness ratio relative to current practice ($ / LYS)
Automated LBC slide reading	$194 835
Manual LBC slide reading—$2.40 incremental cost	$126 315
Manual LBC slide reading—$10.90 incremental cost	$385 982

More details are presented below on the predicted costs and effects (Table 52), process outcomes (Table 53), and health outcomes (Table 54).

Both manual and automated LBC are predicted to result in a benefit in terms of the average discounted years of life saved in Australian women: between approximately 1 h (manual LBC) and 3 h (automated LBC). The benefit is achieved at an increased cost (see Table 52 and Table 57), attributable mainly to the increased cost of cytology, although there is also a predicted increase attributable to the additional colposcopies and biopsies associated with the new technologies.

Both manual and automated LBC are predicted to reduce the number of unsatisfactory smears, but in both cases this reduction is countered by an increase in the total number of satisfactory smears. For example, the analysis assumes that manual LBC is more sensitive but less specific at a threshold of pLSIL than conventional cytology (see Table 29). Both of these properties will tend to increase the number of smears with manual LBC, and therefore will also tend to increase the number of colposcopies, biopsies and treatments.

Both manual and automated LBC are predicted to result in a 17 per cent reduction in the total number of unsatisfactory smears conducted within the program per year (a reduction of approximately 7000 smears), but the total number of satisfactory smears would increase by 1.6 to 2.4 per cent (29 960–45 263) because of differences in test characteristics and the associated flow-on effects. In net terms, therefore, this would result in 22 763 to 38 346 more smears performed annually if liquid-based technology were introduced.

The use of manual LBC is predicted to result in an additional 6770 colposcopies and 3273 biopsies. This would be accompanied by an estimated 23 fewer cancer cases and 6 fewer cancer deaths annually.

The use of automated LBC is predicted to result in an additional 10 788 colposcopies and 5154 biopsies. This would be accompanied by 68 fewer cancer cases and 19 fewer cancer deaths annually.

Table 52 Predicted costs, effects, and incremental cost-effectiveness ratios, by cytology test technology.

Strategy	Discounted lifetime costs (5% discount rate)	Discounted life years (5% discount rate)	Incremental discounted life years (min) compared with current practice	ICER vs current practice ($ / LYS)	ICER vs next most cost-effective strategy ($ / LYS)
Current practice (conventional cytology)	$297.55	29.69575564	–	–	–
Manual LBC ($2.40 incremental cost)	$311.52	29.69586622	0.000111 (58)	$126 315	$126 315
Automated LBC	$366.31	29.69610857	0.000353 (186)	$194 835	$226 100
Manual LBC ($10.90 incremental cost)	$340.23	29.69586622	0.000111 (58)	$385 982	Dominated [a]

[a] Strategy is said to be dominated as it is more expensive than a strategy with equal or greater effectiveness, in this case Manual LBC at the lower incremental cost

Table 53 Predicted process outcomes in Australia, by cytology test technology (per annum).

Strategy	No. satisfactory smears	No. unsatisfactory smears	No. colposcopies	No. biopsies	CIN 2/3 treatment
Current practice (CC)	1 860 071	41 842	64 684	31 356	17 270
Manual LBC ($2.40 incremental cost)	1 890 031	34 644	71 454	34 630	18 004
Automated LBC	1 905 334	34 925	75 472	36 510	19 021
Manual LBC ($10.90 incremental cost)	1 890 031	34 644	71 454	34 630	18 004

Table 54 Predicted health outcomes in Australia, by cytology test technology (per annum).

Strategy	Cancer cases	Cancer incidence (ASR [a] per 100 000 women, all ages)	Cumulative lifetime risk of cancer (to age 85 years)	Cancer deaths	Cancer mortality (ASR [a] per 100 000 women, all ages)	Histologically confirmed high grades (all ages)
Current practice (CC)						
Predicted	712	6.58	0.64%	188	1.72	12 608
Actual [b]	*718*	*6.9*	*0.66%*	*216*	*2.0*	*14 469* [b]
Manual LBC ($2.40 incremental cost)	689	6.37	0.62%	182	1.67	12 921
Automated LBC	644	5.95	0.58%	169	1.55	13 694
Manual LBC ($10.90 incremental cost)	689	6.37	0.62%	182	1.67	12 921

[a] Age-standardised rate, standardised to the Australian 2001 population. [b] Refer to Table 48 and (AIHW 2008).

Overall, liquid-based technologies could be expected to increase the number of colposcopies by 6770–10 788 annually (10%–17%), the number of biopsies by 3273–5154 (10%–16%), and the number of treatments for CIN 2/3 by 735–1751 (4%–10%). This is in part because both manual and automated LBC are assumed to be more sensitive than CC, and thus more true high-grade lesions will be detected (314–1086 additional cases). This will in turn result in a reduction in incident cancers (23–68 cases), and cancer deaths (6–19 deaths). Because survival from cervical cancer is relatively high, the predicted impact on cancer deaths is less than that on cancer incidence.

The impact of the new technologies on average LYS in the population is 0.0001–0.0004 LYS, or approximately 1–3 h over a lifetime, when discounted at 5 per cent. Whether this outcome is considered to be an improvement in overall health outcomes depends on the trade-off between the benefit from a reduction in cancer incidence and the harms associated with increased investigations among the women who are not destined to develop cancer (Table 55). This trade-off could be captured in an analysis incorporating quality-adjusted survival (QALYs). However, such an analysis has not been conducted, because of paucity of data to inform QALY weights, which are required for the sequence of health states and events modelled over the duration of the model. This issue is discussed further in the Discussion under 'Quality-of-life issues', on page 103.

Table 55 summarises how process outcomes are related to averting cancers and deaths.

Table 55 Additional interventions required to prevent one incident cancer case and one cancer death.

Preventing one additional:	Requires an additional:			
	Smears	**Colposcopies**	**Biopsies**	**CIN2/3 treatments**
—cancer case				
Automated LBC	566	159	76	26
Manual LBC	990	295	142	32
—cancer death				
Automated LBC	1982	558	266	91
Manual LBC	3534	1051	508	114

It should be emphasised that these modelled outcomes are based on the most favourable assumptions for the new test characteristics, and in particular a higher than expected sensitivity for detecting high-grade lesions by automated LBC. Thus, for automated LBC, the impact on cancer incidence and deaths may be overestimated.

The modelled analysis also predicts that if automated LBC were to be assessed in the context of the screening program already using manually read LBC at the lower reimbursement level ($2.40 incremental cost), then the further addition of automated reading would be associated with an additional $226 100 per LYS (Table 52). This decrease in the cost-effectiveness of automated LBC when considered in relation to manually read LBC rather than in relation to current practice results from the fact that some of the benefits associated with automated LBC (such as a reduction in unsatisfactory smears) would already have been realised by the adoption of manual LBC. In the analysis, automated LBC reading was assumed to be more accurate than manual LBC (both more sensitive and more specific), but in the cost-effectiveness assessment the improved test accuracy was calculated in relation to the increase in cost associated with automated LBC ($36, compared with $22 for the lowest cost associated with manual

LBC). Manual LBC at $10.90 incremental cost is less cost-effective than automated LBC using ThinPrep Imager, and it is always less cost-effective than manual LBC at a $2.40 incremental cost, as it is assumed to have the same test accuracy but a higher cost.

Although we found cost-effectiveness ratios in terms of cost per LYS due to prevention of cervical cancer to be unfavourably high for both manual and automated reading of LBC, liquid-based sampling has other benefits which were not taken into account in the modelled analysis. These include the capacity to perform reflex testing for HPV, *Chlamydia trachomatis* or *Neisseria gonorrhoeae* from the same sample (see 'Other relevant considerations', page 99).

Sensitivity analysis

Sensitivity analyses were performed to determine how robust the conclusions of the analysis were to various model assumptions. In this analysis, the ICERs were most sensitive to assumptions concerning the cost of the new technology, the discount rate, the test characteristics, the recommended screening interval, and the likelihood that CIN 3 lesions will progress to cancer. The analysis found that assuming a more rapid progression from CIN 3 to cancer will tend to improve cost-effectiveness of more sensitive technologies. For similar reasons, a model which did not assume that some CIN 3 lesions regressed (which is not the case in the current analysis) would favour a more sensitive test. However, such an assumption would not be supported by the evidence.

We also investigated the effect of varying the model assumptions of post-treatment natural history and the use of HPV as a test of cure. These were found to have a minimal impact on the calculated cost-effectiveness ratios of the new technologies (<2%).

Table 56 summarises the range over which model parameters were varied for sensitivity analysis. For more information on the sources used for baseline values and ranges, see 'Summary of data sources for the baseline evaluation model', page 67. The results of the sensitivity analysis are summarised in Figure 13 and Figure 14.

In this analysis, the ICERs were most sensitive to assumptions concerning the cost of the new technology, the discount rate, the test characteristics, the recommended screening interval, and the likelihood that CIN 3 lesions will progress to cancer. The analysis found that assuming a more rapid progression from CIN 3 to cancer will tend to improve cost-effectiveness of more sensitive technologies. For similar reasons, a model which did not assume that some CIN 3 lesions regressed (which is not the case in the current analysis) would favour a more sensitive test. However, such an assumption would not be supported by the evidence.

We also investigated the effect of varying the model assumptions of post-treatment natural history and the use of HPV as a test of cure. These were found to have a minimal impact on the calculated cost-effectiveness ratios of the new technologies (<2%).

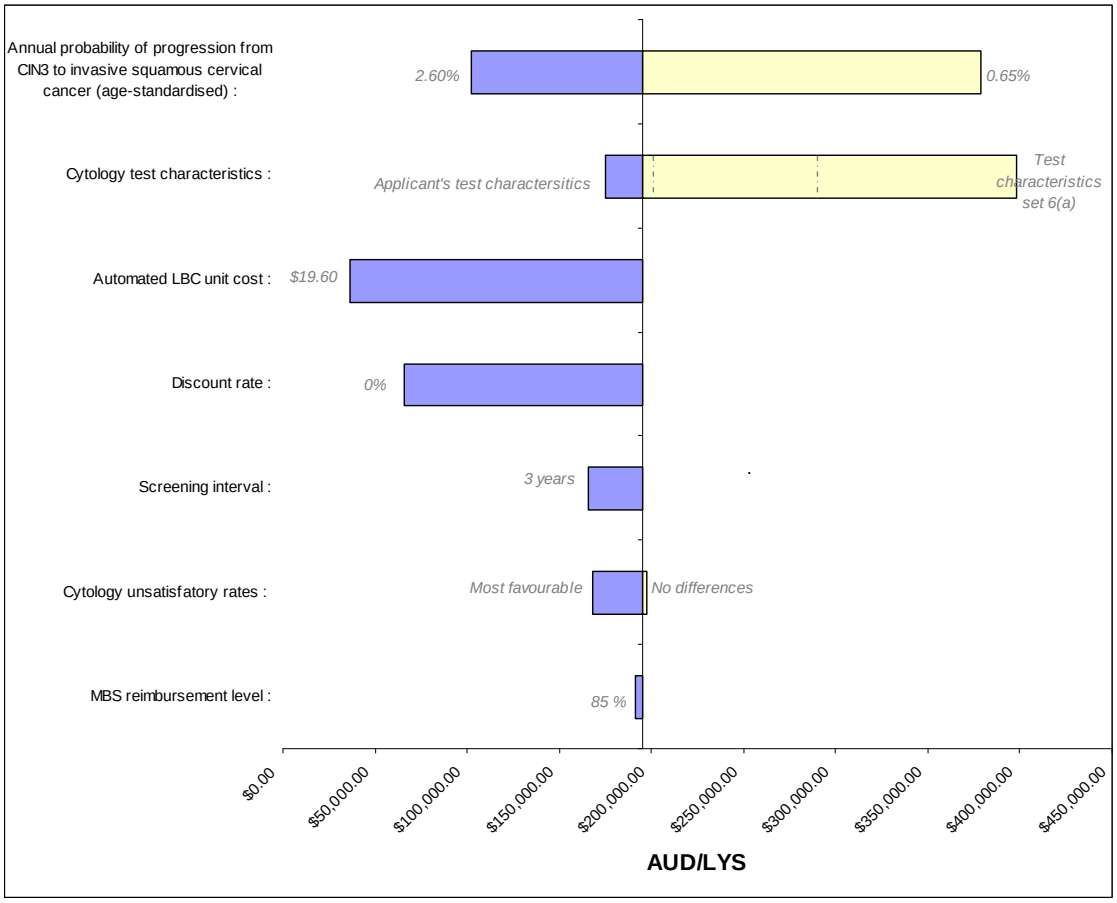
Table 56 Summary of range of parameter values used for sensitivity analysis.

Model parameter		Baseline value	Range used in sensitivity analysis
Cytology unsatisfactory rates (see Table 30)			
Conventional		2.2%	0.5%–5% [a]
LBC manual		1.8%	0.5%–2.57% [a]
Automated LBC		1.8%	0.5%–2% [a]
Cytology test characteristics		As described in *Appendix K Derivation of test probability matrices for cytology* (p 153)	
HC-II positivity by model health state (when used as a test of cure) (see Table 31)			
Normal	(PCR –ve, normal cytology)	1.4%	1.4%–4.2%
HPV (no CIN)	(PCR +ve, normal cytology)	49.7%	49.7%–92.5%
CIN 1	(confirmed by histology or cytology if no histology)	84.2%	69.4%–98.9%
CIN 2	(confirmed by histology)	94.4%	90.9%–97.9%
CIN 3+	(confirmed by histology)	94.4%	90.9%–97.9%
Annual probability of progression from CIN 3 to invasive squamous cervical cancer (age-standardised)		1.3% (all ages)	0.65%–2.6% (all ages)
Manual LBC unit cost		$22.00	$19.60–$30.50
(incremental cost compared with CC)		($2.40)	($0–$10.90)
Automated LBC unit cost		$36.00	$19.60–$36.00
(incremental cost compared with CC)		($16.40)	($0–$16.40)
Discount rate		5%	0%–5%
Recommended screening interval		Two-yearly recommendation—compliance informed by analysis of registry data	Three-yearly recommendation—assumed compliance is informed by data from settings where 3-yearly screening is recommended
MBS reimbursement level		100%	85%

Abbreviations: LBC = liquid-based cytology; HC-II = Hybrid Capture II HPV test; PCR = polymerase chain reaction; CIN = cervical intraepithelial neoplasia; CC = conventional cytology.

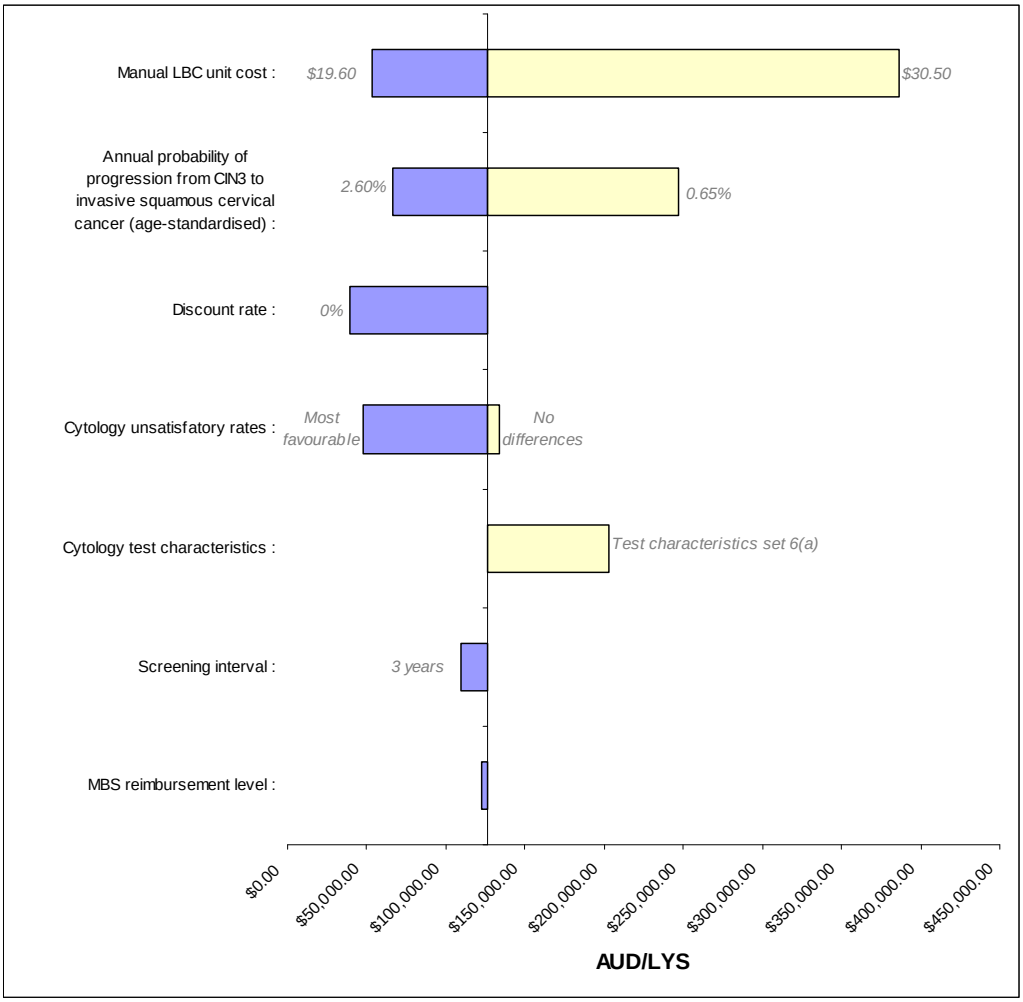
[a] Within these ranges, the sensitivity analysis of the relative performance of the test technologies was constrained such that the unsatisfactory rate of conventional cytology was always ≥ the unsatisfactory rate of manual LBC or automated LBC.

Figure 13 Sensitivity analysis for cost-effectiveness ratios of automated LBC.



(MBS reimbursement level varied from 100% down to 85% in this analysis.)

Figure 14 Sensitivity analysis for cost-effectiveness ratios of manual LBC (assuming $2.40 incremental cost for LBC).



(MBS reimbursement level varied from 100% down to 85% in this analysis.)

Financial implications

The annual financial implications of the decision to publicly fund LBC with manual reading and automation-assisted screening are presented in Table 57.

Annual cytology costs to Medicare were calculated by using estimates of potential utilisation (Table 53 and page 9), and the costs of each cytology technology (using a 100% MBS reimbursement level). The potential utilisation of conventional cytology was estimated at 1 901 913 tests annually, assuming management according to the 2005 NHMRC guidelines and incorporating HPV testing as a test of cure. Current utilisation in Australia is somewhat higher than this estimated value, at 2.1 million smears in 2007, which could in part be explained by transitional issues in the implementation of the 2005 NHMRC Guidelines and in post-treatment management with HPV test-of-cure.

Table 57 summarises the estimated utilisation of manual or automated LBC should these technologies be publicly funded. Their use is expected to reduce the unsatisfactory rate to 1.8 per cent, thus reducing recalls for primary screening. Use of manual LBC is expected to increase the number of low-grade positive test results by 0.9 per cent, and use of automated LBC is expected to increase it by 1.24 per cent, increasing follow-up

testing and management. Both manual and automated LBC are expected to result in a net increase in the number of Pap smears. Modelling predicts that the reduction in repeat tests following unsatisfactory samples would be countered by an increase in the number of satisfactory smear tests overall.

The costs of sample collection (i.e. consultation costs) will remain the same, irrespective of the cytological preparation or reading method used. Therefore, the incremental cost per patient of automated LBC is $16.40 ($36.00 less the current $19.60), and for LBC with manual reading is $2.40 to $10.90 (based on Applicant-requested reimbursements).

Automated reading of LBC slides is likely to result in increased efficiency via faster throughput of cytology slides. This increased efficiency would not, however, have an impact on the cost-effectiveness ratio of the technology unless it were reflected in the reimbursement level of the test.

The net financial impact to the Australian health care system was calculated on the basis of the modelled incremental lifetime costs. For this calculation, no discounting was applied, and the modelled costs for each age group were multiplied by the number of women in Australia of that age in 2007. Screening participation by age was incorporated into the model, and so is already accounted for in the age-specific costs. Total 'societal' costs (using 100% MBS reimbursement level) and the net impact on health care expenditure were considered. The final estimated cost includes all pathology, treatment, and consultation fees, but does not include screening program overheads. The estimates do not consider possible future population growth, changes in population age distribution, or further changes to the current Australian screening guidelines.

By the above method, the total cost of screening and management under current practice was estimated as approximately $166 million. If new technologies were introduced, the net annual costs were estimated as $173.4 million for manual LBC at the $2.40 incremental price, $189.7 million for manual LBC at the $10.90 incremental price, and $203.5 million for automated LBC. These represent cost increases of $7.3 million, $37.4 million, and $23.6 million each year, respectively, or increases of between 4.5 and 22.5 per cent.

It should be borne in mind that should the new technologies be introduced, it is likely that some tests would continue to be performed using conventional cytology, at least in a transition phase, and therefore the actual annual financial impact may be lower than the above estimates.

Table 57 Financial implications of current and new test technologies—annual utilisation and costs.

Strategy	Annual cytology utilisation		Annual MBS cytology costs [a]		Annual cost to health system [a]	
	Total	Incremental	Total ($ million)	Incremental ($ million)	Total ($ million)	Incremental ($ million)
Current practice (CC)	1 901 913	–	37.3	–	166.1	–
Manual LBC ($2.40 incremental cost)	1 924 675	22 763	42.3	5.1	173.4	7.3
Automated LBC	1 940 259	38 346	69.8	32.6	203.5	37.4
Manual LBC ($10.90 incremental cost)	1 924 675	22 763	58.7	21.4	189.7	23.6

[a] 100% Medicare reimbursement level.

Other relevant considerations

Adjunctive pathogen testing

The collection of cervical cytology samples into an LBC medium provides the opportunity for reflex testing of a range of pathogens, in addition to any impact on detecting CIN or compatibility with automated screening technologies. If a cytology sample is collected into an LBC medium, testing can be conducted for HPV, *Chlamydia trachomatis* or *Neisseria gonorrhoeae*.

Workforce implications

Other factors that may play a role in the consideration of funding automated LBC technology are the implications for the workforce involved in cervical cancer screening.

There is an increasing shortage of trained cytotechnologists in Australia (Advisory Panel expert opinion). The job satisfaction of cytotechnologists has also been reported to be low (Dowie et al. 2006b). Technologies which decrease cytology slide screening time and increase productivity may aid in addressing workforce shortages by decreasing staff requirements (Australian Health Technology Advisory Committee 1998). The included studies demonstrated a significant decrease in slide reading time with automated LBC compared to manual reading of conventional slides. This review did not address a comparison of slide reading times for manual reading of conventional slides versus manual reading of LBC slides, however any advantage gained with this technology may be relatively minor (Dowie et al. 2006a).

In the near future, the introduction of the HPV vaccine in Australia will lead to a decrease in the prevalence of HPV and pre-cancerous cytological abnormalities and may also alter the distribution of cytological abnormalities (Schiffman 2007), increasing technical difficulties for cytotechnologists manually screening slides even further.

Discussion

LBC

Data from two systematic reviews designed to compare the performance of LBC to conventional cytology have indicated in subgroup analyses that there may be variations in accuracy and unsatisfactory rates according to proprietary name.

A recent, high-quality systematic review (Arbyn et al. 2008) did not demonstrate a statistically significant difference in sensitivity to detect CIN 2+ for LBC preparation methods compared to conventional Pap cytology (HSIL threshold, LBC:conventional sensitivity ratio 1.05, 95% CI 0.95–1.16). This conclusion did not vary if only studies conducted in a screening population, or of paired design were considered. This review reported LBC was less specific than conventional cytology at a test threshold of pLSIL (ASCUS+, LBC:conventional specificity ratio 0.91, 95% CI 0.84–0.98). Study design and proprietary name of the LBC technology contributed to study heterogeneity for specificity only. A separate HTA and meta-analysis demonstrated that LBC classified significantly more slides as LSIL+ than conventional cytology, but classification of slides as HSIL+ did not significantly differ (Krahn et al. 2008).

This recent HTA report from CADTH concluded that LBC may have a lower unsatisfactory rate than conventional cytology, but that the estimate of difference varied between studies (Krahn et al. 2008). Possible reasons for the variations include differences in reporting terminology and screening practices. An earlier systematic review by Davey et al. (2006) found no difference in the unsatisfactory rate between LBC and conventional cytology in a meta-analysis of 48 datasets. A summary of findings in this systematic review is provided in Appendix I. This review stimulated debate in the literature as some prominent individual studies had indicated a decrease in the unsatisfactory rate with the use of LBC. In particular, the experience of introducing LBC into the cervical cancer screening program in the United Kingdom reduced unsatisfactory rates from 9.1 per cent with conventional Pap slides to an average of 1.6 per cent with LBC (Moss et al. 2004; 87% reduction, $P < 0.0001$; National Institute for Clinical Excellence 2003). However, this reduction came from a base rate of unsatisfactory conventional smears that is much higher than that seen in the Australian screening program (9.1% UK vs 2.1%–2.3% [1] Australia). Davey et al. (2006) concluded that large randomised trials are needed to assess the comparative performance of LBC and conventional cytology.

A recent large RCT of 45, 174 women reported a decrease in the unsatisfactory rate for LBC (ThinPrep) versus conventionally prepared slides (2.6% LBC vs 4.1% conventional, relative frequency 0.62, 95% CI 0.56–0.69). The unsatisfactory rate in the conventional arm of this Italian study (4.1%) was higher than that seen in the Australian screening program (2.1% [1]), but more similar to Australian rates than were the UK rates.

[1] Victoria 2007, 2.1% (11 259/546 012) unsatisfactory (pers. comm., Cathryn Wharton, Victorian Cervical Cytology Register). NSW, fourth quarter of 2006, 2.3% (3843/163 568) unsatisfactory (pers. comm., Grace Kwan, Cancer Institute NSW).

A modelled analysis of cervical cancer screening, diagnosis and treatment is necessary to explore the potential benefits and trade-offs of using LBC with manual slide reading instead of conventional cytology.

Modelled analysis estimated that the replacement of conventional cytology in Australia with manual LBC would result in 23 fewer cancer cases, and 6 fewer cancer deaths. It also predicted, however, that LBC would substantially increase investigations and treatments. It estimated an additional 22 763 smears, 6770 colposcopies, 3273 biopsies, and 735 treatments for CIN 2/3. This analysis was based upon favourable assumptions about test specificity. Whether this outcome is considered as an improvement in overall health outcomes is dependent upon the trade-off between the benefit from a reduction in cancer incidence and the harms associated with increased investigations for the women who are not destined to develop cervical cancer.

Automated LBC

No eligible studies addressing the primary review question of the comparative accuracy of the FocalPoint system for reading LBC slides compared to manual reading of conventional slides were identified.

Two studies of AutoPap addressed secondary review questions. A single fair-quality study providing level III-2 evidence of accuracy comparing AutoPap-assisted reading of conventional slides with manual reading of conventional slides did not provide any evidence of an advantage, disadvantage or equivalence. Similarly, a study of highly limited applicability comparing the AutoPap system to manual reading of LBC did not provide any evidence of an advantage, disadvantage or equivalence.

Two studies found that automation-assisted reading of LBC slides with the ThinPrep Imager system was associated with a significant decrease in slide reading time and a significant increase in the number of slides classified with an abnormality compared to conventional cytology. One study indicated that the number of slides classified as unsatisfactory was also significantly decreased with the ThinPrep Imager system

A summary of the body of evidence for the comparative accuracy of the ThinPrep Imager with conventional cytology is shown in Table 58. One fair-quality study providing level III-2 evidence of accuracy indicated that automation-assisted reading of LBC slides with the ThinPrep Imager system detected significantly more high-grade lesions than manual reading of conventional cytology (Davey et al. 2007a). This was associated with an increase in the number of women classified as having low-grade lesions on cytology.

A second fair-quality study indicated a non-significant trend toward an increase in the number of true positive cases detected by the ThinPrep Imager system compared to conventional cytology (Roberts et al. 2007). This study is considered to provide less robust evidence than the Davey et al. (2007a) study as there is a possibility of verification bias. The number of discordant positive slides verified with each technology is not reported. The findings of this study are difficult to interpret given that the head-to-head direct comparison of conventional cytology, automation-assisted reading of LBC and manual reading of LBC indicated a significant advantage of manual LBC but not automation-assisted LBC over that of manual reading of conventional cytology.

There was inconsistency between the numbers of false positive biopsy results at a pHSIL test threshold in the two studies comparing the ThinPrep Imager system to conventional

cytology. There was no significant difference in the study by Davey et al. (2007a), but there were significantly more false positive findings with the ThinPrep Imager system in Roberts et al. (2007).

Table 58 Body-of-evidence matrix for accuracy of ThinPrep Imager vs conventional cytology.

Component	A Excellent	B Good	C Satisfactory	D Poor
Evidence base				2 level III studies with some risk of bias
Consistency			Some inconsistency reflecting genuine uncertainty around clinical question—detection of CIN 2+	Evidence is inconsistent—FP cases
Clinical Impact				Unknown [a]
Generalisability	Populations studied in body of evidence are the same as the target population			
Applicability		Applicable to Australian health care context with few caveats [b]		

[a] Relates to impact of automated LBC on patient outcomes. Requires modelling of linked evidence to determine. [b] Some limitations related to differences from current cytological terminology.

Three studies comparing the accuracy of ThinPrep Imager reading of LBC slides to manual reading of LBC slides did not demonstrate a significant difference in the number of true cases detected. In two of three studies the ThinPrep Imager identified significantly fewer false positive cases for high-grade lesions at a test threshold of pHSIL or HSIL, respectively. There is a possibility of verification bias in one of these studies.

In summary, a single study conducted in an Australian setting has demonstrated an advantage of ThinPrep Imager reading of LBC slides over conventional imaging in terms of number of high-grade cases detected. A second Australian study detected a non-significant increase in the number of cases. This is associated with an increase in the cytological classification of women with low-grade lesions, and a decrease in the unsatisfactory rate and slide processing time. It is unclear whether any advantage is attributable to automation-assisted reading system, to LBC alone, or a combination of both.

A modelled analysis of cervical cancer screening, diagnosis and treatment is necessary to explore the potential long-term benefits and trade-offs of using automated LBC instead of conventional cytology.

Modelled analysis estimated that the replacement of conventional cytology in Australia with automated LBC would result in 68 fewer cancer cases, and 19 fewer cancer deaths, although as favourable assumptions were made about test characteristics, these could be overestimates. This would be associated with an increase in investigations and treatments, specifically a predicted additional 38 346 smears, 10 788 colposcopies, 5154 biopsies, and 1751 treatments for CIN 2/3. Whether this outcome is considered as an improvement in overall health outcomes is dependent upon the trade-off between the benefit from a reduction in cancer incidence and the harms associated with increased investigations for the women who are not destined to develop cervical cancer.

Cost-effectiveness analysis

A modelled analysis based on favourable assumptions regarding test characteristics found that manual LBC was associated with a cost of $126 315 to $385 982 per LYS, depending on the level of reimbursement for LBC, and that automated LBC was associated with a cost of $194 835 per LYS.

The results presented are based on the current screening program in Australia without taking into account potential changes resulting from HPV vaccination. The findings may be different for different screening populations, for different screening programs, or once the anticipated impact of vaccination on the incidence of cervical neoplasia and precursor lesions occurs. The current Australian screening program of biennial screening of women aged 18–69 is more intensive than many programs internationally. As more tests are performed at a primary screening level in Australia, annual incremental costs associated with the new technologies are higher. If changes to the Australian screening program are considered in the future, and as changes due to vaccination are realised, reassessment of the cost-effectiveness of these technologies, using similar methods, would be warranted as part of any review of different screening strategies and technologies.

Quality-of-life issues

This modelled analysis focuses on the implications for overall survival of introducing new cytology technologies. An analysis incorporating quality-adjusted survival (QALYs) has not been conducted, because of paucity of data to inform QALY weights. The calculation of a QALY depends on the relative weighting attached to an entire sequence of health states and events modelled over the duration of a lifetime. This will be influenced by an individual's perceived trade-off of benefits against potential downsides for each health state. For this reason the cost per QALY gained may be either higher or lower that the reported cost per LYS from the current analysis. Additional data on QALY weights are required before we can adequately capture the likely survival and quality-of-life trade-offs in a QALY.

Comparison with previous findings

Comparison with previous MSAC findings (reference 12a)

A previous MSAC assessment of LBC (MSAC 2002c) judged that manual LBC would not be cost-effective. It found that if a reduction in the probability of an unsatisfactory sample of 2.8 per cent with LBC was assumed, an incremental cost of $0.53 for LBC would be budget-neutral.

At the time of the previous MSAC application, there was less published data on the accuracy of manual LBC relative to conventional cytology, and so equivalent test characteristics for LBC and conventional cytology were assumed. Data have since been published which indicate that manual LBC has a lower specificity for the detection of CIN 2+ at a test threshold of pLSIL, and that sensitivity at these thresholds is not significantly improved. Lower specificity and equal sensitivity implies that the test will result in higher costs but no improvement in outcomes, and incremental cost-effectiveness will therefore be reduced.

The previous MSAC models used estimates for the probability of an unsatisfactory sample for each test type based on earlier data {Roberts, 1997 3 /id}. The use of more

recent data resulted in different estimates for unsatisfactory smear rates, as described in 'Cytology unsatisfactory rates' on page 69. The assumed decrease in unsatisfactory rates used in the previous models was 2.8 per cent, whereas we assumed a range of 0 to 2.4 absolute percentage points for manual LBC. The base rate previously used for CC (3.5%) is higher than is reflected in recent registry data (2.1%–2.3%). A greater reduction in unsatisfactory specimens will tend to favour LBC. We considered this in the sensitivity analyses, and found that under extreme assumptions about unsatisfactory rates (a reduction from 5% to 0.5%), manual LBC was associated with a cost of $47 470 per LYS, assuming an incremental cost of $2.40. Automated LBC was associated with a cost of $168 410 per LYS under these extreme assumptions about unsatisfactory rates. In practice, an improvement in unsatisfactory rates of this magnitude is highly unlikely to be achieved in Australia, because the current rate of unsatisfactory conventional cytology samples is approximately 2.2 per cent.

The model submitted with the previous MSAC assessment, and the alternative model also described in that assessment, differ significantly from those used here and from that provided in the current application. Previously, the decision models used had a 2-year time horizon and were therefore unable to calculate lifetime costs and effects. The model structures were also unable to capture the underlying disease processes, and the detailed screening, diagnostic, treatment, and follow-up processes which occur.

A further difference between the previous and the current evaluation is that management guidelines and cytology reporting standards have been updated since the previous application was made.

Comparison with results in other countries

Manually read LBC has been found to be cost-effective and has been adopted in a number of countries, including the United Kingdom and New Zealand. We investigated the underlying reasons why results were different from those found here for Australia, and identified the following key factors:

1. Differences in the reduction in the rate of unsatisfactory samples

   Before the introduction of LBC in comparable countries such as New Zealand and the United Kingdom, the rate of unsatisfactory smears associated with conventional cytology was substantially higher than that currently experienced in Australia. In our sensitivity analysis for unsatisfactory cytology rates, we found that assuming a higher unsatisfactory smear rate for conventional cytology and a reduction in unsatisfactory samples of 2.4% absolute percentage points, the cost per LYS reduced by 30 per cent. At a reduction of 4.5 per cent, the cost per LYS was reduced by 62 per cent. Under the latter scenario, which was the most favourable to LBC, the cost per LYS was $47 470. Therefore, differences in the unsatisfactory rate for conventional cytology, and the reduction in unsatisfactory rates attributable to LBC, experienced in other countries were a significant factor in explaining differences in the cost-effectiveness of LBC between Australia and these countries.

2. Differences in screening intervals

   New Zealand has a recommended screening interval of 3 years. The United Kingdom has a recommended screening interval of 3 years for women aged 25–49 years, and 5 years for women aged 50–64 years. We found that cost-

effectiveness was moderately sensitive to this parameter: modelling screening behaviour in Australia assuming a 3-yearly interval recommendation increased the cost-effectiveness of the new technologies.

Therefore, Australia differs from other comparable countries in key areas which have a substantial impact on the cost-effectiveness of liquid-based technologies.

Conclusions

Safety

Automated LBC and manual LBC are considered safe procedures.

Effectiveness

No studies have assessed the impact of LBC with manual or automated slide reading on incidence or mortality rates of invasive cervical cancer compared to conventional cytology. The present review therefore relies on evidence about the relative accuracy of manual or automated LBC to detect precancerous cervical lesions to draw conclusions about its relative effectiveness. This 'linked evidence' approach is justified by existing evidence that early detection and treatment of precancerous cervical lesions leads to a reduction in the incidence and mortality of cervical cancer.

Manual LBC

High-quality systematic reviews and a large randomised trial have indicated that liquid-based cytology compared to conventional cytology

- provides no statistically significant increase in sensitivity or specificity
- provides no statistically significant difference in sensitivity (HSIL, LSIL or pLSIL thresholds) or specificity (HSIL or LSIL thresholds) for the detection of CIN 2+
- reduces the specificity for the detection of CIN 2+ at a test threshold of pLSIL
- classifies more slides as positive for low-grade lesions
- reduces the rate of unsatisfactory smears

Automated LBC

A single Australian fair-quality study providing level III-2 evidence of accuracy indicated that automation-assisted reading of LBC slides with the ThinPrepImager system compared to manual reading of conventional cytology

- detects statistically significantly more true high-grade lesions
- increases the number of slides classified as having low-grade lesions on cytology
- did not increase the number of false positive cases at a pHSIL test threshold
- reduces the rate of unsatisfactory slides

A second fair-quality Australian study providing level III-2 evidence of accuracy of the ThinPrepImager system, that contains some possibility of verification bias, found

- no statistically significant difference in the number of high-grade lesions detected compared to manual reading of conventional cytology

- a statistically significant increase in the number of high-grade lesions detected by manual LBC compared to conventional cytology

- statistically significantly more false positive findings at a pHSIL test threshold compared to manual reading of conventional cytology

Two studies indicated that the use of the ThinPrep Imager system significantly decreased slide reading time compared to manual reading of conventional cytology.

Three studies comparing ThinPrep Imager reading of LBC to manual LBC concurrently did not find any significant increase in the detection of true positive cases with automated LBC compared to manual LBC. The ThinPrep Imager was associated with a decrease in the detection of false positive cases in two of three studies.

It is concluded that in an Australian setting, automated LBC reading with the ThinPrep Imager system

- detects at least as many CIN 2+ lesions as conventional cytology, and may detect more

- increases the cytological classification of women with low-grade lesions

- reduces the unsatisfactory slide rate

- reduces slide processing time.

Economic considerations

Cost-effectiveness

The main findings of the economic evaluation are that both technologies under consideration—automated LBC and LBC with manual reading—had high cost-effectiveness ratios, which would appear to be unfavourable in Australia at the currently requested level of reimbursement.

Cost-effectiveness ratios were estimated from lifetime costs and LYS, both of which were discounted at 5 per cent per annum. Automated LBC cost $194 835 per LYS. The cost of manual LBC varied depending on the level of reimbursement, but ranged from $126 315 per LYS ($2.40 incremental cost) to $385 982 per LYS ($10.90 incremental cost).

The findings are sensitive to assumed relative test accuracy, differences in the unsatisfactory smear rate, assumptions about disease natural history (particularly high-grade regression and progression) and the recommended screening interval. Favourable assumptions were made about the accuracy of the new technologies. On the basis of these favourable assumptions, both technologies would result in an improvement in LYS, but at a substantially higher cost, due mainly to direct cytology test costs, but also to follow-up costs for an increased number of test positives.

Financial impact

Modelled analysis predicted that direct replacement of conventional cytology with automated LBC would be associated with an annual increase in health system costs of $37.4 million. Incremental health system costs associated with manual reading of LBC ranged from $7.3 million to $23.6 million per annum, depending on the level of reimbursement for manual LBC.

Appendix A MSAC terms of reference and membership

MSAC's terms of reference are to:

- advise the Minister for Health and Ageing on the strength of evidence pertaining to new and emerging medical technologies and procedures in relation to their safety, accuracy, effectiveness and cost-effectiveness and under what circumstances public funding should be supported;
- advise the Minister for Health and Ageing on which new 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 and/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 AHMAC.

The membership of MSAC at the March 2009 meeting comprised a mix of clinical expertise covering pathology, nuclear medicine, surgery, specialist medicine and general practice, plus clinical epidemiology and clinical trials, health economics, consumers, and health administration and planning:

Member	Expertise or affiliation
Professor Robyn Ward (Chair)	medical oncology
Dr William Glasson (Deputy Chair)	ophthalmology
Associate Professor Frederick Khafagi (Deputy Chair)	nuclear medicine
Associate Professor John Atherton	cardiology
Professor Justin Beilby	health research
Professor Jim Butler	health economics
Professor Peter Cameron	trauma and emergency medicine
Associate Professor Kirsty Douglas	health research
Dr Kwun Fong	thoracic medicine
Professor Richard Fox	medical oncology
Professor John Horvath	Commonwealth Chief Medical Officer
Professor Helen Lapsley	health economics
Mr Russell McGowan	consumer health issues
Professor Julia Potter	pathology
Dr Ian Prosser	haematology
Dr Judy Soper	radiology
Dr Graeme Suthers	genetics/medical oncology
Dr Shiong Tan	general practice
Professor Ken Thomson	radiology
Professor Andrew Wilson	public health physician
Dr Caroline Wright	colorectal surgery

Appendix B Advisory Panel and Health Technology Assessors

Advisory panel for application 1122: Automation-assisted and liquid-based cytology for cervical cancer screening

Member	Nomination/Expertise or affiliation
Professor Brendon Kearney (Chair)	Health Administration and Planning
Associate Professor David Allen	Royal Australian and New Zealand College of Obstetricians and Gynaecologists nominee Gynaecological Oncologist
Dr Louise Farrell	Royal Australian and New Zealand College of Obstetricians and Gynaecologists nominee Obstetrician & Gynaecologist, special interest in colposcopy
Professor Suzanne Garland	Royal College of Pathologists of Australia nominee Clinical Microbiologist and Sexual Health Physician
Dr Huw Llewellyn	Royal College of Pathologists of Australia nominee Population health
Associate Professor Danielle Mazza	Royal Australian College of General Practitioners General Practice
Dr Ewa Piejko	General Practice
Associate Professor Marion Saville	Australian Society of Cytology nominee Gynaecological pathology
Dr Simon Towler	Co-opted member/ W.A. Department of Health Health Administration and Planning/Population Health
Ms Diane Walsh	Consumer Health Forum nominee Consumer Health Issues

Health Technology Assessors

Dr Suzanne Dyer, Project Manager **Dr Sally Lord,** Epidemiologist **Ms Sally Wortley**, Project Officer	NHMRC Clinical Trials Centre, University of Sydney
Dr Kirsten Howard, Health Economist	Screening and Test Evaluation Program, University of Sydney
Dr Karen Canfell, Research Fellow **Ms Megan Smith,** Program Manager **Ms Jie Bin Lew**, Senior Research Programmer **Ms Prudence Creighton**, Research Programmer **Ms Yoon Jung Kang**, PhD student	Cancer Epidemiology Research Unit, Cancer Council NSW
Dr Mark Clements, Research Fellow	National Centre for Epidemiology and Population Health, Australian National University

Appendix C FIGO Staging of cervical cancer

The Federation Internationale de Gynecologie et d'Obstetrique (FIGO) classification system for staging cervical cancer:

Stage I
Stage I is carcinoma strictly confined to the cervix; extension to the uterine corpus should be disregarded. The diagnosis of both Stages IA1 and IA2 should be based on microscopic examination of removed tissue, preferably a cone, which must include the entire lesion.
Stage IA: Invasive cancer identified only microscopically. Invasion is limited to measured stromal invasion with a maximum depth of 5 mm and no wider than 7 mm.
Stage IA1: Measured invasion of the stroma no greater than 3 mm in depth and no wider than 7 mm diameter.
Stage IA2: Measured invasion of stroma greater than 3 mm but no greater than 5 mm in depth and no wider than 7 mm in diameter.
Stage IB: Clinical lesions confined to the cervix or preclinical lesions greater than Stage IA. All gross lesions even with superficial invasion are Stage IB cancers.
Stage IB1: Clinical lesions no greater than 4 cm in size.
Stage IB2: Clinical lesions greater than 4 cm in size.
Stage II
Stage II is carcinoma that extends beyond the cervix but does not extend into the pelvic wall. The carcinoma involves the vagina, but not as far as the lower third.
Stage IIA: No obvious parametrial involvement. Involvement of up to the upper two thirds of the vagina.
Stage IIB: Obvious parametrial involvement, but not into the pelvic sidewall.
Stage III
Stage III is carcinoma that has extended into the pelvic sidewall. On rectal examination, there is no cancer-free space between the tumour and the pelvic sidewall. The tumour involves the lower third of the vagina. All cases with hydronephrosis or a non-functioning kidney are Stage III cancers.
Stage IIIA: No extension into the pelvic sidewall but involvement of the lower third of the vagina.
Stage IIIB: Extension into the pelvic sidewall or hydronephrosis or non-functioning kidney.
Stage IV
Stage IV is carcinoma that has extended beyond the true pelvis or has clinically involved the mucosa of the bladder and/or rectum.
Stage IVA: Spread of the tumour into adjacent pelvic organs.
Stage IVB: Spread to distant organs.

Source: TNM Classification of Malignant Tumours. L. Sobin and Ch Wittekind (eds.), UICC International Union against Cancer, Geneva, Switzerland, pp 155–157; 6th ed. 2002.

Appendix D MSAC 2002 and 2003 reviews

Author/year Country	Objective of report Databases & dates searched	Type of analysis, number & date of included studies	Population considered in included studies Test comparison	Conclusions/recommendation	Quality assessment
MSAC (Medical Services Advisory Committee 2003) Reference 12c Australia	To conduct a review of the literature on computer-assisted image analysis for cervical screening cytology <u>Search</u> • To Sept 2002 • Cochrane Library, MEDLINE, PreMEDLINE, Current Contents, Biological Abstracts, CINAHL, EBM reviews, CancerLit, HTA databases & websites • AutoCyte SCREEN and PAPNET from 1997 • AutoPap from 2000	Systematic review <u>Included studies</u> Three systematic reviews: • AHTAC 1998 • Broadstock NZHTA 2000 • Hartmann et al. 2001 Nil primary studies	Population Women undergoing cervical screening Test comparison Computer-assisted analysis as an adjunct to, or replacement for, primary manual screening vs Manual reading alone Considered use in primary screening, rescreening and triage	Authors' conclusion There is a lack of evidence that computer-assisted image analysis is as effective as conventional manual screening for cervical screening cytology. Recommendation As there is insufficient evidence to draw conclusions on the appropriate use of computer-assisted image analysis vs manual cytology screening, there are no grounds to change current funding arrangements.	Quality: Explicit review questions: NR Explicit & appropriate eligibility criteria: yes [a] Explicit & comprehensive search strategy: yes Quality of included studies appraised: yes Methods of study appraisal reproducible: yes Heterogeneity between studies assessed: N/A Summary of main results clear and appropriate: yes Applicability: No information presented on currently commercially available, integrated, automated LBC slide-reading systems Excluded studies where computer-assisted image analysis used for triage of patients with ASCUS or AGUS result. No studies excluded considered technologies currently commercially available
Results: The HTA report by Broadstock 2000 (NZHTA) included only studies of AutoPap, system which is used in conjunction with conventional slides. The AHTAC 1998 report also reviewed AutoPap, in addition to PAPNET and AutoCyte Screen, which are used for rescreening conventional slides. The systematic review by Hartmann et al. 2001 (AHCPR) did not identify any studies that met the selection criteria. The two included HTAs by Broadstock 2000 (NZHTA) and AHTAC 1998 both 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.					
MSAC (Medical Services Advisory Committee 2002b)	To conduct a review of the literature and cost-effectiveness analysis of LBC for cervical	Systematic review and modelled economic analysis. <u>Included studies</u>	Population Women undergoing cervical screening	Authors' conclusion There is insufficient evidence to draw conclusions regarding the diagnostic	Quality: Explicit review questions: NR Explicit & appropriate eligibility criteria:

Reference 12a Australia	screening cytology <u>Search</u> • Jan 2000 to April 2002 • Cochrane Library, MEDLINE, PreMEDLINE, Current Contents, Biological Abstracts, CINAHL, EBM reviews, HTA databases & websites	5 HTAs/systematic reviews: • Broadstock NZHTA 2000 • Hartmann et al. 2001 • Sulik 2001 • Moseley & Paget 2002 • Bernstein 2001 meta-analysis (ThinPrep) 7 primary studies: • 2 split-sample studies • 1 RCT • 2 historical control studies • 1 study with historical & concurrent controls • 1 prospective independent sample study	Test comparison LBC with manual reading as a replacement for the Papanicolaou smear	characteristics of LBC and Pap smears for cervical screening. The lack of high-quality evidence on the performance of LBC does not permit evaluation of whether it is equal or superior in effectiveness to Pap smears. Recommendation Since there is currently insufficient evidence pertaining to LBC for cervical screening, the MSAC recommends that public funding should not be supported at this time for this screening test.	yes [a] Explicit & comprehensive search strategy: yes Quality of included studies appraised: yes Methods of study appraisal reproducible: yes Heterogeneity between studies assessed: N/A Summary of main results clear and appropriate: yes Applicability: Included systematic reviews generally pooled data from different manufacturers and included split-sample studies. Sensitivity and specificity from primary studies in high risk populations.

Results:

<u>Systematic reviews</u>

Secondary studies of LBC concluded that there is insufficient high-quality data or evidence to suggest that LBC is better than the Pap smear for cervical screening.

However, Sulik et al. (2001) suggest that there may be a role for LBC for women who have had abnormal Pap test results or who are at a high risk of cervical cancer owing to infrequent screening. Bernstein et al. (2001) deduced that the LBC test improved sample adequacy and led to improved diagnosis of LSIL and HSIL, but results comparing LBC and Pap tests were not evaluated against a histological reference test.

All authors noted that the most frequent study design was the split-sample method, and that many of the clinical studies examined were funded partially or completely by manufacturers of LBC technologies.

<u>Primary studies</u>

All 7 studies included problems with the calculation of sensitivity and specificity.

Authors extracted test parameters from 2 studies with reference standard of cervical histology:

- Bergeron et al. (2001), AutoCytePrep, HSIL+ test threshold, LBC Se 41.7%–82.6%, Sp 52.2%–90.2%; Pap Se 39.4%–86.8% Sp 47.8%–98.9%; Unsat/ASCUS/AGUS+ LBC Se 82.6%, Sp 52.2%; Pap Se 86.8% Sp 47.8%.
- Park et al. (2001), ThinPrep, ASCUS+ test threshold, LBC Se 82.8%, Sp 62.0%; Pap Se 89.6%, Sp 52.1%.
- Studies failed to meet several validity criteria and had probable presence of non-appraisable bias. Included high-risk populations.

Five other studies did not include reference standard for negative tests, thus sensitivity & specificity not considered valid. The relative TPR and FPR (LBC/Pap) were:

- Anton et al. (2001), manufacturer NR, relative TPR 1.17 ($P < 0.01$), relative FPR 1.73
- Bai et al. (2000), ThinPrep, ASCUS+ rTPR 1.25 ($P < 0.005$), rFPR ($P < 0.005$); HSIL+ rTPR 1.52 ($P < 0.025$), rFPR 0.82 ($P < 0.025$)
- Obwegeser et al. (2001), ThinPrep, HSIL+ rTPR 0.91 ($P < 0.0005$)
- Limitations related to several validity criteria, no verification of negative results, lack of definition of time limit for histological verification and interpretation of overseas classification systems.

Abbreviations: AHTAC = Australian Health Technology Advisory Committee, NR = not reported, NZHTA = New Zealand Health Technology Assessment, Se = selectivity, Sp = specificity; rTPR =relative true positive rate, rFPR = relative false positive rate [a] Application of criteria unclear—see applicability.

Appendix E Databases searched

Table 59 Electronic databases and websites searched.

1. International electronic databases—searched to August 2008
NHS Centre for Reviews and Dissemination databases/ International Network of Agencies for Health Technology Assessment (INAHTA) Economic Evaluation Database (EED) Database of Abstracts of Reviews of Effectiveness (DARE) Heath Technology Assessment (HTA) Cochrane Database of Systematic Reviews and Cochrane Controlled Trials Register
2. Individual HTA Agencies—searched to November 2007
AUSTRALIA
Australian Safety and Efficacy Register of New Interventional Procedures—Surgical (ASERNIP-S)
Centre for Clinical Effectiveness, Monash University
Health Economics Unit, Monash University
Medical Services Advisory Committee
AUSTRIA
Institute of Technology Assessment / HTA unit
CANADA
Agence d'Evaluation des Technologies et des Modes d'Intervention en Santé (AETMIS)
Canadian Agency for Drugs and Technologies in Health (CADTH) (previously Canadian Coordinating Office for Health Technology Assessment, CCHOTA)
Centre for Health Economics and Policy Analysis (CHEPA), McMaster University
Centre for Health Services and Policy Research (CHSPR), University of British Columbia
Institute for Clinical and Evaluative Studies (ICES)
Institute of Health Economics (Alberta)
DENMARK
Danish Institute for Health Technology Assessment (DIHTA)
FINLAND
FINOHTA
FRANCE
Haute Autorité de Santé (HAS; French National Authority for Health)
GERMANY
German Institute for Medical Documentation and Information (DIMDI) / HTA
THE NETHERLANDS
Health Council of the Netherlands Gezondheidsraad
NEW ZEALAND
New Zealand Health Technology Assessment (NZHTA)
NORWAY
Norwegian Centre for Health Technology Assessment (SMM)
SPAIN
Catalan Agency for Health Technology Assessment (CAHTA)

SWEDEN
Swedish Council on Technology Assessment in Health Care (SBU)
Center for Medical Health Technology Assessment
SWITZERLAND
Swiss Network on Health Technology Assessment (SNHTA)
UNITED KINGDOM
Health Technology Board for Scotland
National Health Service Health Technology Assessment (UK) / National Coordinating Centre for Health Technology Assessment (NCCHTA)
University of York NHS Centre for Reviews and Dissemination (NHS CRD)
National Institute for Clinical Excellence (NICE)
UNITED STATES
Agency for Healthcare Research and Quality (AHRQ)
US Blue Cross / Blue Shield Association Technology Evaluation Center (TEC)
3. Trials registries
Clinical trials Searched February 2008
NIHR Health Technology Assessment Programme—research in progress Searched February 2008
World Health Organization International Clinical Trials Registry Platform (ICTRP) Searched November 2007

Appendix F Included studies

Author/year Country	Objective of report Databases & dates searched	Type of analysis, number & date of included studies	Population considered in included studies Test comparison	Conclusions/recommendation	Quality assessment
HTAs and systematic reviews					
CADTH (2008) (Krahn et al 2008) Technology Report Issue 103 Canada	To assess: • the effectiveness and cost-effectiveness of LBC compared to conventional cytology • which subpopulations and population-based parameters influence estimates of effectiveness or cost-effectiveness (screening participation rates, HPV prevalence) • how HPV testing affects the cost-effectiveness of LBC • the budget impact of adopting LBC + HPV triage for cervical cancer screening <u>Search</u> • Nov 2002 to June 2006 • BIOSIS Previews, CancerLit, EMBASE, MEDLINE, Cochrane Library	Update systematic review, Bayesian meta-analysis and modelled economic analysis <u>Included studies</u> 21 secondary studies: • 19 HTAs/systematic reviews/economic evaluations identified in update 108 primary studies: • 44 identified in update • 20 direct comparison studies • 49 LBC studies • 47 split-sample • 31 two-cohort studies • 66 studies with unsatisfactory rate	Population Women undergoing cervical screening, sexually active and ≥ 15 years Test comparison LBC with manual reading, or LBC with HPV triage of ASCUS or AGUS versus conventional cytology with manual reading	Authors' conclusion (LBC only) No statistical differences in sensitivity and specificity between LBC and CC. LBC is estimated to be on average 6% more sensitive and 4% less specific than CC across a range of cytological thresholds. The health economic evidence suggests that two-year screening strategies using HPV triage, ± LBC, represents the best use of resources for cervical cancer screening	Quality: high Explicit review questions: yes Explicit & appropriate eligibility criteria: yes Explicit & comprehensive search strategy: yes Quality of included studies appraised: yes (no requirement of histology reference standard for high quality) Methods of study appraisal reproducible: yes Heterogeneity between studies assessed: yes Summary of main results clear and appropriate: yes Applicability: LBC manufacturer data not pooled Accuracy data has limited applicability, as main analysis includes studies of any quality with cytology or consensus reading as a reference standard. Influence of reference standard validity and quality studies only explored separately. High-quality studies did not require histology as a reference standard No subgroup analyses for unsatisfactory rates by study quality/design, and non-comparative studies included

Author/year Country	Objective of report Databases & dates searched	Type of analysis, number & date of included studies	Population considered in included studies Test comparison	Conclusions/recommendation	Quality assessment
Results: Previous systematic reviews Six systematic reviews concluded that LBC is more effective than CC; 5 concluded that there was insufficient evidence to draw conclusions about LBC effectiveness; 3 were equivocal in their conclusion Meta-analysis—accuracy The meta-analysis of data from 20 head-to-head studies including 68 114 participants showed no statistical differences in sensitivity and specificity between LBC and CC. The average sensitivity difference for LBC relative to CC was 6.4% (95% CI: –6.5% to +18.8%). The average specificity difference for LBC relative to CC was –4.0% (–19.8% to +10.6%). Neither trade-off estimate was statistically significant as indicated by the CI boundaries that included the value of zero (i.e. no difference). Sensitivity analyses were conducted considering high-quality studies, studies with histology as a reference standard, ThinPrep, and different test and reference standard thresholds. The trade-off estimates from the main analysis were similar to those from the sensitivity analysis that was restricted to 6 high-quality studies. One of these studies used independent cytological review rather than histology as a reference standard. Restricting data to the 13 studies in which a histology classification from a combination of colposcopy and biopsy was used as a reference standard changed the trade-off estimates. In this sensitivity analysis, the average sensitivity difference for LBC relative to CC was 1.2% (–16.6% to +16.4%), and the average specificity difference was –0.6% (–18.9% to +17.3%). The trade-off estimates for ThinPrep (*n* = 17) were consistent with those of the main analysis, most likely because most of the data came from studies of this technique. Meta-analysis—unsatisfactory rates Data from 44 ThinPrep studies. Overall 0.95% (6674/704 813) LBC slides and 1.04% (13 664/1 316 318) conventional slides were unsatisfactory. A Bayesian meta-analysis gave pooled unsatisfactory estimates of 2.24% (95% CI 1.20%–3.29%) for LBC and 3.04% (1.92%–4.16%) for conventional slides, with a difference of –0.81 (–1.87% to +0.24%). Statistically significant heterogeneity was observed between studies in this outcome. Data from 15 studies of SurePath. Overall, 2539 (0.42%) of 597 565 LBC slides were unsatisfactory, whereas 9598 (1.39%) of 692 406 CC slides were unsatisfactory. A Bayesian meta-analysis gave pooled unsatisfactory estimates of 0.82% (95% CI 0.14%–1.51%) for LBC and 3.31% (0.97%–5.67%) for conventional slides, with a difference of –2.49 (–4.43% to –0.55%). On average, the percentage of unsatisfactory slides from LBC samples was less compared with that from CC samples. The estimates from individual studies were heterogeneous, reflecting differences in screening programs, smear settings, and practices. In studies with large sample sizes (i.e. small CIs around the difference estimates of LBC vs CC), the difference estimates between the two techniques were close to zero. Economic analysis Utilises accuracy data from meta-analysis of 20 studies, considering LBC increase in sensitivity of 6.4% (95% CI: –6.5% to +18.8%), decrease in specificity of 4.0% (–19.8% to +10.6%). CC screening every 2 years was the least expensive option, but life expectancy was reduced compared with annual CC screening. Two-yearly screening with LBC vs CC 2-yearly cost ~$31 000 per LYS or $29 000 per QALY gained. Annual CC screening would be the next best option to be considered an improvement from LBC screening every 2 years. Compared with LBC screening every 2 years, annual CC screening was more expensive than LBC screening every 2 years but extended life expectancy by a small margin. Annual CC screening dominated other screening strategies. The next most effective option was annual LBC screening, at an incremental cost of $147 000 per LYS vs LBC every 2 years, or $149 000 per QALY.					

Author/year Country	Objective of report Databases & dates searched	Type of analysis, number & date of included studies	Population considered in included studies Test comparison	Conclusions/recommendation	Quality assessment
Arbyn et al. 2008 Belgium	To compare test performance characteristics (accuracy) of conventional Pap tests and LBC cervical samples *Search* • PubMed (MEDLINE & PreMEDLINE) & EMBASE Jan 1991 – May 2007 • tables of contents of 5 gynaecology & 4 cytopathology journals • reference lists *Inclusion criteria* • Comparative studies (concomitant split-sample or direct-to-vial, or 2-cohort design) • All subjects receive reference standard verification by colposcopy ± biopsy for CIN 2+ if indicated • RCTs with ≥ 90% complete follow-up confirmation of positive women • No language restrictions	9 primary studies: • 7 with concomitant testing (4 split-sample) • 1 two-cohort study • 1 RCT • 6 ThinPrep, 1 AutoCyte, 1 DNA Citoliq, 1 CellSlide	Population • Most studies: women with previous cervical abnormalities • 2 studies: screening population Test comparison • Direct comparison of LBC vs conventional cytology, both with manual reading • 3 test thresholds: ASCUS+, LSIL+, HSIL+ Outcomes • Accuracy for CIN 2+: absolute & relative sensitivity & specificity	Authors' conclusion Liquid-based cervical cytology is neither more sensitive nor more specific for detection of high-grade CIN than the conventional Pap test.	Quality: high Explicit review questions: yes Explicit & appropriate eligibility criteria: yes Explicit & comprehensive search strategy: yes Quality of included studies appraised: yes Methods of study appraisal reproducible: yes Heterogeneity between studies assessed: yes Summary of main results clear and appropriate: yes Applicability: Most studies in women with previous abnormalities due to criteria for all subjects receiving high-quality reference standard. Subgroup analysis conducted.

Study quality

- Studies appropriately addressed 14 to 22 of 25 quality items in Standards for Reporting of Diagnostic Accuracy

Accuracy (CIN 2+)

Absolute sensitivity, specificity, DOR [7 studies]

HSIL+: LBC Se 57.1 (46.3–67.2), Sp 97.0 (93.8–98.6), DOR 42.9 (20.6–89.2); conventional Se 55.2 (45.5–64.7), Sp 96.7 (95.6–97.5), DOR 36.2 (21.2–61.9)

LSIL+: LBC Se 79.1 (70.1–86.0), Sp 78.8 (69.8–85.7), DOR 14.1 (7.7–25.7); conventional Se 75.6 (66.5–83.0), Sp 81.2 (71.9–88.0), DOR 13.4 (7.1–25.4)

ASCUS+: LBC Se 90.4 (82.5–95.0), Sp 64.6 (50.1–76.8), DOR 17.3 (8.9–33.5) ; conventional Se 88.2 (80.2–93.2), Sp 71.3 (58.3–81.6), DOR 18.5 (9.8–35.0)

1. Pooled sensitivity varied substantially by cytologic cutoff. The pooled sensitivities for conventional Pap were not significantly lower. The specificity of LBC dropped by decreasing cutoff.
2. Specificities of conventional Pap at cutoff HSIL+ and LSIL+ were in the same range as LBC.

[REVIEWER COMMENTS: These values were not considered accurate]

Author/year Country	Objective of report Databases & dates searched	Type of analysis, number & date of included studies	Population considered in included studies Test comparison	Conclusions/recommendation	Quality assessment

Relative sensitivity [9 studies] and specificity [8 studies], LBC : conventional

HSIL+: Se 1.05 (0.95–1.16), Sp 0.99 (0.98–1.01); Italian trial: Se 1.07 (95% CI 0.71–1.26)

LSIL+: Se 1.03 (0.96–1.11), Sp 0.97 (0.94–1.01); Italian trial: Se 1.03 (95% CI 0.74–1.43)

ASCUS+: Se 1.03 (0.97–1.09) [7 studies], Sp 0.91 (0.84–0.98) [6 studies]

- Overall, LBC was minimally, but not significantly, more sensitive in detecting underlying CIN 2+ than conventional Pap
- Clinical setting, split-sample vs direct-to-vial design and type of reference standard did not contribute to explaining inter-study variation in test accuracy
- Significant heterogeneity for relative specificity due to study design and LBC—this was due to one study: higher specificity ratio in the two-cohort study (1.01, 95% CI 1.00–1.02; Taylor et al. 2006) and lower specificity ratio with the DNA Citoliq system (0.97, 95% CI 0.95–0.99; Longatto-Filho et al. 2005).

Table 4. Subgroup Meta-Analyses: Variation of Relative Accuracy of Liquid Compared With Conventional Cytology* According to Study Characteristics

	Pooled Sensitivity Ratio (95% CI)	No. of Studies	P†	Pooled Specificity Ratio (95% CI)	No. of Studies	P†
Clinical setting			**.784**			**.792**
Screening‡	1.03 (0.76–1.40)	2	.821	1.00 (0.99–1.00)	1	–
Follow-up/high risk‡	1.05 (0.94–1.17)	8	.063	0.99 (0.97–1.01)	8	.001
Study design			**.742**			**.001**
Concomitant testing	1.07 (0.95–1.20)	7	.053	0.99 (0.97–1.00)	7	.052
Two-cohort/RCT	0.94 (0.71–1.26)	2	.421	1.01 (1.00–1.02)	1	–
Split sample/direct-to-vial			**.999**			**.999**
Split-sample	1.06 (0.92–1.22)	4	.065	1.00 (0.98–1.01)	3	.190
Direct-to-vial	1.05 (0.87–1.27)	5	.195	0.98 (0.96–1.01)	5	<.001
Gold standard			**.876**			**.999**
Colposcopy and histology if indicated	1.07 (0.60–1.64)	1	–	–	0	–
Complete colposcopy, histology if indicated	1.10 (0.94–1.28)	6	.027	0.99 (0.98–1.01)	6	.001
Complete histology	0.96 (0.84–1.11)	2	.601	0.98 (0.89–1.08)	2	.021
LBC system			**.405**			**.001**
ThinPrep	1.07 (0.92–1.23)	6	.047	1.00 (0.99–1.01)	5	.053
AutoCyte	0.95 (0.81–1.11)	1	–	0.94 (0.87–1.01)	1	–
CellSlide	1.27 (0.75–2.15)	1	–	1.00 (0.95–1.05)	1	–
DNA-Citoliq	1.14 (0.85–0.51)	1	–	0.97 (0.95–0.99)	1	–

CI, confidence interval; RCT, randomized controlled trial; LBC, liquid-based cytology.
* At cutoff high-grade squamous intraepithelial lesion or worse for cervical intraepithelial neoplasia 2 or worse.
† *P* corresponding with test for interstudy heterogeneity; bold type indicates between-group heterogeneity.
‡ Study of Coste et al, 2003, contributed a screening group and a follow-up group.

Summary receiver operating characteristic (SROC) regression

- SROC regression, using ThinPrep as reference, identified a lower DOR for AutoCyte at cut-off HSIL+
- The contrast in DOR between CC and LBC was not influenced by the number of quality issues of Standards for Reporting of Diagnostic Accuracy addressed in the individual reports (HSIL+ *P* = 0.84, LSIL+ *P* = 0.37, ASCUS+ *P* = 0.65)

Previous meta-analysis

The authors report that a previous meta-analysis (2003), which included studies with incomplete verification (confirmation of ≥ 80% positive, ≥ 5% normal results), with boot-strapping found no significant difference in sensitivity or specificity (data not shown).

Author & year Setting, *N*	Study objective & design	Study population	Study quality and applicability
Primary studies: Auto LBC vs conventional cytology			
Davey et al. 2007a Australia One pathology laboratory Aug 2004 – June 2005 *N* = 55 164 consecutive samples	*Objective:* To compare the accuracy of LBC using computerised TPI with manually read conventional cytology *Study design* Diagnostic accuracy study with split-sample pairs *Index test* ThinPrep LBC with computerised TPI reading *Comparator test* CC with manual reading *Reference standard* Histology for discordant cytology to 6 months, using (1) register results; (2) blinded (to cytology and register results) re-reading of histology slides; and (3) more severe result from (1) or (2) No reference standard for 254/844 discordant samples *Test threshold* Test: CIN 1+ Reference standard: CIN 2+ *Primary outcomes* Accuracy for squamous lesions *Secondary outcomes* Unsatisfactory rate, distribution of squamous cytology, accuracy for glandular lesions *Relevant subgroups* Nil *Cytological classification* AMBS 1994	*Inclusion criteria* • Any age • Electing LBC sample • Screening and diagnostic population *Exclusion criteria* • Nil *Patient characteristics* NR	Quality: fair Q2 Accuracy data level III-2 Prospective: yes Consecutive: yes Explicit selection criteria: yes Test threshold specified: yes Reference standard • valid • not applied to all participants; applied to most discordant participants with bias testing Test interval adequate for reference standard and comparator (6 months) Tests reported blinded to ref standard & comparator Ref standard reported blinded to tests Uninterpretable/intermediate results reported: yes Study withdrawals explained: yes Sufficient data for relative TP & FP: yes Applicability: applicable P1 Applicable population • mixed screening & diagnostic population Applicable comparator: yes Applicable intervention: yes

RESULTS:

Test accuracy

On-line appendix: Logistic modelling showed no evidence that the proportions of discordant slides (with at least one result of CIN 1 or more severe) that were verified were associated with whichever test (TPI or CC) gave the more severe result (χ^2 = 0.34, 1 df, *P* = 0.56). A test for interaction showed no evidence that the relative odds of verification varied across categories of discordance (χ^2 = 3.7, 10 df, *P* = 0.96).

There was no evidence that knowledge of which test gave the abnormal result was associated with the odds of verification (χ^2 = 0.50, 1 df, *P* = 0.48).

Accuracy for CIN 2+ lesions

Incremental TP, register ref std

Test threshold CIN 1+:

TPI 71 (133–62) more TPs, 170 more biopsies (380–210; 3.08 per 1000) than conventional (99 additional biopsies negative for high-grade lesion) (χ^2 *P* = 0.0000). TPI detected 1.29 additional cases per 1000 women screened. TP:FP for discordant positives: TPI 1:1.85 (133:247), conventional 1:2.37 (62:148); for additional positives 71:99 (1:1.39)

Test threshold CIN 1+, overall

TPI 71–106 more TPs (1.29–1.92 per 1000 women), 126–170 more biopsies (2.28–3.08 per 1000) than conventional (41–170 additional biopsies negative for high-grade lesion)

Test threshold inconclusive high grade+:

TPI 45 (141 – 96) more TPs (0.82 [0.27–1.36] per 1000 women), Pearson's χ^2 *P* = 0.0034), 41 more biopsies (262 – 221; 0.74 [–0.02 to +1.51] per 1000; Pearson's χ^2 *P* = 0.0615), 4 fewer biopsies negative for high-grade lesion (–0.07 [–0.63 to +0.48] per 1000). FPs 4 fewer (TPI 121, conv 125) Pearson's χ^2 *P* = 0.798. TP:FP for discordant positives: TPI 1:0.86 (141:121), conventional 1:1.30 (96:125); additional TP:FP 45:–4 (1:–0.09)

Test threshold atypia with HPV+ (LSIL)

Biopsies in 15% (99/656) of discordant findings. TPI 77 (132 – 55) more TPs (1.40 [–0.63 to +0.48] per 1000 women, Pearson's χ^2 *P* < 0.0001), 208 more biopsies (429 – 221; 3.77 [2.90–4.65] per 1000), 131 additional biopsies negative for high-grade lesion. FPs = TPI additional 131 (297 – 166), 2.37 [1.63–3.13] per 1000. TP:FP for discordant positives: TPI 1:2.25 (132:297), conventional 1:3.02 (55:166); additional TP:FP 77:131 (1:1.70)

Test threshold CIN 1+, blinded re-reading as ref std

TPI 85 more TPs (153 – 68), 126 more biopsies (280 – 154) than conventional (41 additional biopsies negative for high-grade lesion) (χ^2 *P* = 0.0000), additional FPs (127 – 86)

TP:FP for additional positives 85:41 (1:0.48)

Test threshold CIN 1+, most severe histology result as ref std

TPI 106 more TPs (196 – 90), 170 more biopsies (380 – 210) than conventional (64 additional biopsies negative for high-grade lesion [184 – 120]). Additional TP:FP 106:64 (1:0.60)

Glandular lesions to detect AIS+

41.3% (26/63) total +ve on either test verified

Test threshold atypia

TPI: TP = 6, FN = 0, FP = 8, TN = 12; conv: TP = 6, FN = 0, FP = 16, TN = 4

Test threshold inconclusive, high grade to be excluded

TPI: TP = 5, FN = 1, FP = 3, TN = 17; conv: TP = 6, FN = 0, FP = 9, TN = 11

Test characteristics Unsatisfactory rate TPI 1.8%; conventional 3.1% (χ^2 $P < 0.001$) Yield Total pLSIL+: TPI 7.4%, conventional 6.0% *pLSIL (atypia):* TPI 3.06%, conventional 2.84% (χ^2 $P = 0.036$) *LSIL (atypia with HPV + CIN 1):* TPI 3.07% (1617), conv 1.97% (1038) *pLSIL/LSIL (atypia, atypia with HPV or CIN 1):* TPI 6.1% (3226), conventional 4.8% (2532) *pHSIL:* TPI 0.42%, TPM 0.54% *CIN 2:* TPI 0.40%, TPM 0.28% *Inconclusive, high-grade histology to be excluded or CIN 2:* TPI 0.83% (435), conventional 0.82% (430) *CIN 3+:* TPI 0.49% (258), conventional 0.43% (224) Glandular TPI (23) 0.04%, conventional (52) 0.09%			
Davey et al. 2007b Diagnostic cytopathology Australia One pathology laboratory Aug 2004 – Feb 2005 *N* = 1645 reading sessions; 52 385 CC slides, 24 645 TPI slides	*Objective:* To compare slide reading times of LBC using computerised TPI with manually read CC *Study design* Retrospective, cross-sectional study. Sub-study of diagnostic accuracy study *Index test* ThinPrep LBC with computerised TPI reading (Aug 2004 – Feb 2005) *Comparator test* CC with manual reading (Aug 2004 – Dec 2004) *Reference standard* N/A *Test threshold* N/A *Primary outcomes* Within-reader differences in slide reading times (slides/h, ave min per slide). Time spent not screening deducted from time recorded *Secondary outcomes* Effect of cytologist experience on reading times *Relevant subgroups* Those trained vs untrained to read TPI	*Inclusion criteria* • Routine (screening and diagnostic) population *Exclusion criteria* • CC slide-reading sessions: any sessions including non-CC reading • Session values > 3 SDs from mean for particular reader • 38 (2.3%) sessions excluded (34 CC, 4 TPI) *Cytologist characteristics* • 41 readers • Paired data for 20 readers • Same readers & slides as accuracy study (pers. correspondence)	Quality Intervention level III-2 Sub-study in diagnostic accuracy study Prospective: no Consecutive: no (excluded some whole sessions) Explicit selection criteria: yes Test threshold specified: N/A Reference standard N/A Study withdrawals explained: yes Sufficient data for contingency table: N/A Applicability Applicable population • mixed screening & diagnostic population Applicable comparator: yes Applicable intervention: yes

RESULTS:

Test characteristics

Process outcomes

Paired data (20 readers)

581 CC sessions, 379 TPI sessions

Mean min/slide (95% CI): TPI 4.71 (4.38–5.04), CC 10.61 (9.73–11.49); mean difference 5.90 (5.04–6.75; median 5.5, range 3.3–9.8), $P < 0.001$, paired *t*-test

Mean slides/h (95% CI): TPI 13.31 (12.25–14.36), CC 6.12 (5.76–6.49), mean within-reader difference 7.18 (6.17–8.20; median 7.0, range 4.2–13.5); $P < 0.001$, paired *t*-test

Subgroups

Training: No difference in CC mean reading times for those trained ($n = 20$, 10.61 min, 9.73–11.49; 6.1 slides/h) vs not trained ($n = 21$, 685 sessions, 10.61 min, 10.03–11.18; 5.9 slides/h) to read TPI ($P = 1.00$ reading time, $P = 0.38$ reading rate; independent *t*-test; similar result with random effects analyses NR)

Speed of reading: Slower readers had greater increases in speed (slide/min and min/slide) with TPI than faster readers ($P < 0.001$, Altman–Bland plots & linear regression)

Experience: More experienced readers read CC slides more quickly, but experience did not affect TPI reading times ($P = 0.41$ mins/slide, $P = 0.48$ slides/min), CC-reading times ($P = 0.30$, $P = 0.58$ TPI trained readers; $P = 0.12$, $P = 0.11$ TPI untrained readers) or within-reader differences ($P = 0.16$ mins/slide, $P = 0.34$ slides/min linear regression). Weak evidence for effect of experience on reading times of CC over all 41 readers ($P = 0.08$ mins/slide, $P = 0.09$ slides/min)

Roberts et al. 2007 Australia One pathology laboratory Feb 2005 – April 2005 *N* = 11 416 samples + 103 seeded LBC cases	*Objective:* To compare the accuracy of LBC using the computerised ThinPrep Imager (TPI) with manually read ThinPrep slides (TPM) and also manually read conventional slides (secondary aim). *Study design* Three-arm diagnostic accuracy study with split-sample pairs and seeded LBC cases. Data reported separately for routine and seeded cases. *Index test* ThinPrep LBC with computerised ThinPrep Imager reading *Comparator test* 1) Manual reading of ThinPrep LBC 2) conventional cytology with manual reading *Reference standard* Histopathology within 9 months for high-grade and possible high-grade lesions, no reference standard for 12/134 TPI (9.0%), 15/180TPM (8.3%), 10/93 conventional (10.8%) (χ^2-test $P = 0.834$). None for low-grade positive, negatives. *Test threshold* Test: HSIL/pHSIL (including possible glandular high grade) Reference standard: high grade (defined as HSIL, AIS or carcinoma) *Primary outcomes* Accuracy for high-grade lesions	*Inclusion criteria* • Routinely received samples (*N* = 11 416) • HSIL, AIS or carcinoma (*N* = 103) • Likely screening and diagnostic population—unclear *Exclusion criteria* • Slides not read by imager (3.7%) due to poor cellularity, excessive blood, technical problems *Patient characteristics* NR	Quality: fair Q2 Prospective: yes Consecutive: no Explicit selection criteria: no Test threshold specified: yes Reference standard • valid • not applied to all participants; applied to majority positive participants Test interval adequate • reference standard—yes (9 months) • comparator - yes Tests reported blinded to ref standard & comparator Ref standard reported blinded to tests - NR (probably no) Uninterpretable/intermediate results reported: total NR Study withdrawals explained: yes TP slides rejected by imager were excluded from analysis (3.7%) Sufficient data for relative TP&FP: yes McNemars' test

	Secondary outcomes Productivity: average screening time per slide *Relevant subgroups* Nil *Cytological classification* Australian Modified Bethesda System 2004		Applicability: limited P2 Applicable population - unclear Applicable comparator: yes Applicable intervention: yes

RESULTS:

Test accuracy (for detection of CIN 2+/CIN 3+)

Accuracy for CIN 2+ lesions

Reported sensitivities invalid as not based on full reference standard for negative findings (possible FN). Total positive cytology results unclear

TPs, including seeded cases (N = 11 519):

TPI = 172, TPM = 184. TPM identified additional 12 TPs over TPI; additional 1.0 cases per 1000 women screened

Test threshold pHSIL/HSIL (including glandular), excluding seeded cases (N = 11 416):

TPs: TPI = 80, TPM = 95, conventional = 63 (including possible high-grade glandular lesions) (reviewer χ^2 TPI 0.15, TPM 0.01). TPI identified additional 17 TPs over conventional (1.49 cases per 1000). TPM identified additional 32 TPs over conventional, additional 15 TPs over TPI (2.8 cases per 1000). $\chi^2 < 0.05$ for (i) comparison of automated LBC vs manual LBC vs manual Pap test; and (ii) comparison of manual LBC vs CC. χ^2 *P* = 0.255 for automated LBC vs manual LBC

Additional biopsies (Tables VII & VIII): TPI vs conv 39 (122–83), 3.42 per 1000.

FPs, bx confirmed: TPI = 20 HSIL + 22 pHSIL = 42, TPM = 24 + 46 = 70, (χ^2 *P* = 0.0079) conventional = 8 + 12 = 20 (χ^2 *P* = 0.0051). TPI additional 22 over conventional (additional 1.93 per 1000), TPM additional 55 over conventional (4.8 per 1000).

Total TP:FP TPI 1:0.53, TPM 1:0.74, conventional 1:0.32. TPI vs conventional χ^2 *P* = 0.114; TPM vs conv χ^2 *P* = 0.005.

Additional TP:FP TPI vs conv 17:22 (1:1.29); TPI vs TPM 15:28 (1:1.87)

Test threshold pHSIL/HSIL (excluding glandular), excluding seeded cases (N = 11 416):

TPs: TPI = 78, TPM = 91, conventional = 60 (excluding possible high-grade glandular lesions) (reviewer χ^2 TPI 0.12, TPM 0.011). TPI additional 18 cases over conventional (1.58 per 1000). χ^2 *P* = 0.3155 for automated LBC vs manual LBC

Test threshold pLSIL/LSIL (including glandular), excluding seeded cases (N = 11 416):

TPs: TPI = 99, TPM = 106, conventional = 81 (including possible high-grade glandular lesions) (reviewer χ^2 TPI 0.18, TPM). TPI identified additional 18 TPs over conventional (1.58 cases per 1000). TPM identified additional 25 TPs over conventional, additional 7 TPs over TPI (0.6 cases per 1000).

Test threshold pLSIL/LSIL (excluding glandular), excluding seeded cases (N = 11 416):

TPs: TPI = 97, TPM = 102, conventional = 78 (including possible high-grade glandular lesions) (reviewer χ^2 TPI, TPM). TPI identified additional 19 TPs over conventional (1.66 cases per 1000). TPM identified additional 24 TPs over conventional, additional 5 TPs over TPI (0.4 cases per 1000)

Test characteristics Yield Total pLSIL+: TPI 7.9%, conventional 4.2% *pLSIL:* TPI 3.8%, TPM 4.5%, conventional 1.8% *LSIL:* TPI 2.9%, TPM 2.7%, conventional 1.6% *pLSIL/LSIL:* TPI 6.7%, TPM 7.2%, conventional 3.3% *Possible high grade:* TPI 0.43% (49), TPM 0.72% (82), conventional 0.33% (38) *High grade:* TPI 0.74% (85), TPM 0.86% (98) (χ^2 *P* = 0.33), conventional 0.48% (55) *pHSIL/HSIL:* TPI 1.17% (134), TPM 1.58% (180), conventional 0.81% (93) Screening time (mean min per slide): TPI = 3.42, TPM = 4.71, conventional = 7.40 TPI: 84% slides did not require full manual screening			
Confortini et al. 2003 Italy April 2000 – NR *N* = 14 145	*Objective:* To compare the performance of conventional and AutoPap-assisted reading *Study design* Prospective, paired diagnostic accuracy study, with cost comparison. *Index test* AutoPap 300 QC–assisted reading of conventional smears (read by same cytopathologist who did conventional read) *Comparator test* Manual reading of conventional smears *Reference standard* Histology (biopsy or loop excision) ASCUS favouring reactive process: repeat smear @ 6 mo then colposcopy if necessary ASCUS favouring SIL +: colposcopy HSIL, colposcopy neg: repeat cytology 6 mo Proportion reference standard available NR *Test threshold* Reference standard CIN 2+ *Primary outcomes* • rate abnormal smears • proportion high-grade CIN detected by each test *Secondary outcomes* • costs	*Inclusion criteria* • Smears from Florence screening program	Quality: fair Q2 Prospective: yes Consecutive: yes Explicit selection criteria: no Test threshold specified: yes Reference standard • valid • unclear if results available for all positive patients Test interval adequate • reference standard • comparator—yes (3–4 days) Index test reported blinded to comparator: yes (although same) Tests reported blinded to ref standard: yes Ref standard reported blinded to tests: NR Uninterpretable/intermediate results reported: yes Study withdrawals explained: yes Sufficient data for relative TP & FP: no (FP not reported)

	• reading time cost savings *Cytological classification* • 1991 Bethesda		Applicability: applicable P1 Applicable population yes Applicable comparator: yes Applicable intervention: yes

RESULTS:

Test accuracy (for detection of CIN 2+/CIN 3+)

Test threshold HSIL

Incremental TP rate

- 3 additional TP for conventional (conventional 16, AutoPap 13) (reviewer χ^2 0.58)
- AutoPap: –0.21 cases per 1000 women screened

Test threshold ASCUS-R

Incremental TP rate

- 1 additional TP for conventional (conventional 31, AutoPap 30) (reviewer χ^2 0.90)
- AutoPap: –0.07 cases per 1000 women screened

ASCUS-S referral for colposcopy

Referral for colposcopy AutoPap 117 (147 – 30), conv 148 (179 – 31)—but unclear what proportion underwent colposcopy: invalid as FP; just yield

Test characteristics

Yield

ASCUS-R: TPI 1.28%, conv 2.33%

ASCUS-S/AGUS: TPI 0.54%, conv 0.65%

ASCUS-R/ASCUS-S/AGUS: TPI 1.82% (258), conv 2.99% (423)

LSIL: TPI 0.20%, conv 0.41%

HSIL: TPI 0.13%, conv 0.19%

Carcinoma: TPI 0.007%, conv 0.007%

HSIL+: conventional 0.20% (28), AutoPap 0.16% (23)

ASCUS-R+: conventional 3.6% (509), AutoPap 2.2% (307)

Process outcomes

- *Unsatisfactory*: conventional: 1.3% (188); AutoPap 0.88% (125) ($P < 0.001$, χ^2)
- AutoPap 1818 process review (failed processing > manual readings) 12.8%
- 2398 NFR (16.9%)
- ASCUS-S+ refer to colposcopy, AutoPap 147, conv 179 (χ^2 $P = 0.07$, 1 df)

Resource use

• Slides read: conventional 100%, AutoPap 83.1% (including PR) • 6-month repeat cytology: conventional 2.3%, AutoPap 1.6% ($P < 0.01$, χ^2) • Colposcopy conventional 1.2%, AutoPap 1.0% ($P = 0.07$) • Detection CIN 2+ conventional 0.22%, AutoPap 0.21% Cost-effectiveness Workload 25 000 slides per year: • Overall cost: conventional €259 833, AutoPap €392 959 • Cost per CIN 2+ detected: conventional €4724, AutoPap €7414 Workload 60 000 slides per year (based on best conditions in last part of study, 20% NFR, 10% PR): • Overall cost: conventional €623 600, AutoPap €639 825 • Cost per CIN 2+ detected: conventional €4724, AutoPap €5078 • Differences between conventional and AutoPap decreased as NFR rate increased, with similar costs at 60 000 slides per year and 30% NFR			
Stevens et al. 2004 Australia Jan 2000 – Feb 2000 *N* = 5583	*Objective:* To determine if AutoPap-GS is equivalent to manual reading of conventional Pap for detecting abnormal, unsatisfactory, endocervical component or organisms. Also assessed effectiveness of slide ranking for triaging abnormal smears *Study design* Two-arm retrospective diagnostic accuracy study with split-sample pairs and seeded LBC cases *Index test* AutoPap location-guided screening–assisted practice (AutoPap-GS) (archiving NFR + PapMap FOV locations traced onto review slides + manual review of 10% smears selected as QC) *Comparator test* Manual screening of all conventional smears, rescreening of high-risk patients (~20%) *Reference standard* Concordant diagnosis between the current practice and AutoPap-GS diagnosis. Discordant diagnoses were adjudicated by discrepancy panels of 3 experienced cytology professionals; truth was determined by majority agreement *Test threshold* Low-grade epithelial abnormality (LGEA+) *Unsatisfactory smear* Scant epithelial component, or obscuring blood, inflammation or drying artefact	*Inclusion criteria* Consecutive (6000) *Exclusion criteria* • Broken, cracked, 2 cover-slips, vaginal smears, LBC (422, 7.0%) • AutoPap not for review (986, 16.4%)	Quality: fair Q2 Prospective: no Consecutive: no (NFR excluded) Explicit selection criteria: no Test threshold specified: yes Reference standard • Valid (suboptimal) • Applied to all discordant participants Test interval adequate • reference standard—yes • comparator—yes Tests reported blinded to ref standard—yes Ref standard reported blinded to tests—NR Uninterpretable/intermediate results reported: yes Study withdrawals explained: yes Sufficient data for relative TP & FP: no Applicability: limited P2 Applicable population: unclear Applicable comparator: yes

	precluded interpretation of ≥ 75% epithelial cells *Cytological classification* Australian NHMRC endorsed terminology 1994		Intervention limited (NFR slides)
RESULTS: Test accuracy—data not clearly reported Discordant high-grade epithelial abnormality (HGEA) vs normal (excludes 18 HGEA on AutoPap, low-grade epithelial abnormality [LGEA] on conventional): AutoPap 4 HGEA normal on conventional; conventional 5 HGEA normal on AutoPap Test characteristics <u>Yield</u> *LGEA+:* AutoPap 4.3% (193/4493), conv 5.5% (250/4493) Process outcomes • 986 NFR			
Primary studies: Automated LBC vs manual LBC			
Biscotti et al. 2005 US 4 clinical sites Dec 2000 – July 2001 *N* = 10 742	*Objective:* To demonstrate that TPI is at least equivalent in [accuracy and] adequacy to manual screening of LBC for categories of Bethesda system *Study design* Prospective, masked, paired, diagnostic accuracy study with seeded cases *Index test* TPI reading of ThinPrep LBC slides (≥ 48 days following manual screening). Slides randomised; same cytologist and pathologist performing manual & automated reading *Comparator test* Manual reading of ThinPrep LBC slides (2000 or 3000 processor). Marks removed *Reference standard* Accuracy: Cytological adjudication (3 independent pathologists) for concordant ASCUS+, discordant ≥ 1 level & 5% concordant negative Adequacy: Independent cytologist for all concordant unsatisfactory, discordant adequacy, random 5% of concordant satisfactory or 'satisfactory but limited by' *Test threshold* Test: HPV or reactive cell changes, epithelial abnormalities or malignant cells. Reference standard: ASCUS+ *Primary outcomes* Sensitivity & specificity	*Inclusion criteria* o Routine screening & diagnostic population (10 359) o Seeding with HSIL cases (NR, likely 383, 3.6%) *Exclusion criteria* o Cases that could not be processed by TPI, *n* = 732/10359 slides, 7.1%) (eg, excessive air bubbles) o No consensus diagnosis on adjudicated review (*n* = 6/361 concordant slides) *Patient characteristics* NR o 1.3% (138/10 742) reference standard HSIL+ o 636 ASCUS+ cases	Quality: fair Q2 Prospective: yes Consecutive: no Explicit selection criteria: no Test threshold specified: yes Reference standard • valid (suboptimal) • applied to all participants—yes Test interval adequate • reference standard—yes • comparator—yes Tests reported blinded to ref standard—yes Ref standard reported blinded to tests—NR Uninterpretable/intermediate results reported: yes Study withdrawals explained: no (unclear) TP slides not processed by imager were

	Secondary outcomes Productivity: average slides screened per day and time expended Specimen adequacy *Relevant subgroups* Nil *Cytological diagnosis system* o 1991 Bethesda system		excluded from analysis (7.1%) Sufficient data for relative TP & FP: yes Applicability: limited P2 Applicable population: limited (seeded cases) Applicable comparator: yes Applicable intervention: yes

RESULTS:

Test accuracy

Accuracy for HSIL+

Sensitivity, specificity (95% CI), test threshold HSIL+:

TPM: Se 74.1% (66.0%–81.2%), Sp 99.4% (99.2%–99.6%)

TPI: Se 79.9% (72.2%–86.2%), Sp 99.6% (99.5%–99.7%)

Difference: TPI Se +5.8% (–1.1% to +12.6%), Sp +0.2% (0.06%–0.4%)

Test threshold HSIL, excluding AGUS cases:

TPI = 111 TPs (includes 1 SCC case), TPM = 103 TPs. TPI identified additional 8 TPs over TPM; additional 0.74 cases per 1000 women screened. χ^2 *P* = 0.581 (using 10 742 as denominator, but could be 10 359)

TPI = 32 FPs, TPM = 50 FPs. TPI identified 18 fewer FPs over TPM; 1.68 fewer cases per 1000 women screened

Additional TP:FP 1:–2.25.

Test threshold HSIL, including AGUS cases:

TPI = 111 TPs, TPM = 103 TPs. TPI additional 8 TPs. TPI = 33 FPs, TPM = 52 FPs. TPI identified 19 fewer FPs over TPM; 1.77 fewer cases per 1000 women screened

Additional TP:FP 1:–2.38. Estimated difference in sensitivity = 8/139 = 5.8%. Estimated difference in FPR = –19/9401 = 0.20%

Test threshold LSIL, *excluding AGUS cases:*

TPI = 128 TPs, TPM = 124 TPs. TPI identified additional 4 TPs; additional 0.37 cases per 1000 women screened. TPI = 253 FPs, TPM = 265 FPs. TPI identified 12 fewer FPs over TPM; 1.12 fewer cases per 1000 women screened. χ^2 *P* = 0.59. Additional TP:FP 1:–3. Estimated difference in sensitivity = 4/139 = 2.9% (1.1–7.2). Estimated difference in FPR = –12/9401 = –0.13% (–0.07 to 0.23) (specificity = +0.13%)

Test threshold LSIL, *including AGUS cases:*

TPI = 128 TPs, TPM = 124 TPs. TPI identified additional 4 TPs; additional 0.37 cases per 1000 women screened. TPI = 254 FPs, TPM = 267 FPs. TPI identified 13 fewer FPs over TPM; 1.21 fewer cases per 1000 women screened. Additional TP:FP 1:–3.25.

Estimated difference in sensitivity = 4/139 = 2.9%. Estimated difference in FPR = –13/9401 = –0.14%

Test threshold ASCUS, excluding AGUS

TPI = 137 TPs, TPM = 132 TPs. TPI identified additional 5 TPs over TPM; additional 0.47 cases per 1000 women screened. TPI = 618 FPs, TPM = 591 FPs. TPI identified additional 27 FPs over TPM; additional 2.51 per 1000 women screened. χ^2 *P* = 0.42

Additional TP:FP 1:5.4. Estimated difference in sensitivity = 5/139 = 3.6% (1.6–8.2). Estimated difference in FPR = 27/9401 = 0.29% (0.20–0.42)

Test threshold ASCUS, including AGUS

TPI = 137 TPs, TPM = 132 TPs. TPI identified additional 5 TPs over TPM; additional 0.47 cases per 1000 women screened. TPI = 624 FPs, TPM = 598 FPs. TPI identified additional 26 FPs over TPM; additional 2.42 per 1000 women screened.

Additional TP:FP 1:5.2. Estimated difference in sensitivity = 5/139 = 3.6%. Estimated difference in FPR = 26/9401 = 0.28%

Reference standard AGUS

10 cases (ref std)

TPI = 4 neg, 1 ASCUS, 4 AGUS, 1 SCC. TPM = 2 neg, 2 ASCUS, 3 AGUS, 1 HSIL, 1 SCC. TPI = 5 TPs, TPM = 5 TPs. TPI identified 0 additional TPs

FNs:

Of 428 concordant neg undergoing ref std, 3 (0.7%) were adjudicated as ASCUS

Discordant FNs: 68 TPI, 112 TPM

Accuracy for LSIL+

Sensitivity, specificity (95% CI), test threshold LSIL+:

TPM: Se 79.7% (75.3%–83.7%), Sp 99.0% (98.8%–99.2%). TPI: Se 79.2% (74.7%–83.2%), Sp 99.1% (98.9%–99.3%). Difference: TPI Se –0.5% (–5.0% to +4.0%), Sp +0.09% (–0.1% to +0.3%)

Accuracy for LSIL+

Sensitivity, specificity (95% CI), test threshold ASCUS+:

TPM: Se 75.6% (72.2%–78.8%), Sp 97.6% (97.2%–97.9%). TPI: Se 82.0% (78.8%–84.8%), Sp 97.8% (97.4%–98.1%). Difference: TPI Se 6.4% (2.6%–10.0%), Sp +0.09% (–0.2% to +0.6%)

Test characteristics

Yield, excluding 10 AGUS cases

SCC TPI 3, TPM 5; *HSIL* TPI 139, TPM 147; *LSIL* TPI 238, TPM 236; *AGUS* TPI 6, TPM 9; *ASCUS* TPI 374, TPM 334

Yield (% x/9627), including AGUS cases

SCC TPI 4 (0.04%), TPM 6 (0.06%); *HSIL* TPI 139 (1.44), TPM 148 (1.54); *LSIL* TPI 238 (2.47), TPM 236 (2.45); *AGUS* TPI 10 (0.10), TPM 12 (0.12); *ASCUS* TPI 375 (3.90), TPM 336 (3.49)

Process outcomes

Not processed: TPI = 732/10 742 (6.8%)

Unsatisfactory: TPI = 29/9627 (0.30%), TPM = 66/9627 (0.69%); TPI = 29/10 742 (0.27%), TPM 66/10 742 (0.61%)

Satisfactory but limited by: TPI = 2252/9627 (23.4%), TPM = 2186/9627 (22.7%)

Either of above: TPI = 23.7%, TPM = 23.4%

Screening times

Ave slides per 8-h day: TPI = 151, TPM = 72. *Ave screening time (mins):* TPI = 3.2, TPM = 6.7

Bolger et al. 2006 Ireland 1 clinical site April 2004 – March 2005, second half of cases *N* = 6000	*Objective:* To compare TPI with manual screening of TP slides. Second study using TPI as in standard practice included in review *Study design* Prospective, paired, diagnostic accuracy study *Index test* TPI reading of ThinPrep LBC slides (± manual screening following FOV results) *Comparator test* Manual reading of ThinPrep LBC slides, by different cytologist. Slide markings removed. *Reference standard* Positive and discordant cases: adjudicated cytology (one senior cytologist + one cytopathologist) Negative cases: nil *Test threshold* Test: NR Reference standard: NR *Primary outcomes* Sensitivity & specificity *Secondary outcomes* Nil *Relevant subgroups* Nil *Cytological diagnosis system* British Society for Clinical Cytology (BSCC) guidelines	*Inclusion criteria* • Routine samples • Likely screening and diagnostic population—unclear *Exclusion criteria* • Cases not scanned by TPI (approx 3%) *Patient characteristics* NR • 1.4% (83/6000) reference standard CIN 2+ Data from 1st set 6000 cases not used as they do not represent intended clinical use of TPI	Quality: fair Q2 Prospective: yes Consecutive: yes Explicit selection criteria: no Test threshold specified: no (sens & spec) Reference standard • valid (suboptimal) • applied to all discordant & positive participants Test interval adequate • reference standard—NR • comparator—yes Tests reported blinded to ref standard—yes Ref standard reported blinded to tests—NR Uninterpretable/intermediate results reported: yes Study withdrawals explained: no (no. TP slides not processed by imager NR) Sufficient data for relative TP & FP: yes Applicability: unclear P2 Applicable population: unclear Applicable comparator: yes Applicable intervention: yes
Test accuracy *Sensitivity, specificity* Test & ref std thresholds not reported. Basis of calculation not reported; likely concordant negative results assumed to be TN. Cannot confirm calculations: reliability of reported values unclear TPI: Se 92.84 (95% CI 90.82–94.45), Sp 98.62 (95% CI 98.26–98.91) Manual LBC: Se 94.42 (95% CI 92.57–95.84), Sp 98.77 (95% CI 98.42–99.04) No significant statistical difference in sensitivity or specificity <u>Accuracy for CIN 2+ (HSIL)</u> (excl glandular—but same differences as including glandular as 0%; see below): *Test threshold CIN 2:* TPI = 69 TPs, manual = 74 TPs. TPI identified 5 fewer TP, 0.83 fewer per 1000 women screened. χ^2 *P* = 0.674. TPI = 5 FPs, manual = 7 FPs. TPI identified 2 fewer FPs, 0.33 fewer per 1000 women screened. Additional TP:FP –1:–0.4. Estimated difference in sensitivity = –5/83 = –6.0% (–2.6 to –13.3). Estimated difference in FPR = –2/5845 = –0.03% (0.01–0.12) (specificity = +0.03%)			

Test threshold CIN 1: TPI = 80 TPs, manual = 81 TPs. TPI identified 1 fewer TP, 0.17 fewer per 1000 women screened. TPI = 370 FPs, manual = 375 FPs. TPI identified 5 fewer FPs, 0.83 fewer per 1000 women screened. χ^2 *P* = 0.85 Additional TP:FP –1:–5. Estimated difference in sensitivity = –1/83 = –1.2% (–0.21 to –6.51). Estimated difference in FPR = –5/5845 = –0.09% (–0.04 to –0.21) *Test threshold BNA (borderline nuclear abnormality):* TPI = 82 TPs, manual = 82 TPs. TPI identified 0 additional TPs, 0 fewer per 1000 women screened. TPI = 706 FPs, manual = 711 FPs. TPI identified 5 fewer FPs, 0.83 fewer per 1000 women screened. χ^2 *P* = 0.89 Additional TP:FP N/A. Estimated difference in sensitivity = 0/83 = 0% (0–0.04). Estimated difference in FPR = –5/5845 = –0.09% (–0.04 to –0.21) Accuracy for glandular: TPI = 2 TPs, 0 FPs; TPM = 2 TPs, 0 FPs. 0% difference in sensitivity or specificity. Biopsy confirmation: Of 83 CIN 2+ on adjudicated cytology, histology available for 51: 43 confirmed high-grade, 6 low-grade, 2 negative (8 FN). Not reported contingent on screening method results, ie, unclear if these cases were concordant on TPI and TPM (adjudicated cytology positive predictive value of 84% for biopsy confirmed high-grade disease) Of 2 glandular cases, cytology indicated 1 low-grade cervical glandular intraepithelial neoplasia (cGIN) (no biopsy available), 1 case endometrial (confirmed as endometrial adenocarcinoma on biopsy). Test characteristics Yield (denominator 6000) *BNA:* TPI 5.7%, TPM 5.6%; *CIN 1:* TPI 6.3%, TPM 6.3%; *CIN 2:* TPI 0.92%, TPM 0.95%; *CIN 3:* TPI 0.32%, TPM 0.40% ; HSIL (CIN 2 + CIN 3): TPI 1.23%, TPM 1.35% *Glandular:* TPI 0.03%, TPM 0.03% Process outcomes *Not processed:* NR (approx 3%) *Unsatisfactory:* TPI = 63/6000 (1.05%), TPM = 64/6000 (1.07%)		
Wilbur et al. 2002 US 1 cytopathology laboratory Recruitment period NR *N* = 1275	*Study design* Prospective, paired-sample, masked, diagnostic accuracy study with retrospectively obtained samples *Index test* AutoPap reading of AutoCyte PREP slides (SlideWizard 2), including designation of some slides as NFR. Review slides indicate 10 FOVs *Comparator test* Manual reading of AutoCyte PREP slides *Reference standard* Reviewers aware of slide markings, but not study arm providing markings Discordant and positive slides: cytological adjudication (1 cytopathologist)	*Inclusion criteria* • Retrospective sample of slides • 1049 consecutive routine samples + 251 seeded cases • Likely screening and diagnostic population—unclear *Exclusion criteria* • Cases with process failure by AutoPap (22/1300, 1.73%) • No AutoPap review (2) • Slide read twice (1)	Quality: fair Q2 Prospective: yes Consecutive: no Explicit selection criteria: no Test threshold specified: no (sens & spec) Reference standard • valid (suboptimal) • applied to all discordant & positive participants

	Adequacy discrepancy: cytological adjudication (1 senior cytotechnologist) Concordant neg: nil (assumed TN) *Test threshold* Test: NR Reference standard: NR (likely ASCUS+) *Cytological diagnosis system* 1991 Bethesda system	*Patient characteristics* NR • 9.7% (124/1275) reference standard HSIL+, including 6.8% carcinoma cases	Test interval adequate • reference standard—N/A • comparator—yes Tests reported blinded to ref standard—yes Ref standard reported blinded to tests—yes Uninterpretable/intermediate results reported: no (not contingent on tests) Study withdrawals explained: yes Sufficient data for contingency table: no Applicability: limited P2 Applicable population: not applicable (many seeded cases) Applicable comparator: yes Applicable intervention: no (includes NFR)

RESULTS:

Test accuracy

Sensitivity, specificity

Study reports overall sensitivity of appropriate *triage to pathologist review* was 98.4% for location-guided screening and 91.1% for current practice (not appropriate outcome)

Reference standard threshold unclear—likely to be ASCUS+ for sensitivity & specificity calculations

No true estimates as no ref std of concordant negatives, so true prevalence unknown and spectrum difference from screening population

Accuracy for HSIL+

Test threshold HSIL+

AutoPap = 103 TPs, manual = 73 TPs (χ^2 P = 0.019). AutoPap identified 30 additional TPs. Not applicable to screening population

FPs for HSIL+ test threshold not reported

Estimated difference in sensitivity = 30/124 = 24.2%

Study reports sensitivity (AutoPap 83.1%, manual 58.9%) but not true estimate; no ref std of concordant negatives and spectrum difference from screening population

Test threshold ASCUS+

AutoPap = 122 TPs, manual = 111 TPs, AutoPap identified 9 additional TPs

Test characteristics

Yield

NR according to screening methods <u>Process outcomes</u> No further review following AutoPap 218/1275 (17.1%) *Not processed:* 22/1300 (1.69%) *Unsatisfactory:* NR separately for manual vs auto screening *Auto classified as no further review:* 17.1%

Abbreviations: Se = sensitivity, Sp = specificity.

Appendix G Clinical flowchart

Figure 15Clinical flowchart of cervical cytology testing

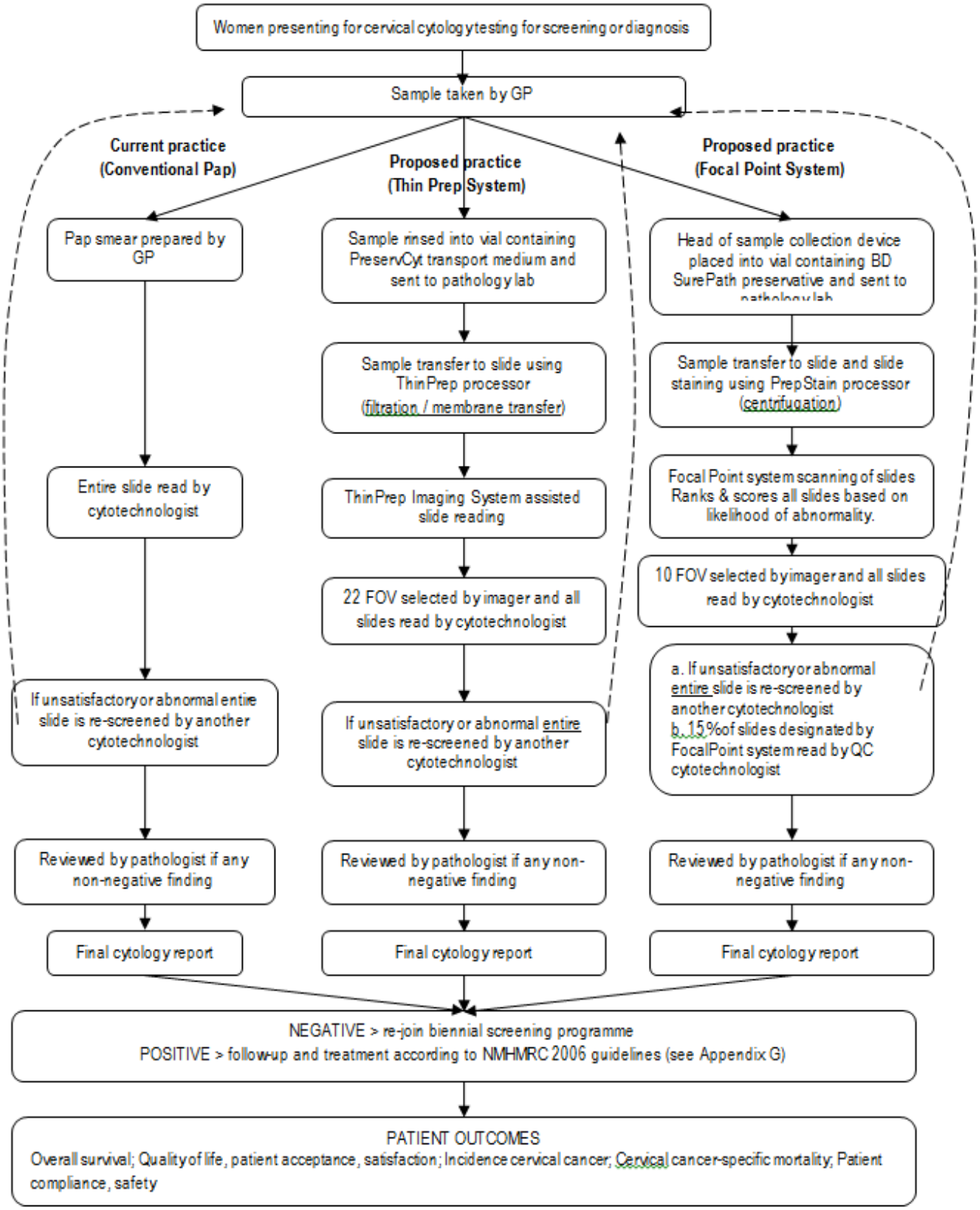
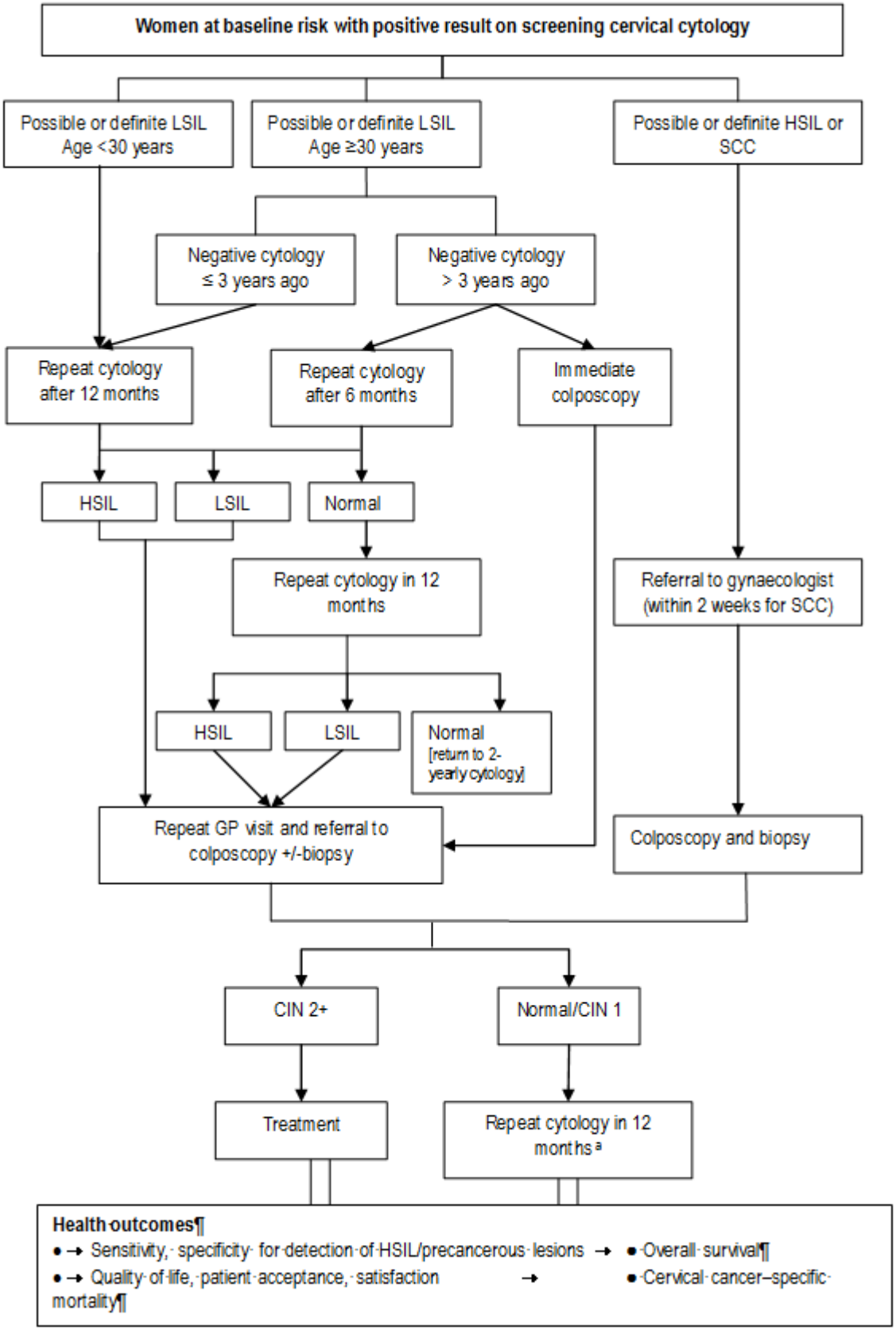


Figure 16 Management of participants testing positive in screening program.



[a] If repeat cytology tests at 12 and 24 months show no abnormalities, these women return to 2-yearly cytology

Appendix H Existing HTA reports 2002–2008

Relevant existing HTAs and systematic reviews published since the previous MSAC reviews (2002–2008).

Organisation Country	Date	Title	Scope
Arbyn et al. (systematic review)	2008	Liquid compared with conventional cervical cytology: a systematic review and meta-analysis	Manual LBC versus conventional cytology (accuracy). Search to May 2007
CADTH 2008 (Krahn et al.) Canada	2008	Liquid-based techniques for cervical cancer screening: systematic review and cost-effectiveness analysis	Manual LBC versus conventional cytology (accuracy & unsatisfactory rates). Search to June 2006
Davey et al. (systematic review) Australia	Jan 2006	Effect of study design and quality on unsatisfactory rates, cytology classifications, and accuracy in liquid-based versus conventional cervical cytology: a systematic review	Manual LBC versus conventional cytology in cervical screening. Accuracy, yield & unsatisfactory rates. Search 1966 to Jan 2004
Danish Centre for Evaluation and Health Technology Assessment Denmark	2005	The use of liquid-based cytology (LBC) and conventional Pap smear (CPS) for cervical screening in Denmark: a health technology assessment	Conventional Pap smear (CPS) versus LBC, including consequences related to automated reading & HPV testing. English summary only
UK NHS R&D HTA programme (Willis et al.) UK	2005	Cervical screening programmes: can automation help? Evidence from systematic reviews, an economic analysis and a simulation modelling exercise applied to the UK	Automated (AutoPap) versus manual reading conventional cytology. Search to end of 2000
UK HTA (Karnon et al.)	May 2004	Liquid-based cytology in cervical screening: an updated rapid and systematic review and economic analysis	Update on effectiveness and cost-effectiveness of manual LBC vs conventional Pap smear Search 1999 – October 2002 Economic analysis based on English Pilot Study report (revised January 2003).
Canadian Coordinating Office for Health Technology Assessment CCOHTA (Noorani et al.) Canada	2003	Liquid-based cytology and human papillomavirus testing in cervical cancer screening	Manual LBC versus conventional cytology. Search to July 2003
German Agency for Health Technology Assessment at the German Institute for Medical Documentation and Information	2003	Efficacy of liquid-based and computer-assisted cervical cytology - medical efficacy, economic efficiency	Considers manual LBC & automated LBC separately & together, including economic analysis Search to Nov 2001
ECRI USA	2003	Automated thin-layer slide preparation systems for cervical cancer screening	Report not freely available
Institute for Clinical Systems Improvement	Aug 2003	Liquid-based cervical cytology	Manual LBC versus conventional cytology Search dates unclear

Appendix I Davey et al. (2006) LBC review

The most recent systematic review comparing unsatisfactory rates before the CADTH HTA was Davey et al. (2006). This review included 48 datasets (from 46 studies) in a meta-analysis of the rate of unsatisfactory slides. Overall, 0.75 per cent (3646/483 050) of LBC slides and 0.81 per cent (5389/662 401) of conventional slides were unsatisfactory. This gave a summary estimate of difference of –0.14 per cent (95% CI –0.33% to +0.06%), and a median of –0.17 per cent (LBC – conventional: interquartile range – 0.98% to +0.37%). There was strong evidence of heterogeneity in the included studies ($P < 0.0001$).

Davey et al. (2006) found that the median of the differences in unsatisfactory slide rates was not significantly different by study quality (high quality –0.07%, medium quality – 0.17%, low quality –0.12%), study design (0% paired studies, –0.17% independent studies) or setting (screening 1.02%, referral clinic 0.03%, mixed screening & referral – 0.24%, unspecified –0.38%). Meta-regression findings were consistent with this, except for a clinically non-significant association in setting (differences ranged from 0.4% in screening to –0.4% in mixed settings). There was no correlation of unsatisfactory rate with year of publication, but this measure may not truly reflect experience with reading LBC slides ($P = 0.65$).

A significant difference in unsatisfactory rates was found according to LBC proprietary name ($P = 0.01$; ThinPrep $n = 32$, AutoCytePrep $n = 7$, CytoRich $n = 8$, Cytoscreen omitted as $n = 1$) and manufacturer ($P = 0.003$; ThinPrep $n = 32$, AutoCytePrep or CytoRich $n = 15$). However, the differences in the proportions of unsatisfactory slides were small (LBC – conventional = 0.12% for ThinPrep and –1.08% for AutoCytePrep or CytoRich). The review did not provide any analyses or inclusion/exclusion criteria by generations of LBC technologies. Studies comparing the earlier ThinPrep beta with the current ThinPrep 2000 were included in the review.

This systematic review concluded that large randomised trials are needed to assess the comparative performance of LBC and conventional cytology.

Appendix J Detailed model data sources

Cervical cancer natural history parameters

Invasive cancer survival parameters

The invasive cancer survival model incorporated states for undiagnosed and symptomatically detected invasive cervical cancer, with separate states for localised cervical cancer and for cervical cancer with regional and distant metastases. We did not directly incorporate FIGO staging in the model, because the available estimates of relative survival after a diagnosis of cervical cancer were calculated according to extent of disease rather than FIGO stage. We used cumulative relative ratios of survival of invasive cervical cancer by extent of disease, based on data from the NSW Central Cancer Registry (NSW CCR) from women diagnosed with cervical cancer between 1991 and 2000, and followed up to 2001 (pers. comm., A/Prof. Dianne O'Connell, Cancer Council NSW). In the model, relative survival after both local and regional disease is assumed to plateau after 10 years, on the basis of projections from the survival data to 10 years. Among women diagnosed with distant metastases, we assumed no survivors after 10 years, based on NSW CCR data which showed that only about 3 per cent of those diagnosed were still alive after 10 years (with a cumulative relative survival probability of approximately 20%).

To model progression of undiagnosed invasive cervical cancer, we adapted previous estimates by Myers et al. (2000) of the rates of progression between undiagnosed FIGO disease states. We also calculated rates of development of symptomatic and therefore diagnosed disease for each possible disease extent from rates in Myers et al. (2000).

Screening parameters

Compliance with screening and management recommendations

The model incorporated information on compliance with screening and management recommendations obtained via analysis of data from the Victorian Cervical Cytology Register (VCCR). During 1995–2007 in Victoria, there were 7 501 419 cytology tests for 1 839 179 women, and 218 171 histology tests for 165 201 women.

Probability of being rescreened

The rescreening probabilities were calculated by using standard cohort analysis methods, taking account of the person-time of follow-up and possible censoring. For each index smear, we calculated the earliest of (i) the time to the next smear, (ii) time to death, (iii) 10 years of follow-up, or (iv) time to 31 December 2007. Smears before 1997 were excluded, as there is uncertainty about the level of individual level matching before 2000. Data from 1997–1999 were included to allow for some estimation of longer-term rescreening probabilities. The follow-up was stratified by 3-monthly periods, with recalculation of age and period for each stratum of follow-up. We then aggregated the person-time and the number of events to calculate rates, and subsequently calculated the interval-specific probabilities of rescreening. Rescreening probabilities over longer time intervals were then reported according to the index smear follow-up recommendation and whether a woman had had a high-grade histology in the previous 5 years.

From this, the VCCR data were used to derive rescreening and management probabilities for 10-year periods after each screening or follow-up investigation. These were then applied to the appropriate section of the model. Beyond these 10 year periods, we assumed that each year, among the remaining women who have still not attended any follow-up, 30 per cent of women aged 15-39 years, 20 per cent of women aged 40-49 years, 10 per cent of women aged 50-59 years and 5 per cent of women aged 60 years or more will finally re-attend.

Screening initiation

Age at screening initiation (i.e. the distribution of ages at first attending for screening) in the model was informed by the national screening recommendations and target ages of the screening program, data from the VCCR, national data on screening participation at young ages, and survey data on women who report they have never had a Pap smear test (ABS 2002; ABS 2006; AIHW 2007b; NHMRC 2005; VCCR 2006).

We initially assessed age at first screening by using VCCR data, as measured by age at first recorded test for a given woman. We did not rely solely on the VCCR data for the distribution of ages at first screening, because many women were over the age of 20 at the time that the National Cervical Screening Program was introduced in 1991. Before this time, opportunistic screening occurred, but we considered that it would not be representative of behaviour in the context of the organised national program. We used the age at first test during 2005–2007 to represent the pattern for an established screening program for younger women.

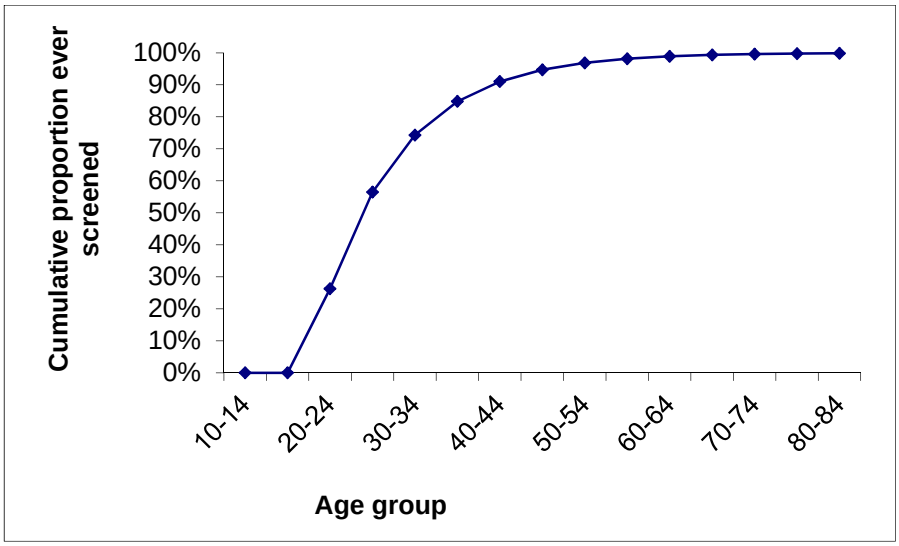
The final model was parameterised as follows. The proportion of women who initiate screening under the age of 20 was based on information from the VCCR. From age 20, the additional proportion starting screening each year was chosen so that:

- the age-standardised rate of women aged 20 or more who are never screened is 12 per cent (ABS 2002; ABS 2006)

- the average proportion of women ever screened in the 20–24 years age group was at least as high as 2-year participation in cervical cancer screening in this age group (AIHW 2007b).

Most women are expected to attend for their first test soon after the age of 20 years, and the median age of first screening is 22 years. We assumed that over 90 per cent of women will have had at least one screening test by the age of 40 years. The model also assumes that a small proportion of women (~3%) will never be screened during their lifetime, even in the context of the current organised screening program. The modelled proportion of women who have ever been screened by age is shown in Figure 17.

Figure 17 Modelled cumulative proportion of women ever screened by age.



Follow-up recommendations specific to the 2005 NHMRC Guidelines

The 2005 NHMRC *Guidelines for the Management of Asymptomatic Women with Screen Detected Abnormalities* generally recommend 12-month cytological surveillance after a low-grade smear. To model compliance with 12-month follow-up, we used the VCCR data to calculate the distribution of probabilities of re-attending over time for women who were given a 12-month follow-up management recommendation.

The 2005 NHMRC Guidelines incorporate a specific recommendation for more active follow-up in a particular group of women with index low-grade smears (these women are termed 'exceptions' here). Exceptions are defined as women with an index LSIL smear who are aged 30 years or older and have no recent history of negative cytology (within 2–3 years). These women should either have immediate colposcopy or a repeat smear in 6 months, at the discretion of the clinician. To estimate the proportion of women assigned to each of these management paths, we used VCCR data to assess management practice since the implementation of the 2005 NHMRC Guidelines.

Low-grade cytology tests from 1 July 2006 were extracted from records of women who (i) were aged 30 years and over, (ii) had no high-grade histology in the previous 5 years and (iii) had no cytology tests in the previous 3 years. Management recommendations for women 30 years and over with an index smear of possible or definite LSIL and no history of negative cytology in the previous 3 years were then assessed. This information is presented in Table 60.

Table 60 Management recommendations for 'exception' cases under the 2005 NHMRC Guidelines—analysis of VCCR data.

Management recommendation	Percentage
No recommendation	0.73
Repeat smear 12 months	25.84
Repeat smear 6 months	8.84
Repeat smear 6–12 weeks	1.70
Colposcopy/biopsy recommended	49.14
Already under gynaecological management	5.88
Referral to specialist	0.39
Other management recommended	1.23
Symptomatic—clinical management required	6.25

To estimate the proportions following each management pathway in the model, we then excluded women with 'No recommendation' and 'Other management recommended' from the analysis. We also excluded women with a recommendation of 'Already under gynaecological management', 'Referral to specialist' and 'Symptomatic—clinical management required', because we assumed that these women would be accounted for elsewhere in the model, or would already be under management for causes unrelated to cervical cancer.

Therefore, the final assumption used in the model was that 28 per cent of exceptions receive a recommendation of a repeat smear in 12 months, 11 per cent receive a recommendation of a repeat smear within 6 months, and 61 per cent are referred for immediate colposcopy.

Test characteristics—cytology

The model requires parameters that specify the relationship between each possible underlying natural history health state at the time of testing, and each possible test result. In practice, many of these relationships are difficult to observe in the context of routine screening. We therefore derived a series of test probability matrices for each of conventional, liquid-based and automated cytology, based on the available evidence.

First, we derived a test probability matrix for conventional cytology (CC). We used published data to estimate the distribution of true underlying health states within each cytology result category, and data from registries to determine Australian reporting rates for each cytology test result. These data can be transformed using Bayesian methods to derive the probability of a particular cytology test result, given an underlying health state. Our final choice of baseline test characteristics for conventional cytology was made on the basis of best and most complete verification of results, use of Australian data where possible, and consistency of relative test performance with published data.

Test probability matrices for manual and automated reading of LBC slides were derived from the baseline CC matrix. We adjusted the distribution of cytology test results to reflect the relative distribution while keeping constant the prevalence of disease implied by the matrix.

We estimated call rates for LBC and automated LBC relative to conventional cytology using data from Davey et al. 2006 and Davey et al. 2007a, and then we applied these ratios to the age-specific call rates for conventional cytology.

Where assumptions had to be made, or constraints could not all be met during the calibration phase, we made assumptions which favoured automated and manual LBC over conventional cytology.

The accuracy of the conventional cytology implied by the selected test probability matrix is detailed in Table 28 (page 69). The calibrated accuracy of automated LBC relative to either conventional cytology, or manual LBC is shown in Table 29 (page 69; see also Table 49 and Table 50 in 'Calibration of the screening model'). A range of values were derived for sensitivity analyses.

The derivation of the test probability matrices is described in detail in 'Appendix K Derivation of test probability matrices for cytology'.

Cytology unsatisfactory rates

We used the available data on cytology unsatisfactory rates in Australia to derive baseline values and ranges for sensitivity analysis, as summarised in Table 61 (reproduced from Table 30, page 70).

For the baseline estimate for conventional cytology, we used aggregate national data from Australian laboratories for 2007 (Royal College of Pathologists Australasia Cytopathology Quality Assurance Programs, 2008). In 2007, 46 451 (2.2%) of 2 135 214 specimens received were reported as unsatisfactory. This figure is consistent with registry data from the VCCR and NSW Pap Test Registry (PTR). In 2007, 11 259 (2.1%) of 546 012 smears in Victoria were reported as unsatisfactory (from our analysis of data from VCCR). During the fourth quarter of 2006, 3843 (2.3%) of 163 568 smears processed in NSW were reported as unsatisfactory (pers. comm., Grace Kwan, Cancer Institute NSW).

These values may be underestimates of the true unsatisfactory rate for conventional cytology, because some samples were split conventional and LBC samples, in which only one sample is required to be satisfactory. For this reason, extensive sensitivity analyses was performed, which was informed by the range of unsatisfactory rates reported by Victorian laboratories in 2006 (0.7% - 4.9%) (VCCR Statistical Report 2006).

We used 1.8 per cent as the baseline unsatisfactory rate for automated LBC slide reading, based on data from Davey et al. (2007a). The range of unsatisfactory rates for automated cytology used in sensitivity analyses was based on data from a pilot study in a Sydney laboratory (pers. comm., Dr Annabelle Farnsworth, Douglass Hanly Moir Pathology) and on the range of values found in the systematic review described in earlier sections of this document (Table 19, page 46; Table 25, page 57), which includes a published analysis from a study conducted in Ireland (Bolger et al. 2006).

No Australian registry data on the unsatisfactory rate for manual LBC were available (because manual LBC is currently used only as an adjunct to conventional cytology), nor were data from published Australian comparative studies of manual and automated LBC reading.

Bolger et al. 2006 found similar unsatisfactory rates for manual and automated reading of LBC slides (1.05% vs 1.07%). Therefore, we used the same unsatisfactory rate for manual LBC as we did for automated LBC (1.8%). This similarity in unsatisfactory rates is consistent with results from a pilot study in a Sydney laboratory (1.98% vs 2.08%, $P = 0.75$, χ^2), which compared manual and automated LBC with CC (pers. comm., Dr Annabelle Farnsworth, Douglass Hanly Moir Pathology). To inform the range for sensitivity analysis, in the absence of Australian data, we used an estimates based the systematic review described earlier in this document ('Unsatisfactory rate', page 32). Data come principally from a meta-analysis included in a recent Canadian HTA report (CADTH; Krahn et al. 2008), and from two more recent studies (Ronco et al. 2007a) (Bolger et al. 2006), not included in that meta-analysis.

During sensitivity analyses, we varied sets of rates for all three tests. The sets were constrained in a way such that the unsatisfactory rate of conventional cytology was always the same or greater than that of automated or manual LBC.

Table 61 Model values for rates of unsatisfactory cytology, by test technology.

Test type	Unsatisfactory rate used in model	Range for sensitivity analysis [a]	Reference
	%	%	
CC	2.2	0.5–5.0	Davey et al. 2007a; data from NSW Pap Test Registry & VCCR
Manual LBC	1.8	0.5–2.57	Davey et al. 2007a; Bolger et al. 2006; pilot study in DHM [b]; Ronco et al. 2007
Automated LBC	1.8	0.5–2.0	Davey et al. 2007a; Bolger et al. 2006; pilot study in DHM

[a] Within these ranges, the sensitivity analysis of the relative performance of the test technologies was constrained such that the unsatisfactory rate of CC was always ≥ the unsatisfactory rate of manual or automated LBC.

[b] Douglass Hanly Moir Laboratories, Sydney.

Modelling assumed that all unsatisfactory tests were repeated immediately (no non-compliance), before any natural history progressions or regressions were applied, and that all repeat tests were satisfactory.

Test characteristics—colposcopy

We derived a test probability matrix for colposcopy based on data on over 21 000 colposcopies supplied by the Royal Women's Hospital in Victoria (pers. comm., Dr Jeffrey Tan, Royal Women's Hospital, Melbourne). This specifies the relationship between each possible underlying natural history health state at the time of testing and the colposcopy result. In the model, this information was used to specify the probability that a biopsy would be taken, according to the underlying health state: 88.4 per cent for the CIN 2–3 threshold and 76.5 per cent for CIN 1 (after exclusion of unsatisfactory colposcopy results) (Table 62).

Table 62 Modelled test characteristics—colposcopy.

Model health state	Probability that a biopsy will be taken at colposcopy (%)
Normal	50.2%
HPV	50.2%
CIN 1	76.5%
CIN 2	88.4%
CIN 3+	88.4%

The data from the Royal Women's Hospital were also used to derive an age-specific probability of unsatisfactory colposcopy (Table 63), which was used in the model.

Table 63 Probability that colposcopy will be unsatisfactory—modelled values.

Age	Probability that colposcopy will be unsatisfactory (%)
15–24	2.01
25–29	2.78
30–34	6.03
35–39	7.5
40–44	12.56
45–49	19.58
50–54	30.98
55–84	45.66

In the model, the probability that a follow-up colposcopy would also be unsatisfactory (ie, when performed after an initial unsatisfactory colposcopy) was specified as 91.74 per cent.

Test characteristics—Hybrid Capture II (HCII)

The test characteristics of HCII were assessed in order to simulate current practice in Australia, where it is used as a test of cure following treatment for CIN 2 and 3.

We have used data from a summary of meta-analyses and from a number of international studies which compared HCII, HPV DNA testing with polymerase chain reaction (PCR) amplification, cytology and histology. The reference standard used for a true HPV positive was PCR positivity for a high-risk HPV type.

HCII test positivity rates for CIN 2 and CIN 3+ were derived from a summary of meta-analyses of HPV testing (Arbyn et al. 2006), and the range for sensitivity analysis was based on the 95% CIs presented in this summary. CIN 2 and CIN 3+ were histologically confirmed.

HCII test positivity rates for CIN 1 were derived from a study in which women from a screening population referred for any cytological atypia had histological confirmation of CIN 1 (Soderlund-Strand et al. 2005). We have used an upper-end estimate for sensitivity analysis based on women with cytological LSIL, but no diagnosis of CIN 2+ within 2 years (Zuna et al. 2005).

HCII test positivity rates and ranges for the Normal and HPV (no CIN) model states were based on HCII and PCR results in cytologically normal women from screening populations (Kulmala et al. 2004; Riethmuller et al. 1999; Safaeian et al. 2007; Sandri et al. 2006; Venturoli et al. 2002; Wahlstrom et al. 2007; Yamazaki et al. 2001).

Treatment failure and post-treatment recurrence

The treatment failure rate within the first year and annual recurrence rates thereafter were obtained from our own review and meta-analysis of data from relevant studies (Baldauf et al. 1998; Bigrigg et al. 1994; Cecchini et al. 2002; Flannelly et al. 1997; Oyesanya et al. 1993; Paraskevaidis et al. 2000; Powell 1996; Wright et al. 1992).

We assumed that 93.6 per cent of women receiving appropriate treatment for a confirmed high-grade lesion will have a successful treatment outcome within the first year (Baldauf et al. 1998; Bigrigg et al. 1994; Cecchini et al. 2002; Flannelly et al. 1997; Oyesanya et al. 1993; Paraskevaidis et al. 2000; Powell 1996; Wright et al. 1992), of whom 15.8 per cent would be HPV-positive after 1 year (MSAC 2005) (Table 64).

We also assumed that a small proportion of both HPV-positive and HPV-negative patients would have recurrent disease 1 year after successful treatment. The rate of high-grade CIN recurrence is assumed to be 1.1 per cent in years 1–2, 0.6 per cent in years 2–3, 0.2 per cent in years 3–4 and 0.1 per cent after the 4th year (Table 65). Of women experiencing recurrent disease, 32.1 per cent are assumed to develop CIN 2, and 67.9 per cent to develop CIN 3 (Cecchini et al. 2002; Flannelly et al. 1997; Rema et al. 2007) (Table 66).

Table 64 Post-treatment HPV status by PCR at 12 months.

HPV status	Baseline %	Reference
HPV –ve	84.2	MSAC 12e report
HPV +ve	15.8	MSAC 12e report

Table 65 Post-treatment CIN 2/3 detection.

Follow-up time	Baseline %	95% CI %
0–1 y	6.4	5.0–7.9
1–2 y	1.1	0.2–1.9
2–3 y	0.6	0.2–1.1
3–4 y	0.2	–
4 y +	0.1	–

Table 66 Proportion of detected lesions that are CIN 2 or 3.

Recurrent lesion detected	Baseline %	Range of combinations %
CIN 2	32.1	18.8–49.0
CIN 3	67.9	81.2–51.0

We derived the relative risk of developing recurrent disease 1 year after the successful treatment according to post-treatment HPV status. Combining the estimated annual

recurrence rate with the proportion HPV positive at 1 year after treatment gives the row and column sum of the 2 × 2 table for post-treatment, recurrent, high-grade CIN given an HPV test result. This table imposes a one-to-one relationship between the associated positive and negative likelihood values, and restricts the range of positive and negative likelihood values obtained from MSAC 12e. We calculated the range of possible sets of pairs ([true positive, true negative], [true negative, false negative]), based on the range of possible pairs of positive and negative likelihood ratios (LR+, LR–), to give upper and lower estimates for the 2 × 2 table. We further restricted the range of solutions to those with a sensitivity of 80–100 and a specificity of 81–100. The baseline matrix for post-treatment recurrence was calculated by using the midpoint of the acceptable range of the positive likelihood (Table 67, Table 68, Table 69).

Table 67 Baseline matrix for post-treatment recurrence.

HPV status after successful treatment	Recurrence %	No recurrence %	Total %
HPV+	0.44	15.36	15.80
HPV–	0.06	84.14	84.20
Total	0.50	99.50	100.00

The relative risk of recurrence is 36.9; the sensitivity of HPV testing for recurrent disease is 88 per cent, and the specificity is 85 per cent.

Table 68 Matrix for post-treatment recurrence (lower estimate for relative risk).

HPV status after successful treatment	Recurrence %	No recurrence %	Total %
HPV+	0.40	15.40	15.80
HPV–	0.10	84.10	84.20
Total	0.50	99.50	100.00

The relative risk of recurrence is 21.4; the sensitivity of HPV testing for recurrent disease is 80 per cent, and the specificity is 85 per cent.

Table 69 Matrix for post-treatment recurrence (upper estimate for relative risk).

HPV status after successful treatment	Recurrence %	No recurrence %	Total %
HPV+	0.47	15.33	15.80
HPV–	0.03	84.17	84.20
Total	0.50	99.50	100.00

The relative risk of recurrence is 74.0; the sensitivity of HPV testing for recurrent disease is 94 per cent, and the specificity is 85 per cent.

Management assumptions made for modelling in cases where guidelines do not specify outcomes

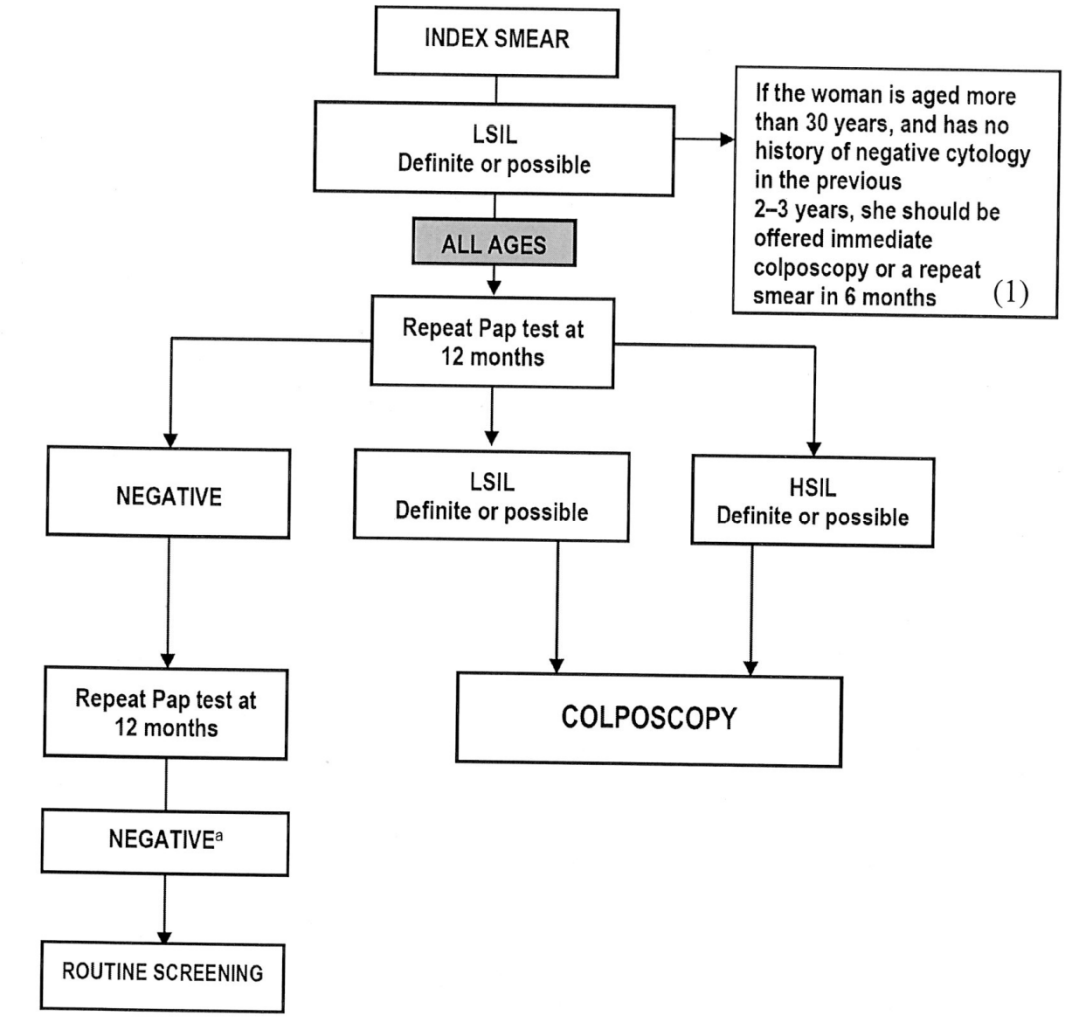
Management of LSIL

- There are some alternative formations in the 2005 NHMRC Guidelines regarding the age cut-off for variations in management of LSIL (possible or definite). For modelling purposes, we have used the two age groups *under 30 years* and *30 years or more* to be consistent with the Summary of Guidelines [p. xi].[2]

[2] Page numbers in [ ] refer to the 2005 Guidelines (NHMRC 2005).

- (Figure 18 (1)) Women who are aged 30 years or more and have no recent history of negative cytology may be given a repeat smear in 6 months following an index LSIL smear. Thereafter, for the model, we assume the same subsequent management path as for women who are given a 12-month follow-up test: that is, an LSIL[3]/ HSIL[4] result is referred to colposcopy, and a negative result requires a follow-up smear in 12 months (18 months from index smear). If this second follow-up smear is negative, then women return to routine screening.

Figure 18 Management of a cytological prediction of possible or definite LSIL.



LSIL = low-grade squamous intraepithelial lesion; HSIL = high-grade squamous intraepithelial lesion

a See Section 6.7 for discussion and a guideline on fluctuating abnormalities

'Expiry' of cytological or histological status

In several places in the model it was necessary to reset or redefine the index smear. Reasons included non-compliance with screening recommendations or colposcopy

[3] LSIL refers to both possible and definite low-grade squamous intraepithelial lesions

[4] HSIL refers to both possible and definite high-grade squamous intraepithelial lesions

referral, or being stuck in a loop of repeated follow-up where the reason for follow-up may not still be relevant. We made the following assumptions:

- Colposcopy participation rates are approximated by probabilities from the VCCR on the rescreening of women referred to colposcopy, stratified by age and current cytology result.
- After an LSIL, in the absence of follow-up screening (i.e. non-compliance with the Guidelines), low-grade cytology results have a **3-year** expiry. That is, if an LSIL smear is not followed up according to the guidelines, then another LSIL which is more than 3 years from the first LSIL is regarded as the new index smear and not referred to colposcopy. (Guidelines p. 40 state that 'referral for colposcopy should be considered for a woman if she has two LSIL reports within a 3-year timeframe, regardless of intervening normal cytology reports.' We assume that this 3-year timeframe also holds in the absence of adequate or any intervening cytology.)
- Low-grade histology has a **3-year** expiry. Histologically confirmed low-grade abnormalities are managed by annual cytological monitoring, with return to routine screening triggered by two negative smears [p. 49]. At the stage of annual monitoring we assume that women with LSIL smears are not referred to colposcopy within 3 years of the histology result, but that after 3 years women with repeated LSIL reports would be referred to colposcopy owing to the possibility of further progression of underlying disease, following the same management as for repeated low-grade cytology.

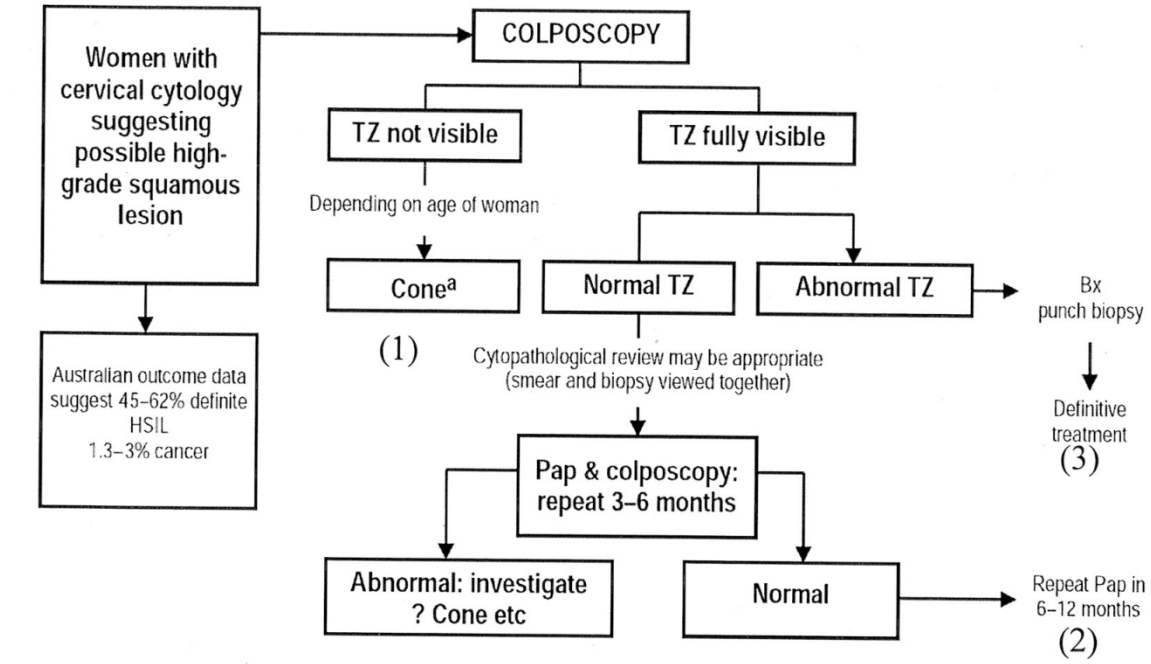
Colposcopy management

HSIL cytology is always referred to colposcopy, and LSIL cytology sometimes, but the guidelines do not give complete details of all possible management paths following colposcopy, leaving some details to the discretion of the practitioner (Figure 19). In these situations we have constructed the model as follows:

- (Figure 19, footnote a) HSIL cytology, unsatisfactory colposcopy (TZ not visible) in women whose fertility is an issue. The Guidelines recommend repeating colposcopy. We define women whose fertility is an issue as women aged ≤45 years. If the second colposcopy is also unsatisfactory then we assume that treatment would be offered: 1/3 will have immediate treatment and 2/3 will choose to delay treatment for 1 year.
- (Figure 19 (1)) HSIL cytology, unsatisfactory colposcopy (TZ not visible) in women whose fertility is not an issue, cone biopsy. We assume the following management pathway following cone biopsy: high-grade cone biopsy results take the same management as follow-up treatment for HSIL; low-grade and negative cone biopsy results take the same management as for histologically confirmed LSIL.
- (Figure 19 (2)) HSIL cytology, followed by satisfactory colposcopy (TZ fully visible) with normal TZ, then normal colposcopy and normal cytology at 3–6 months, then repeat Pap test at 6–12 months. We assume the following management pathway following repeat Pap at 6–12 months: any suspected abnormalities in the repeat Pap smears are referred to colposcopy; negative result requires repeat Pap test 12 months later. Repeat 12 months later, then return to routine screening. This is the same management as for low-grade cytology.

- (Figure 19 (3)) HSIL cytology, satisfactory colposcopy (TZ fully visible), TZ abnormal, punch biopsy showing negative histology. We assume the following management pathway: conservative approach following the same management as for histologically confirmed low-grade squamous abnormalities [p. 49].

Figure 19 Management of women with surgical cytology predicting possible high-grade squamous lesions.



HSIL = high-grade squamous intraepithelial lesion; TZ = transformation zone; Bx = targeted biopsy
a In women where fertility is an issue, repeat colposcopy

- LSIL cytology, satisfactory colposcopy (TZ fully visible), TZ abnormal, punch biopsy showing negative histology. We assume the following management pathway: conservative approach following the same management as for histologically confirmed low-grade squamous abnormalities [p. 49].

Treatment and post-treatment modelling assumptions

- Participation in treatment is not perfect, and 3 per cent of women choose to delay treatment for 1 year.

- The impact of treatment failure is estimated by the fraction of treated women who receive treatment twice, according to reported treatment failure rates.

- After successful treatment for high-grade disease, women return to health states that are histologically negative, but they may still be HPV-infected. Thus, allowable health states immediately after treatment are Well and HPV.

- Cone biopsy is given to women whose TZ is not fully visible. If a woman is later found to have had low-grade or negative histology, she returns to a Well state of health.

- Cone biopsy has the same properties as LEEP (loop electrosurgical excision procedure) when used for high-grade disease, including treatment success rate and post-treatment recurrence rate, but has different costs.

- Natural history parameters are altered in women after treatment for high-grade disease owing to the high risk of recurrence in this group. The exception is women treated with cone biopsy who are then found to have low-grade or negative histology—they are given the same natural history parameters as the untreated population.

Women who have been treated for histologically confirmed high-grade disease by LEEP or cone biopsy are followed up by annual cytology and HPV testing until both tests are negative for 2 years in a row, then follow-up is cytology at the routine interval (2 years). Thereafter, in this subpopulation, we made the following assumptions (for simplicity):

- All women with a subsequent high-grade cytology result will be retreated, regardless of colposcopic investigative findings or HPV positivity or negativity.

- Women with low-grade cytology will be referred to colposcopy. Unsatisfactory colposcopy will be treated with cone biopsy. Satisfactory colposcopy with abnormal TZ will be treated on the basis of biopsy: women with high-grade histology are treated by LEEP, and women with low-grade or negative histology are followed up with annual cytology and HPV testing until both tests are negative twice consecutively. Satisfactory colposcopy with normal TZ is also followed up with annual double testing.

- Cytology, colposcopy and biopsy have the same test characteristics in women after treatment as for the untreated population.

- After testing negative in both cytology and HPV tests for 2 years in a row, women return to 2-yearly screening with cytology alone. We assume thereafter that any abnormality is referred to colposcopy, that high-grade cytology is treated regardless of colposcopic findings, and that follow-up after the second treatment is done by annual cytology and HPV testing, as for follow-up after initial treatment.

Appendix K Derivation of test probability matrices for cytology

Introduction

The model requires probabilities that specify the relationship between each possible underlying natural history health state at the time of testing and each possible test result. These probabilities of receiving any test result given a true underlying health state form a table called the 'test probability matrix'. A test probability matrix is required for CC, LBC and automated cytology (based on characteristics of the ThinPrep Imager).

We assume that the sensitivity and specificity of each screening test is independent of the age of the woman at the time of testing, and hence that the test probability matrix is age-independent. Although underlying disease prevalence varies with age, the model reflects this variation in the distribution of health states in the population at each age: the test probability matrix is not required to take prevalence into account.

In each test probability matrix there are five possible underlying health states (Table 70). For each underlying health state there are five possible test results: negative, possible LSIL (pLSIL), definite LSIL (dLSIL), possible HSIL (pHSIL) and definite HSIL (dHSIL), each defined according to the AMBS 2004 classifications. The test probability matrix gives complete information on test characteristics, and the test parameters for sensitivity, specificity, positive predictive value and negative predictive value can be secondarily derived for thresholds at any health state (such as high-grade lesions [CIN 2+] or all CIN lesions) and for any testing threshold (such as pLSIL or pHSIL).

Table 70 Definition of model health states.

Model health state	Gold standard definition
Normal	PCR negative; negative histology
HPV	PCR positive; negative histology
CIN 1	CIN 1 histology
CIN 2	CIN 2 histology
CIN 3+	CIN 3 histology or cervical cancer

CIN, cervical intraepithelial lesion; PCR, polymerase chain reaction.

The probabilities that appear in the test probability matrix cannot all be observed directly and must be derived from a range of data sources. They also depend on observed cytology abnormality rates, since these determine the operating sensitivity/specificity trade-off (operating point on the ROC [receiver operating characteristic] curve). We have estimated these parameters by using a variety of data sources (Arbyn et al. 2008; Coste J. 2003; Krahn et al. 2008; Medical Services Advisory Committee 2002a; Nanda et al. 2000; Sherman et al. 2006; Victorian Cervical Cytology Registry 2006) and have found it difficult to reconcile estimates derived from registry data or studies based on a general screening population with estimates derived from studies or meta-analyses which were based on selected trial populations. In this section we explain various approaches towards these parameter estimates, the difficulties involved, and reasons for the final choice of a baseline test probability matrix used in the model.

We expect that relative differences between the matrices for conventional cytology, LBC and automated LBC will drive differences in outcomes. In this section we also explain

our methods for deriving matrices for LBC and automated LBC from the baseline matrix for conventional cytology. Sensitivity analysis of the model includes testing the assumptions in relative differences between both the test probability matrices and changes to the baseline matrix.

Data summary

ALTS

The ASCUS and LSIL Triage Study (ALTS) was a major randomised trial conducted in the US to assess the relative benefits of various protocols for the management of low-grade smear abnormalities (Schiffman & Adrianza 2000; Schiffman & Solomon 2003; Solomon D et al. 2001; Walker et al. 2006). Enrolment for ALTS commenced in 1997, and the TBS1991 classification system was originally used. Later reanalysis used the TBS2001 system (Sherman et al. 2006).

ALTS involved cluster-randomising women with low-grade abnormalities into three arms, which were managed according to different strategies: conservative management (cytology only), immediate colposcopy, and HPV DNA triage with HCII testing. An HSIL cytology result in the cytology arm or an HPV DNA-positive result in the HPV triage arm triggered colposcopy referral. All women were followed up for 2 years, with cytology every 6 months and referral for colposcopy for an HSIL result. At the 2-year exit visit all women underwent colposcopy with biopsy for any suspected low- or high-grade abnormalities. Within ALTS, a QC panel reviewed cytology and histology diagnoses, and therefore inter-observer variability in diagnosis was minimised. We use the most severe histologically confirmed abnormality obtained in the 2-year period as a proxy for the underlying health state at the time of enrolment. Women participating in the study had a mean age of 29 years and a median age of 25 years.

ALTS is not a sufficient study for our purposes in that the population is not representative of the entire screening population: women with HSIL smear results or negative smear results were not eligible for entry into ALTS. There are some data on high-grade cytology, as a QC panel reviewed cytology slides from the time of enrolment, and some low-grade slides were reanalysed as ASC-H or HSIL, but these are not representative of high-grade smears in the general screening population. The results for low-grade cytology, however, should be representative of the screening population, as women were enrolled directly from this population on the basis of an entry ASCUS or LSIL smear. An advantage of ALTS is that there is essentially complete verification of low-grade cytology with colposcopy ± biopsy.

From the data in ALTS, we were able to estimate the breakdown of various cytology test results into true underlying health states. Methods 3 and 4 (below) use data from ALTS for all cytological abnormalities. Methods 5 and 6 use data from ALTS for low-grade abnormalities.

We assessed the effect of using lower estimates of the CIs presented in ALTS (Sherman et al. 2006) for the proportions of women with low-grade cytology and high-grade histology, and of using upper estimates of the CIs for the proportions of women with high-grade cytology and high-grade histology. This had the aim of achieving sensitivity for detection of CIN 2+ at the HSIL threshold that approached published estimates (Arbyn et al. 2008).

Registry data

We used data from the Victorian Cervical Cytology Register (VCCR; data provided by VCCR, July 2008) and NSW Pap Test Registry (NSW PTR; pers. comm., Robyn Godding, Cancer Institute NSW).

Registry data on histology for women with high-grade cytology was extracted from the VCCR data for 2005 and 2006. These data are more directly representative of high-grade cytology results in an Australian screening population than ALTS data on initially low-grade results which were reclassified as high-grade after enrolment.

For evaluating the concordance between histology and cytology, we restricted the histology data to 2005 and 2006. For a given woman with a given histology test, we identified the satisfactory cytology test (or tests, if on the same day) that was immediately before and within 6 months of the histology test. If there was no such cytology test, then we considered whether there was a satisfactory normal smear within 2 years of the histology test. If there were multiple tests on the same day, then we used the most severe histological result. The histology diagnosis codes were grouped according to the 6 August 2008 draft of the *Standardised Cervical Screening Data Dictionary 2008* (supplied by AIHW) with the following alterations: (i) high-grade and invasive/malignant codes were grouped together, and (ii) HPV-related cervical abnormalities were reported separately. Results were reported only for the cervical specimens, excluding the endocervical specimens.

When we combined registry data with ALTS data, we restricted analysis to the age-group 25–29 years to be as consistent as possible with the study population of ALTS. Although the final test probability matrix is assumed to be applicable to all ages and does not depend on the prevalence of health states, intermediate steps in the derivation do depend on the prevalence of health states. It is therefore important to match age-groups as far as possible when combining data from different study populations.

Method 7 uses data from the VCCR for women of all ages in place of ALTS data. Data were taken from histology findings following a cytology report during 2006 (VCCR 2006, Table 5.1). The translation of histology categories presented in this report and model health states are shown in Table 71.

Table 71 Correlation between model health states and histology categories in the VCCR Statistical Report, 2006, Table 5.1.

Model health state	Histology categories
Normal, HPV	Normal; benign; HPV effect
CIN 1	Low grade, not otherwise defined; CIN 1
CIN 2	CIN 2; 50% of high grade, not otherwise defined; 50% of CIN 2/3
CIN 3+	CIN 3; 50% of high grade, not otherwise defined; 50% of CIN 2/3; cancer—micro-invasive; cancer—invasive other; cancer—invasive squamous

Coste

One study (Coste 2003) included in the meta-analysis by Arbyn et al. (2006) had a screening population arm and was designed for the disease status of all participants to be verified. The French Society of Clinical Cytology's Study Group carried out a cross-sectional study during 1999–2000 in which conventional, LBC and HPV testing were performed simultaneously on all participants against a reference standard of colposcopy and histology. We used data from the 'optimised interpretation' results. Conventional

cytology slides were read blind twice, and in cases of disagreement, read again to reach consensus. If there was no consensus then an independent expert reading was given. There were two groups of participants: women referred for colposcopy (n = 828) and women attending for routine screening (n = 1751). We used data for conventional cytology from the screening population arm only. The size of the screening population arm verified by colposcopy and biopsy is larger than the same population in a similar Italian trial (Ronco et al. 2007a) and was chosen for this reason. Davey et al. rate this study as a high-quality study (Davey et al. 2006). Information on HPV status in a subset of women was available (de Cremoux et al. 2003).

Data sources for negative cytology estimates

In all studies of test accuracy there is difficulty involved in histological verification of women with negative smears. Where it has been necessary to indirectly estimate the underlying health states of women with normal cytology owing to a lack of data or lack of verification, FN test results were derived through evidence of the sensitivity of LBC or CC at various cytology thresholds (Arbyn et al. 2008; Krahn et al. 2008; Nanda et al. 2000). TN results were calculated as the remainder, or through estimates of the specificity of LBC or CC at various cytology thresholds (Table 72).

Table 72 Sensitivity and specificity estimates at cytological thresholds of ASCUS, LSIL and HSIL (TBS1994 terminology).

Source (detection threshold)	Sensitivity			Specificity		
	ASCUS	LSIL	HSIL	ASCUS	LSIL	HSIL
CC						
Nanda et al. (CIN 1+)	0.68 [0.31–0.92] [a]	0.62 [0.18–0.98] [a]	–	0.75 [0.17–0.99] [a]	0.90 [0.09–1.0] [a]	–
Nanda et al. (CIN 2+)	–	0.81 [0.23–0.99] [a]	0.53 [0.18–0.92] [a]	–	0.77 [0.06–0.99] [a]	0.96 [0.64–1.0] [a]
Arbyn et al. (CIN 2+)	0.88 (0.80–0.93)	0.76 (0.67–0.83)	0.55 (0.46–0.65)	0.71 (0.58–0.82)	0.81 (0.72–0.88)	0.97 (0.96–0.98)
LBC						
CADTH (CIN 1+)	0.84 (0.73–0.95)	0.71 (0.61–0.82)	–	0.78 (0.63–0.93)	0.86 (0.73–0.98)	–
CADTH (CIN 2+)	0.89 (0.78–1.00)	0.79 (0.67–0.89)	–	0.76 (0.49–1.00)	–	–
Arbyn et al. (CIN 2+)	0.90 (0.82–0.95)	0.79 (0.70–0.86)	0.57 (0.46–0.67)	0.65 (0.50–0.77)	0.79 (0.70–0.86)	0.97 (0.94–0.99)
CADTH (CIN 3+)	0.88 (0.74–1.00)	0.71 [b]	0.57 [b]	0.92 (0.90–0.93)	0.96 [b]	0.99 [b]

Estimate (95% CI) or [a] Median [range];. [b] Only one study was included at this threshold. Numbers in **bold** type are estimates used in derivation of matrices.

These values have also been used to check the test characteristics implied by the test probability matrix. The Arbyn et al. meta-analysis also reports on relative sensitivity and specificity for conventional cytology and LBC (Table 73). As we expect that relative differences between the test probability matrices for conventional cytology and LBC will drive the outcomes of the model, these data have been used to validate the matrices.

Table 73 Sensitivity and specificity of LBC, relative to conventional cytology, for detection of CIN 2+ at cytological thresholds of ASCUS, LSIL and HSIL (TBS1994 terminology).

Source	Sensitivity			Specificity		
	ASCUS	LSIL	HSIL	ASCUS	LSIL	HSIL
Arbyn	1.03 (0.97–1.09)	1.03 (0.96–1.11)	1.05 (0.95–1.16)	0.91 (0.84–0.98)	0.97 (0.94–1.01)	0.99 (0.98–1.01)

We also estimated the probability of a normal test result given underlying HPV infection by using the cross-sectional prevalence of HPV in cytologically normal non-Indigenous

Australian women aged 25–29 years, assumed to be 22.7 per cent on the basis of interim WHINURS data from July 2006 (pers. comm., Prof. Suzanne Garland). This age group was chosen because it is close to that of the ALTS study and is therefore assumed to have similar underlying disease prevalence.

Distribution of abnormal cytology reports in the screening population

We have used the reporting rates for cytology in the screening population of Australia to derive matrices specific to Australia, or to check whether matrices derived independently of the distribution of abnormal cytology reports are applicable to Australia (Table 74). We used cytology distributions from the VCCR (July 2008) and NSW PTR (pers. comm., Robyn Godding, Cancer Institute NSW) for:

- women of all ages, VCCR whole of 2006 (VCCR 2006)
- women of all ages, NSW PTR fourth quarter of 2006
- women aged 25–29 years, VCCR during 2005 and 2006
- women aged 25–29 years, NSW PTR fourth quarter of 2006.

Table 74 Cytology report distributions, adjusted to exclude unsatisfactory smears.

	Cytology report				
Source	Normal	pLSIL	dLSIL	pHSIL	dHSIL+
VCCR (all ages)	93.9%	2.2%	3.0%	0.4%	0.5%
VCCR (25–29 y)	87.0%	3.7%	6.6%	0.8%	1.8%
PTR (all ages)	93.6%	2.7%	2.2%	0.5%	0.9%
PTR (25–29 y)	89.1%	4.0%	4.1%	0.9%	1.9%

These registry data show some differences in the reporting rates of Victoria and NSW. Therefore, our approach was to separately apply NSW and Victorian data for the relevant age group, and compare the two results by sensitivity analysis.

Nomenclature

Many of our data sources report cytology results using the Bethesda System 1991 or 2001, and one source of Australian data (MSAC 2002a) using the 1994 Australian NHMRC-endorsed terminology. We have mapped these results to AMBS 2004, taken from the 2005 Australian National Guidelines for Cervical Screening (NHMRC 2005), with the additional assumption that TBS1991 ASCUS includes TBS2001 ASCUS and TBS2001 ASC-H (Table 75).

Table 75 Comparison of the Australian Modified Bethesda System (AMBS 2004) with previous Australian terminology and the Bethesda System (TBS2001).

AMBS 2004	Australian NHMRC-endorsed terminology 1994	TBS2001	Incorporates
Squamous abnormalities			
pLSIL	Low-grade epithelial abnormality	ASCUS	Non-specific minor squamous cell changes; changes that suggest, but fall short of, HPV/CIN 1
LSIL	Low-grade epithelial abnormality	LSIL	HPV effect, CIN 1
pHSIL	Inconclusive, possible high-grade squamous abnormality	ASC-H	Changes that suggest, but fall short of, CIN 2, CIN 3 or SCC
HSIL	High-grade epithelial abnormality	HSIL	CIN 2, CIN 3
SCC	High-grade epithelial abnormality	SCC	SCC

Reproduced from NHMRC Guidelines 2005.

Health state approximations

Health states had to be approximated due to lack of appropriate data (Table 76).

Table 76 Approximations for model health states.

Data source	Model health state	Approximation to gold standard definition
ALTS	Normal	HCII negative; absence of CIN 2 or worse histology
	HPV and CIN 1	For each test result (excluding negative) we assigned proportions into normal, CIN 2, and CIN 3+ health states. The remaining unassigned proportion was divided between HPV and CIN 1 health states according to a proportional breakdown based on Australian cytology and histology correlation data
VCCR	HPV	HPV histology
	Normal	Normal histology
Coste	Normal and HPV	After assigning proportions into histology categories for CIN 1, CIN 2, and CIN 3 or worse, the remainder was split between normal and HPV states. The proportion assigned to the HPV health state is based on the proportion with CIN 1, according to the Australian histology and cytology data (MSAC 2002a), and the remainder is assumed to be health state normal

Australian cytology and histology correlation data

We use data collected for a previous MSAC evaluation (MSAC 2002a) to estimate remaining unknown parameters. These data cross-tabulate histology and cytology for all women with histology performed within 6 months of an abnormal cytology result. Data for 1999 were provided from all States and Territories except NSW and Queensland.

We used the correlation data to estimate the proportional breakdown into CIN 1 and HPV health states for all abnormal cytology results. Where these data have been used, histological HPV is a proxy for HPV health status.

Methods

We found it difficult to find sources of data on a screening population that were consistent with data from meta-analyses. A summary of our methods for combining data sources is presented in Table 77. Details are provided in supplementary material to this appendix on page 164. In many cases it was impossible to obtain a coherent matrix: sensitivity estimates used to derive parameters for negative cytology resulted in meaningless negative estimates. This indicates that the characteristics of the test in a

study may be different from those of a test used in routine practice, or that the population in which it was being used may not be representative of a screening population. The second last column of Table 77 indicates whether or not the specific method generated a coherent matrix. The final column indicates whether or not the matrix was considered a candidate for sensitivity analysis. We have also used the matrices reported in the Applicant's submission to MSAC in our sensitivity analysis.

The matrix obtained via Method 6(b) was initially chosen as the baseline matrix, but this matrix was calibrated in a parallel MSAC assessment (MSAC reference 39: HPV triage), so we used the calibrated version for consistency between the two reports.

Of the matrices in Table 77, Method 6(b) combined the sources of data with the best level of verification against an appropriate reference standard. Of our sources of data, Coste (2003) had the best follow-up for women with negative cytology; ALTS (Sherman et al. 2006) had the best follow-up of women with low-grade cytology; and we expect that Australian registry data will be the most representative of women with high-grade cytology in Australia, for which follow-up is routine. We used data from the age group 25–29 years to be consistent with the ALTS population, but when this is combined with cytology reporting rates in this age group, the resulting test probability matrix is age-independent. The NSW screening rates for women aged 25–29 years produce very similar operating characteristics to VCCR rates, but NSW rates produce a test probability matrix with sensitivities closer to the published estimates. We included the matrix produced by applying Victorian rates in a sensitivity analysis of the model. This method does not require us to use lower estimates from any data source.

Estimating negative cytology results via Coste (2003) avoids the difficulties in applying sensitivity estimates from Arbyn et al. (2008). Methods derived from sensitivity estimates from Arbyn et al. at a pHSIL threshold often produced matrices with unrealistically high sensitivity (100.00%). We expect that this is largely because of differences in the operating characteristics of cytology when different populations are studied: since Arbyn et al. consisted largely of studies based on women referred for further investigation (eg, colposcopy), it may not be appropriate to apply the sensitivity estimates reported in the meta-analysis to a test probability matrix which must apply to a screening population.

Points on the ROC graph generated by Method 6 are consistent with Figure 2 in Davey et al. (2006) for screening populations. As can be expected, points on the ROC from Arbyn et al. are more consistent with Figure 2 in Davey et al. for studies with populations of women referred for investigation. The specificity of cytology is much higher in a screening population than in a population referred for further investigation.

Table 77 Data sources for various methods of deriving test probability matrices for CC or LBC.

Method	Data sources							
	pLSIL	dLSIL	pHSIL	dHSIL	Negative	Distribution of cytology	Coherent matrix?	Used in sensitivity analysis?
1 (CC)	Arbyn + Coste	Arbyn + Coste	Arbyn + Coste	Arbyn + Coste	Arbyn + Coste	n/a	Yes	No—not representative
2(a) (CC)	Coste	Coste	Coste	Coste	Coste	Coste	Yes	No—no HPV health state
2(b) (CC)	Coste	Coste	Coste	Coste	Coste	NSW (all ages)	Yes	No—no HPV health state
2(c) (CC)	Coste	Coste	Coste	Coste	Coste	VCCR (all ages)	Yes	No—no HPV health state
3(a) (LBC)	ALTS	ALTS	ALTS	ALTS	CADTH; WHINURS	NSW 2006 Q4 25–29 y	Yes	No—inconsistent
3(b) (CC)	ALTS	ALTS	ALTS	ALTS	Via 2(a) (LBC) disease distribution	NSW 2006 Q4 25–29 y	Yes	No—inconsistent
4(a) (CC)	ALTS	ALTS	ALTS	ALTS	Arbyn; Nanda; WHINURS	NSW 2006 Q4 25–29 y	No; no with Arbyn lower estimate	–
4(b) (CC)	ALTS, lower estimate of CI	ALTS, lower estimate of CI	ALTS	ALTS	Arbyn; Nanda; WHINURS	NSW 2006 Q4 25–29 y	No; yes with Arbyn lower estimate	No—see Supplementary Material
4(c) (CC)	ALTS, lower estimate of CI	ALTS, lower estimate of CI	ALTS, upper estimate of CI	ALTS, upper estimate of CI	Arbyn; Nanda; WHINURS	NSW 2006 Q4 25–29 y	No; yes with Arbyn lower estimate	No—see Supplementary Material
5(a) (CC)	ALTS	ALTS	VCCR 25–29 y	VCCR 25–29 y	Arbyn; Nanda; WHINURS	VCCR 25–29 y	No; no with Arbyn lower estimate	-
5(b) (CC)	ALTS	ALTS	VCCR 25–29 y	VCCR 25–29 y	Arbyn; Nanda; WHINURS	NSW 2006 Q4 25–29 y	No; yes with Arbyn lower estimate	Yes, with lower Arbyn estimate
5(c) (CC)	ALTS, lower estimate of CI	ALTS, lower estimate of CI	VCCR 25–29 y	VCCR 25–29 y	Arbyn; Nanda; WHINURS	VCCR 25–29 y	No; yes with Arbyn lower estimate	No—see Supplementary Material
5(d) (CC)	ALTS, lower estimate of CI	ALTS, lower estimate of CI	VCCR 25–29 y	VCCR 25–29 y	Arbyn; Nanda; WHINURS	NSW 2006 Q4 25–29 y	Yes; yes with Arbyn lower estimate	No—see Supplementary Material
6(a) (CC)	ALTS	ALTS	VCCR 25–29 y	VCCR 25–29 y	Coste; WHINURS	VCCR 25–29 y	Yes	Yes
6(b) (CC)	ALTS	ALTS	VCCR 25–29 y	VCCR 25–29 y	Coste; WHINURS	NSW 2006 Q4 25–29 y	Yes	Yes
6(c) (CC)	ALTS, lower estimate of CI	ALTS, lower estimate of CI	VCCR 25–29 y	VCCR 25–29 y	Coste; WHINURS	VCCR 25–29 y	Yes	No—see Supplementary Material
6(d) (CC)	ALTS, lower estimate of CI	ALTS, lower estimate of CI	VCCR 25–29 y	VCCR 25–29 y	Coste; WHINURS	NSW 2006 Q4 25–29 y	Yes	No—see Supplementary Material
7(a) (CC)	VCCR (all ages)	VCCR (all ages)	VCCR (all ages)	VCCR (all ages)	VCCR (all ages)	VCCR (all ages)	Yes	No—not representative
7(b) (CC)	VCCR (all ages)	VCCR (all ages)	VCCR (all ages)	VCCR (all ages)	Arbyn; Nanda; WHINURS	VCCR (all ages)	No; no with Arbyn lower estimate	-
8 (CC)	Cytyc	Cytyc	Cytyc	Cytyc	Cytyc	–	Yes	Yes
8 (auto LBC)	Cytyc	Cytyc	Cytyc	Cytyc	Cytyc	–	Yes	Yes

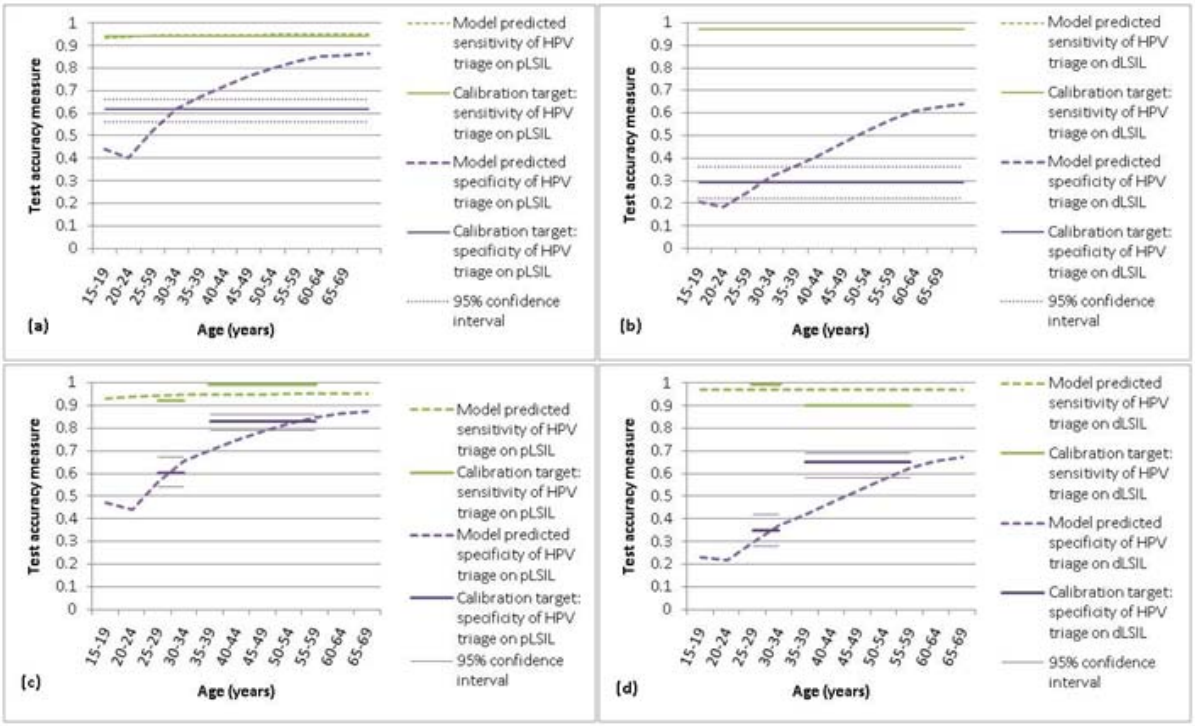
Calibration of baseline matrix for conventional cytology

The matrix of test characteristics for conventional cytology obtained via Method 6(b) was calibrated for HPV triage of pLSIL and dLSIL cytology. Calibration targets were the

sensitivity and specificity for detection of CIN 2+ of triage of pLSIL and dLSIL cytology by HPV testing, and age-specific HPV test positivity rates in women with pLSIL and dLSIL cytology (Arbyn et al. 2006; Ronco et al. 2007b; Sherman et al. 2002). Each target varied depending on whether conventional cytology or LBC was used. There were no sources of information on HPV triage of automated cytology from which to obtain calibration targets.

Differences in the test characteristics of conventional cytology and LBC produce differences in the test accuracy of HPV triage depending on whether conventional cytology or LBC is used, and whether triage is of pLSIL or dLSIL cytology. Accordingly, calibration of the model took into account reported values in each of these categories. Graphs comparing these calibration targets and model predictions using calibrated matrices are presented in Figure 20 and Figure 21. It was not necessary to calibrate the test probability matrix for LBC once conventional cytology was calibrated: the method described below for derivation of the LBC matrices generated results in close agreement with reported values.

Figure 20 Age-specific test accuracy (sensitivity and specificity) of HPV triage for detection of CIN 2+ in women with pLSIL or dLSIL cytology.

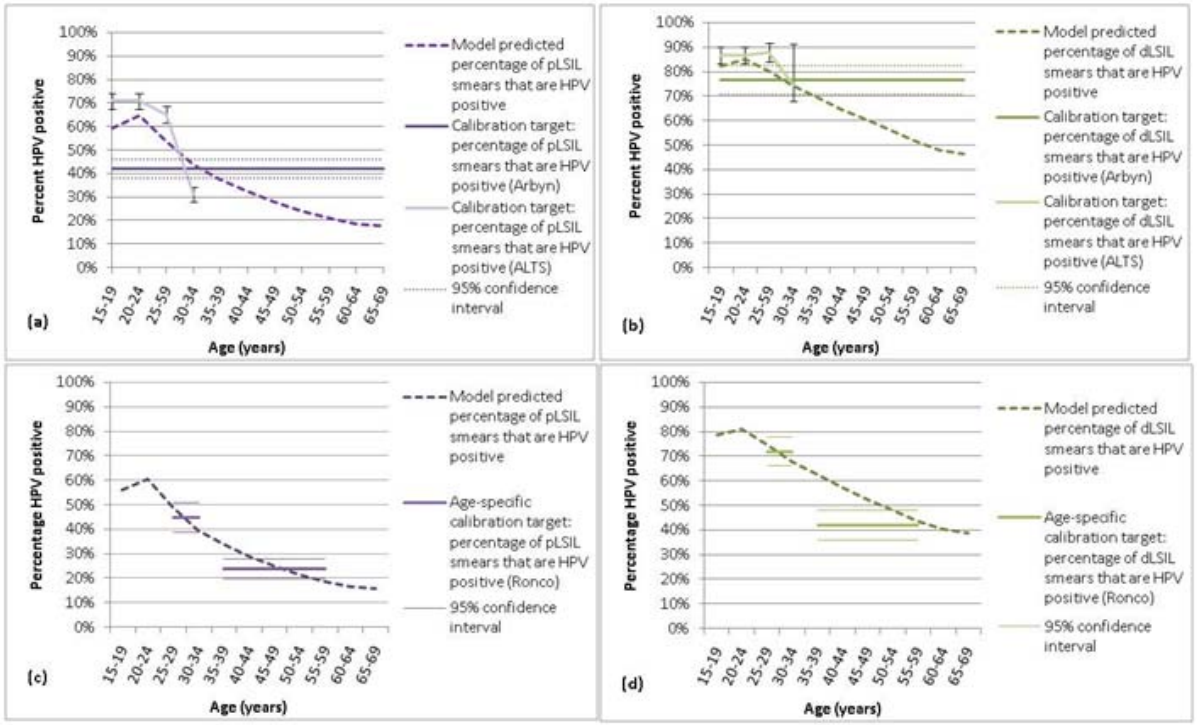


(a) Triage of pLSIL as determined by conventional cytology; (b) triage of dLSIL as determined by conventional cytology; (c) triage of pLSIL as determined by LBC; (d) triage of dLSIL as determined by LBC. Note that data from Arbyn et al. are not age-specific; data from Ronco et al. are limited to two age groups. CIs for sensitivity are not shown for clarity.

Model outputs for the sensitivity of HPV triage agree closely with reported values in the literature (Arbyn et al. 2006), in both pLSIL smears (Figure 20 a) and dLSIL smears (Figure 20 b) as determined by conventional cytology. The corresponding model outputs for pLSIL (Figure 20 c) and dLSIL (Figure 20 d) as determined by LBC are consistent with reported values (Ronco et al. 2007b), although the extremely wide CIs (not shown on graph) diminish the worth of these values as calibration targets. There is very little age dependence in the sensitivity of HPV triage in the model.

On the other hand, the specificity of HPV triage varies highly with age in the model. Unfortunately, age-specific targets in the literature were limited. We aimed for consistency with the available information as far as was possible.

Figure 21 Age-specific percentages of women testing positive in HPV triage for pLSIL and dLSIL cytology.



(a) Triage of pLSIL as determined by CC; (b) triage of dLSIL as determined by CC; (c) triage of pLSIL as determined by LBC; (d) triage of dLSIL as determined by LBC. Note that in (a) and (b), the age-specific calibration target derived from ALTS is approximate owing to the difference in age groupings. The calibration target from Arbyn et al. is not age-specific. The calibration target from Ronco et al. is limited to two age groups.

Model outputs for the percentage of women who test positive at HPV triage are highly age-specific. Again, age-specific targets in the literature were limited, making complete calibration impossible. Model outputs reflect the distribution of reported values in the literature, in both pLSIL (Figure 21 a) and dLSIL (Figure 21 b) as determined by CC, and in both pLSIL (Figure 21 c) and dLSIL (Figure 21 d) as determined by LBC.

Derivation of LBC and automated LBC matrices

The baseline matrix for CC, calibrated for HPV triage, was adapted to be appropriate for LBC and automated LBC by adjusting the cytology test yields while keeping constant the prevalence of disease implied by the matrix. We adjusted the matrix at the stage of derivation where age-specific data were used, and therefore needed to estimate call rates in the age group 25–29 years. Registry data were not a reliable source for estimating LBC test yields, as type of test is not systematically recorded, most LBC tests are split samples, and the test result is given as the worst of both tests, without recording which test was used for the result.

We estimated call rates for automated LBC relative to CC using data from Table 2 in Davey et al. 2007a (see also Table 19), and then we applied these ratios to the age-specific call rates for conventional cytology.

We estimated call rates for LBC relative to CC similarly, based on the difference in total slide classifications as reported in the earlier LBC review section 'Test yield', on page 33. There was no evidence of an increase in HSIL+, so the ratios for pHSIL and dHSIL were set to 1. The increase in LSIL+ ranged from 0.66 to 1.3 per cent. As there was no evidence of an increase in HSIL+, this increase was assumed to be entirely due to an increase in dLSIL test yield. The absolute increase in dLSIL rates was estimated as 1 per cent (mid-range), which was then applied to the NSW PTR rates in 2006 for women of all ages (Table 74) to calculate the relative dLSIL yield in Table 78. The absolute increase in dLSIL results predicted by the model was 0.87 per cent, as the predicted dLSIL yield by CC was lower than the NSW value of 2.2 per cent. This value is still within the range discussed on page 33. We assumed no increased test yield of pLSIL, on the basis of the mixed evidence presented in Davey et al. 2006. This is a favourable assumption.

Table 78 Relative cytology test yields for automated LBC and LBC compared with conventional cytology.

Relative test yields	Negative	pLSIL	dLSIL	pHSIL	dHSIL
Automated LBC : CC	0.99	1.08	1.56	0.78	1.27
LBC : CC	0.99	1.00[a]	1.45	1.00[b]	1.00[b]

[a] Assumption—favourable to LBC, made in light of mixed evidence. Consistent with Davey et al. 2006.
[b] Assumption, based on review results (see page 33).

We adjusted the test yields in the CC model according to these ratios to estimate test yields for automated LBC and LBC. Applying ratios to each test result individually and then summing over all test results gives a total percentage of slightly over 100 per cent. We adjusted each call rate proportionally so the total was 100 per cent. Test yields varied with choice of CC test probability matrix in the sensitivity analysis and were recalculated in each scenario where the CC test probability matrix varied from the baseline matrix.

Table 79 Estimated test yields for automated LBC and LBC in women aged 25–29 years, baseline values.

Estimated test yields	Negative	pLSIL	dLSIL	pHSIL	dHSIL
Automated LBC	89.9%	3.0%	4.5%	0.6%	1.9%
LBC	90.5%	3.1%	4.1%	0.8%	1.5%

The estimates for test yields in the age group 25–29 years for each cytology type were calibrated in order to predict correctly the relative test yields over all ages (Table 79).

For automated LBC, we adjusted the intermediate matrix for CC by applying these test yields and adjusted the values of the cells proportionally to keep the disease distribution constant. We firstly adjusted the values in cells for high-grade cytology results in CIN 2+ women so that the relative true and false positive rates for automation-assisted slide reading, compared toto those for manual reading of conventional cytology and LBC, would be consistent with the estimates in Table 14 and Table 17. Following this adjustment, we:

1. split the excess or deficient disease distribution for health state CIN 2+ between all remaining cytology test results; and proportionally increased or decreased the

pHSIL and dHSIL parameters for health states Normal, HPV and CIN 1 so that test yields for pHSIL and dHSIL were met

2. split the excess or deficient disease distribution for health state CIN 1 between cytology test results Negative, pLSIL and dLSIL; and proportionally increased or decreased the pLSIL and dLSIL parameters for health states Normal and HPV so that test yields for pLSIL and dLSIL were met

3. increased or decreased negative cytology parameters for health state Normal so as to preserve the disease distribution in this health state.

For LBC, we adjusted cytology results for women with CIN 2+ health state so that relative sensitivity between CC and LBC as reported in Arbyn et al. {Arbyn, 2008 4899 /id} was conserved. We then followed steps 1 to 3 as for automated cytology. When running sensitivity analyses of test characteristics, we recalculated matrices for automated LBC and LBC for each CC matrix, with the exception of the Applicant's test probabilities, for which both were provided.

Supplementary material

This supplementary material elaborates on the methods used to derive test probability matrices for conventional cytology as summarised in Table 77.

Method 1

The first method attempted to derive a test probability matrix from estimates of sensitivity and specificity at different thresholds. The sensitivities and specificities reported by Arbyn et al. at cytological thresholds of ASCUS, LSIL and HSIL were expanded into a matrix of test probabilities with two health state sets: CIN 2 or worse (which included CIN 2 and CIN 3+) and CIN 1 or less (which included Normal, HPV and CIN 1). The breakdown followed the proportions from the French trial (Coste 2003). The breakdown of likely cytology results for HPV health state is based on results for CIN 1 with proportions according to the Australian histology/cytology correlation data (MSAC 2002a).

Both Arbyn et al. and Coste use TBS1991 terminology, and additional assumptions would need to be made regarding the breakdown into pLSIL and pHSIL from TBS1991 ASCUS reported results. Before making any further assumptions, we derived the test probability matrix shown in Table 80.

Table 80 Partial test probability matrix derived from Arbyn et al.'s meta-analysis and additional assumptions (%).

Test result/ health state	Negative	pLSIL & pHSIL [a]	dLSIL	dHSIL	*Total*
Normal	77.24	10.09	10.49	2.18	*100*
HPV	71.84	6.12	19.06	2.99	*100*
CIN 1	12.94	15.64	56.70	14.72	*100*
CIN 2+	11.80	12.60	20.40	55.20	*100*

[a] Both reported as ASCUS in TBS1991.

For a test with these characteristics there is a unique underlying disease distribution which will produce a target cytology distribution: multiplying the inverse of this test probability matrix by the target distribution gives the underlying disease distribution. If we set the target test yield distribution as either the NSW PTR 2006 Q4 all ages distribution (as in Method 2(c)) or the VCCR data all ages distribution (as in Method 2(b)), then an impossible disease distribution is implied, involving negative 'proportions' of the population, or 'proportions' greater than 1 (Table 81).

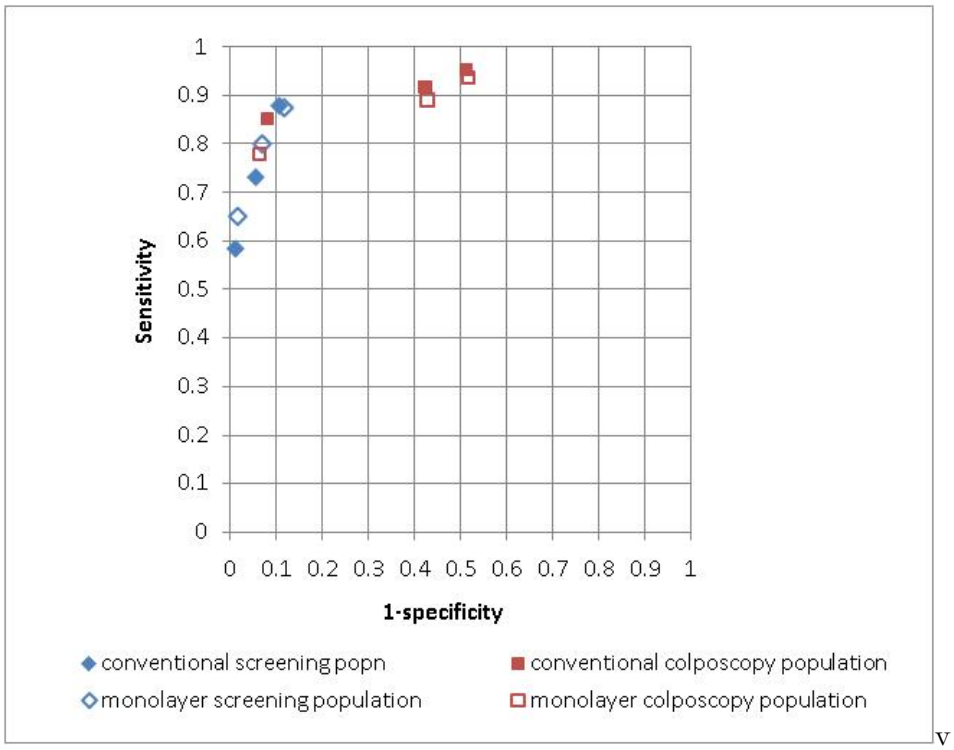
Table 81 Implied underlying disease distribution by various derivation methods.

Distribution of disease/histology in the screening population	Negative	HPV	CIN 1	CIN 2+
Proportion of screening population; derived from test yields applied in Method 2(c)	0.05	1.32	–0.42	0.05
Proportion of screening population; derived from test yields applied in Method 2(b)	0.16	1.19	–0.41	0.06
Proportion of screening population; derived from Method 2(a) prevalence estimates (all ages)	0.795	0.147	0.035	0.023

These 'proportions' indicate that it is difficult to reconcile the sensitivities and specificities estimated by Arbyn et al. with the operating characteristics of screening population-based cytology. This may be because of differences in the populations included in the meta-analysis. SROCs of optimised interpretation data in Coste 2003 and our transformation of these data indicate that screening population studies operate in a consistently different region from studies of participants referred for investigation (usually colposcopy), as shown in Figure 22

Figure 22. A recent review (Davey et al. 2006) presents information consistent with this observed difference in cytology test performance.

Figure 22 Test characteristics associated with cytology in different populations (screening and colposcopy-referred populations).



Source: Coste; results of optimised interpretation.

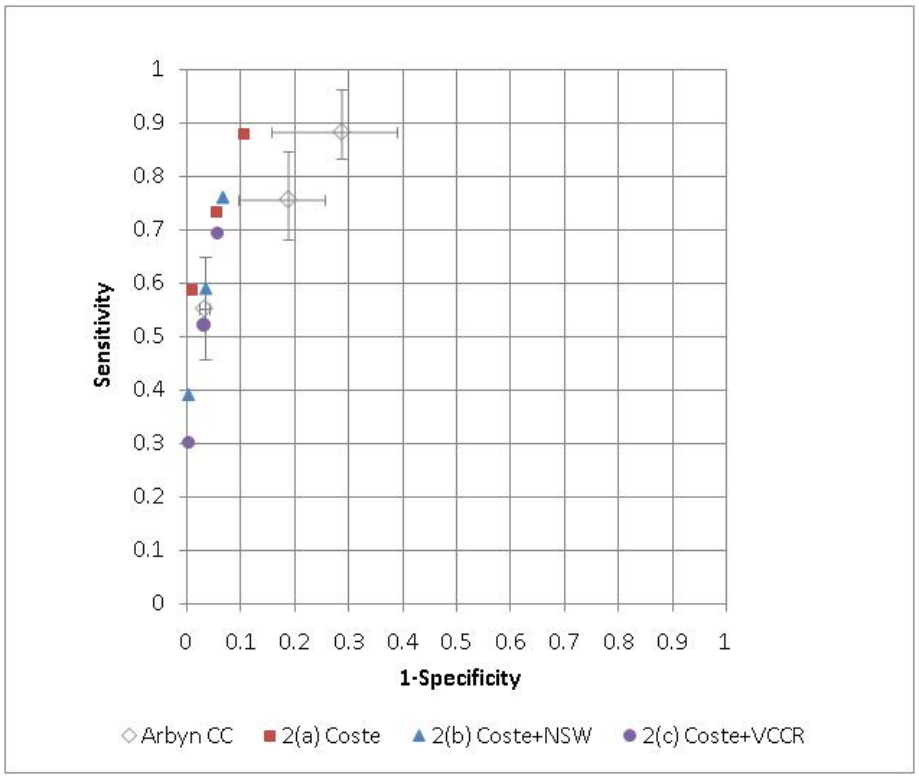
When we used the disease distribution in the screening population implied by Coste 2003 (Method 2(a)) with HPV prevalence in histologically normal women in the same study (de Cremoux et al. 2003), the proportion of abnormal cytology results was unrealistically high (27% across all ages).

Matrices derived via Method 1 were not used in sensitivity analyses.

Method 2

Method 2 uses data from Coste where a reference standard of colposcopy and colposcopically directed histology is available for all cytology results. Applying Australian cytology reporting rates to the Coste data according to Methods 2(b) and 2(c) results in lower sensitivity than the Coste data alone (Method 2(a)) (Figure 23). This indicates that the lower estimate of CI in Arbyn et al. for the sensitivity of CC for detection of CIN 2+ at HSIL threshold may be more appropriate as an estimate of cytology sensitivity in Australia.

Figure 23 Summary ROCs obtained via Method 2, compared to that described in Arbyn et al.



Method 3

Method 3 is based on ALTS data, and was an attempt to derive a test probability matrix for LBC directly. Health state breakdown for women with negative cytology is approximated via published sensitivity estimates for LBC (Krahn et al. 2008).

These estimates for LBC weighted by the distribution of cytology reports in NSW (PTR, fourth quarter of 2006, women aged 25–29 years) imply an estimate of the distribution of disease in the population and provide a method for deriving the test probability matrix for CC by assuming the same disease distribution and deriving false negative results secondarily. Note that these data on distribution of cytology reports are for all cytology, but are dominated by conventional. Using this estimate for LBC is inaccurate since LBC is known to increase low-grade cytology reports relative to conventional cytology.

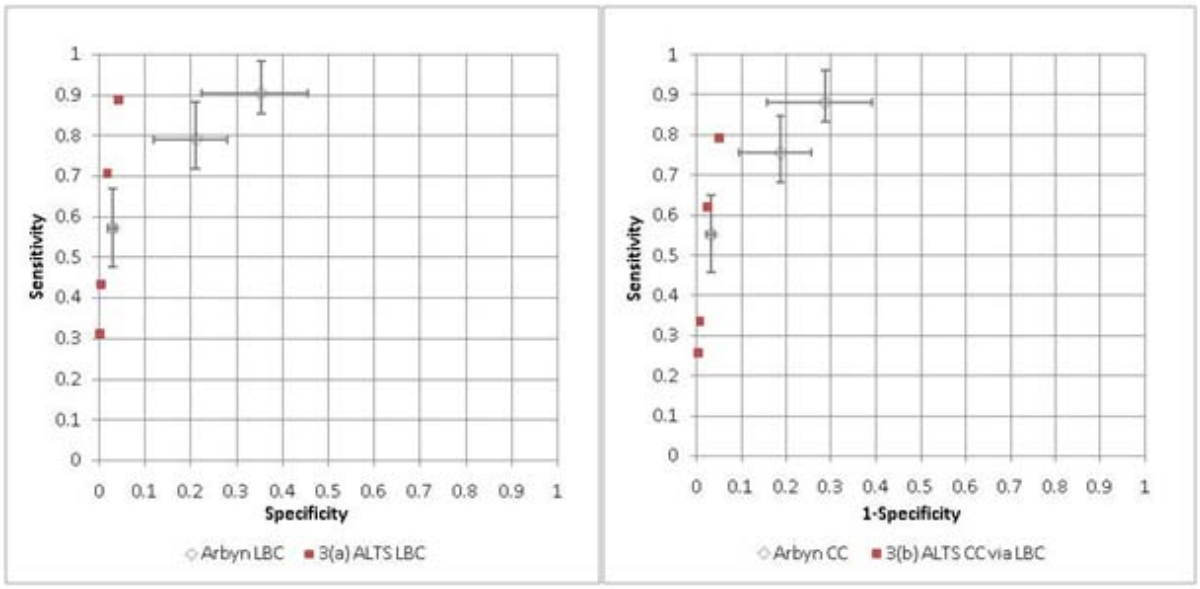
We calculated sensitivity and specificity at various thresholds for CC and LBC, and then compared these with published estimates (Arbyn et al. 2008) at corresponding thresholds.

Table 82 Absolute and relative sensitivity and specificity for detection of CIN 2+ for cytology test characteristics based on Method 3.

	Sensitivity				Specificity			
	dHSIL+	pHSIL+	dLSIL+	pLSIL+	dHSIL+	pHSIL+	dLSIL+	pLSIL+
CC	0.26	0.34	0.66	0.79	0.99	0.98	0.95	0.92
LBC	0.31	0.43	0.73	0.89	0.99	0.99	0.96	0.92
LBC/CC	1.21	1.29	1.10	1.12	1.002	1.004	1.004	1.004

Even allowing for differences in nomenclature systems, these estimates for relative sensitivity are outside the range of CIs given in estimates by Arbyn et al. These results in relation to Arbyn et al.'s meta-analysis are presented on an SROC curve (Figure 24).

Figure 24 SROC obtained via Method 3, compared to that described in Arbyn et al.



Moreover, it proved impossible to apply Arbyn et al.'s sensitivity estimates in place of CADTH estimates with these data.

Due to the difficulty involved in deriving test probability matrices for CC and LBC with accurate absolute sensitivity and specificity, we derived matrices for LBC based on published relative differences once a matrix for CC had been chosen. This general method ensured that relative differences as published in Arbyn et al. were preserved.

Methods 4–6

Method 4 used ALTS data as per Method 3(b), with the exception that the disease distribution in women with negative cytology was parameterised via published sensitivity estimates (Figure 25). Methods 5 and 6 used ALTS data only for low-grade cytology, and used routinely collected Victorian data on histology of high-grade cytology {Victorian Cervical Cytology Registry, 2006 4976 /id} (Figure 26).

Methods 4 and 5 proved impractical unless lower estimates of the ALTS data (Sherman et al. 2006) or of Arbyn et al.'s sensitivity data (Arbyn et al. 2008) were applied. The results from Methods 1–4 suggest that it is reasonable to lower the estimates from Arbyn et al. rather than those from ALTS. Only those methods using unaltered ALTS data were considered as candidates for sensitivity analysis.

Method 6 uses estimates for negative cytology results derived from Coste (2003) for histologically defined health states and WHINURS for the HPV health state. No assumptions are made about the sensitivity of cytology; this avoids applying sensitivity estimates that are not based on screening populations to the test probability matrix, and avoids making assumptions about the applicability of the lower estimates of sensitivity in Arbyn et al. The points on the SROC graph from Method 6 are consistent with the ROC curves for cytology based on a screening population implied by Davey et al. (2006).

Figure 25 SROC for Method 4, compared to that described in Arbyn et al.

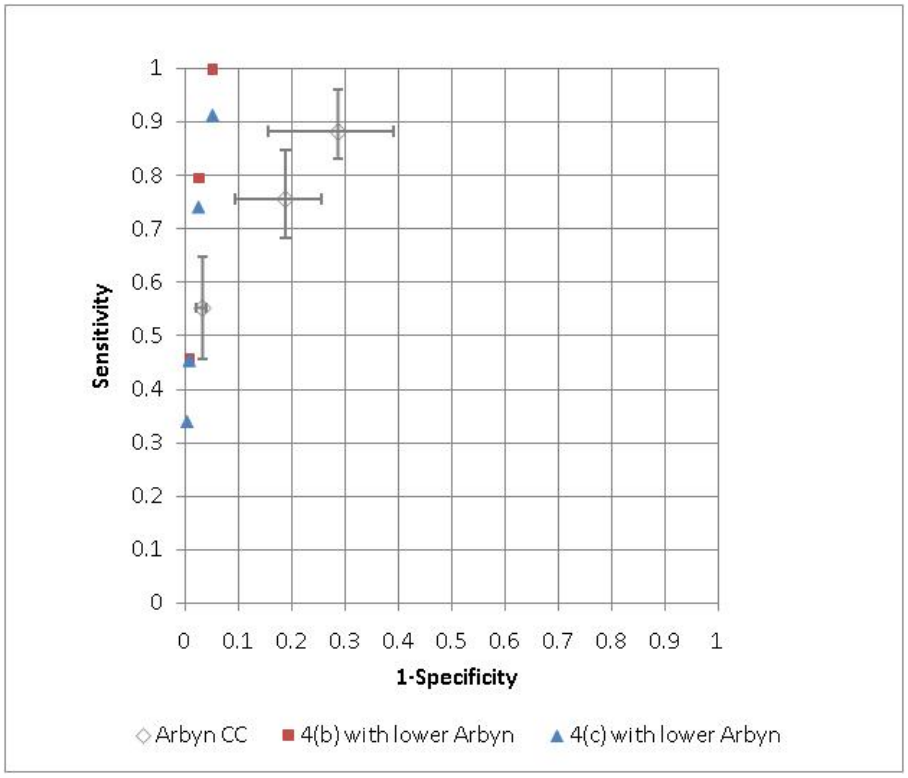
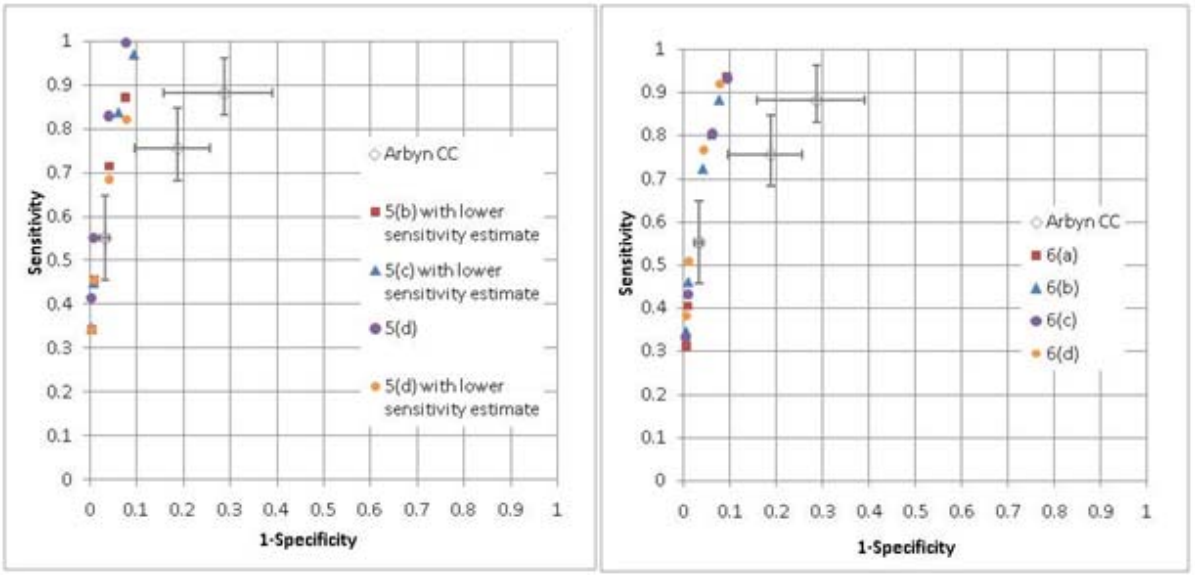


Figure 26 SROC for Methods 5 & 6 compared to that described in Arbyn et al.



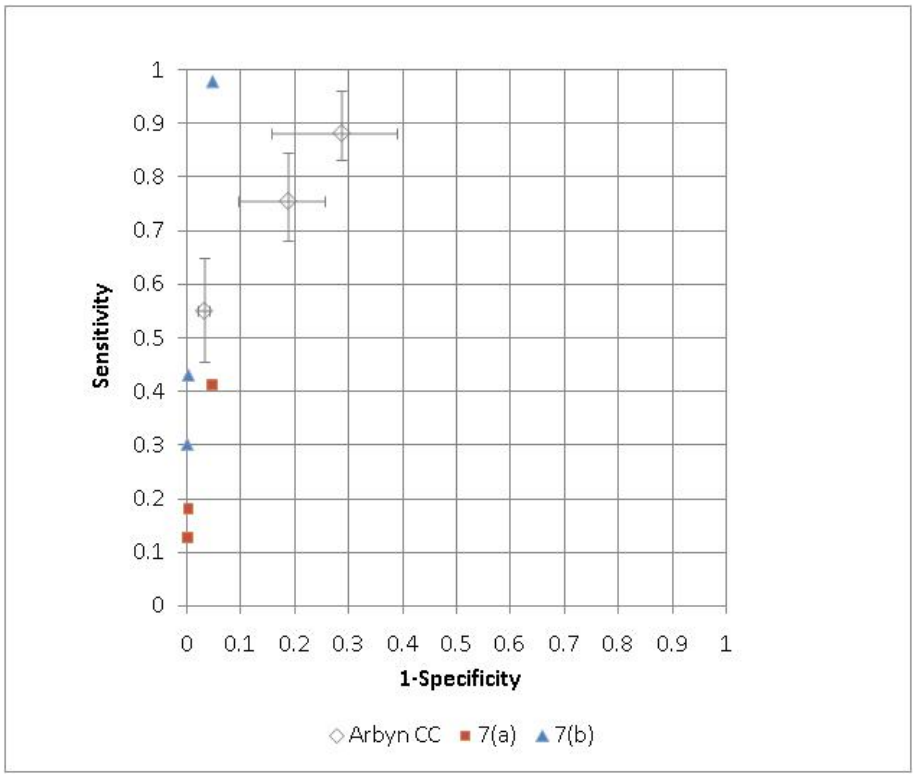
Method 7

Method 7(a) used data entirely from the VCCR; Method 7(b) used published sensitivity estimates {Nanda, 2000 933 /id}to parameterise negative cytology (Figure).

The VCCR data (VCCR 2006) give information on histology in women with negative smears, but this is unlikely to reflect women with negative smears generally. We also estimated the histology breakdown for women with negative cytology via sensitivity

estimates. VCCR data are available for histology in women with low-grade cytology, but as low-grade cytology is not always routinely followed up by at least colposcopy, these results are unlikely to be generally applicable. Matrices derived via Method 7 were not considered to be representative of cytology accuracy in a screening population and were therefore not considered suitable to use in sensitivity analyses.

Figure 25 SROC for Method 7 compared to that described in Arbyn et al.



Method 8

The test probability matrices used in the Applicant's cost-effectiveness model were included in Table 77 because they were used in sensitivity analyses.

Abbreviations

AGUS	Atypical Glandular Cells of Undetermined Significance
AIHW	Australian Institute of Health and Welfare
AIS	Adenocarcinoma *in situ*
ALTS	The ASCUS–LSIL Triage Study Group
AMBS	Australian Modified Bethesda System
ASC-H	Atypical squamous cells, possible high-grade lesion
ASCUS	Atypical squamous cells, undetermined significance
CADTH	Canadian Agency for Drugs and Technologies in Health
CC	Conventional cytology
CCR	NSW Central Cancer Registry
CI	Confidence interval
CIN	Cervical intraepithelial neoplasia
DHM	Douglass Hanly Moir Laboratories
DOR	Diagnostic odds ratio
DRG	Diagnosis-related group
FIGO	International Federation of Gynecology and Obstetrics
FN	False negative
FOV	Field of view
FP	False positive
HGEA	High-grade epithelial abnormality
HPV	Human papilloma virus
HSIL	High-grade squamous intraepithelial lesion
HTA	Health technology assessment
ICER	Incremental cost-effectiveness ratio
LBC	Liquid-based cytology
LEEP	Loop electrosurgical excision procedure

LGEA	Low-grade epithelial abnormality
LSIL	Low-grade squamous intraepithelial lesion
LYS	Life year saved
MBS	Medicare Benefits Schedule
MRI	Magnetic resonance imaging
MSAC	Medical Services Advisory Committee
NFR	Not for review
NHMRC	National Health and Medical Research Council
NR	Not reported
OR	Odds ratio
PBS	Pharmaceutical Benefits Scheme
PCR	Polymerase chain reaction
PET	Positron emission tomography
pHSIL	Possible high-grade squamous intraepithelial lesion
pLSIL	Possible low-grade squamous intraepithelial lesion
PTR	Pap Test Registry
QALY	Quality-adjusted life year
QC	Quality control
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
RCT	Randomised controlled trial
ROC	Receiver operating characteristic
SCC	Squamous cell carcinoma
SIL	Squamous intraepithelial lesion
SROC	Summary receiver operating characteristic
TN	True negative
TP	True positive
TPI	ThinPrep Imager

TPM	ThinPrep Manual
TZ	Transformation zone
US	United States
VCCR	Victorian Cervical Cytology Register
WHINURS	Women, Human Papillomavirus Prevalence, Indigenous, Non-Indigenous, Urban, Rural Study

References

ABS 2002. *National Health Survey, Summary of Results, Australia, 2001*. Canberra: Australian Bureau of Statistics.

ABS 2006. *National Health Survey, Summary of Results, Australia, 2004–05*. Canberra: Australian Bureau of Statistics.

ABS 2007. *Deaths, Australia, 2006*. Canberra: Australian Bureau of Statistics.

ABS and AIHW 2005. *The Health and Welfare of Australia's Aboriginal and Torres Strait Islander Peoples*. ABS Cat. No. 4704.0; AIHW Cat. No. IHW14. Canberra: Commonwealth of Australia.

AIHW 2005. *Cervical screening in Australia 2002–2003*. Canberra: AIHW: Australian Institute of Health and Welfare.

AIHW 2007a. *ACIM (Australian Cancer Incidence and Mortality) Books*. Canberra: Australian Institute of Health and Welfare.

AIHW 2007b, *Cervical screening in Australia 2004–2005*. AIHW cat. no. Can 33.Canberra: AIHW: Australian Institute of Health and Welfare.

AIHW 2008, *Cervical screening in Australia 2005–2006*. AIHW cat. no. Can 36.Canberra: AIHW: Australian Institute of Health and Welfare.

AIHW and AACR 2008, *Cancer in Australia: an overview, 2008*. AIHW cat no CAN 42.Canberra: AIHW: Australian Institute of Health and Welfare & Australasian Association of Cancer Registries.

Anton, R.C., Ramzy, I. et al, (2001). 'Should the cytologic diagnosis of "atypical squamous cells of undetermined significance" be qualified? An assessment including comparison between conventional and liquid-based technologies', *Cancer*, 93, 93-99.

Anttila, A, Hakama, M et al. 2006. Alternative technologies in cervical cancer screening: a randomised evaluation trial, *BMC Public Health*, 6 (-), 252.

Arbyn, M M, Bergeron, C et al. 2008. Liquid compared with conventional cervical cytology: a systematic review and meta-analysis, *Obstet. Gynecol.*, 111 (1), 167–177.

Arbyn, M, Sasieni, P et al. 2006. Clinical applications of HPV testing: a summary of meta-analyses, *Vaccine*, 24 (S3), 78–89.

Australian Government Department of Health and Ageing 2007. *The National HPV Vaccination Program* [Internet]. Available from: Department of Health and Ageing's website [Accessed 11 July 2007].

Australian Government Department of Health and Ageing and National Health and Medical Research Council 2008. *The Australian Immunisation Handbook*. Canberra: Australian Government.

Australian Health Technology Advisory Committee 1998. *Review of automated and semi-automated cervical screening devices*. Canberra: Commonwealth of Australia.

Bai, H., Sung, C.J. & Steinhoff, M.M. (2000). 'ThinPrep Pap Test promotes detection of glandular lesions of the endocervix', *Diagnostic Cytopathology*, 23, 19-22.

Baldauf, JJ, Dreyfus, M et al. 1998. Cytology and colposcopy after loop electrosurgical excision: implications for follow-up, *Obstet. Gynecol.*, 92 (1), 124–130.

Bergeron, C., Bishop, J. et al, (2001). 'Accuracy of thin-layer cytology in patients undergoingcervical cone biopsy', *Acta Cytologica*, 45, 519-524.

Bernstein, S.J., Sanchez-Ramos, L. & Ndubisi, B. (2001). 'Liquid-based cervical cytologic smearstudy and conventional Papanicolaou smears: a metaanalysis of prospective studies comparing cytologic diagnosis and sample adequacy', *American Journal of Obstetrics & Gynecology*, 185, 308-317.

Bigrigg, A, Haffenden, DK et al. 1994. Efficacy and safety of large-loop excision of the transformation zone, *Lancet*, 343 (8888), 32–34.

Biscotti, CV, Dawson, AE et al. 2005. Assisted primary screening using the automated ThinPrep Imaging System, *American Journal of Clinical Pathology*, 123 (2), 281–287.

Blamey, S, Jackson, T et al. 2006. Re: Liquid-based cytology for cervical screening [7], *Australian and New Zealand Journal of Obstetrics and Gynaecology*, 46 (1), 70–71.

Bolger, N, Heffron, C et al. 2006. Implementation and evaluation of a new automated interactive image analysis system, *Acta Cytologica.*, 50 (5), 483–491.

Bosch, FX, Qiao, Y et al. 2006. The epidemiology of human papillomavirus infection and its association with cervical cancer, *International Journal of Gynecology and Obstetrics*, 94 (Supplement 1), S8-S21.

Bossuyt, PM, Irwig, L et al. 2006. Comparative accuracy: assessing new tests against existing diagnostic pathways, *BMJ*, 332 (7549), 1089–1092.

Britt H, Miller GC, Charles J, et al. 2008.General practice activity in Australia 2006–07.AIHW Cat No GEP 21. Canberra: Australian Institute of Health and Welfare.

Brown, AD and Garber, AM 27-1-1999. Cost-effectiveness of 3 methods to enhance the sensitivity of Papanicolaou testing, *JAMA.*, 281 (4), 347–353.

Burchell, AN, Winer, RL et al. 2006. Chapter 6: Epidemiology and transmission dynamics of genital HPV infection, *Vaccine*, 24 Suppl 3 (S52-S61.

Canfell, K, Barnabas, R et al. 2004. The predicted effect of changes in cervical screening practice in the UK: results from a modelling study, *Br J Cancer*, 91 (3), 530–536.

Canfell, K, Clements, M et al. 2008, *Cost-effectiveness of proposed changes to the National Cervical Screening Program.*

Castle, PE, Stoler, MH et al. 2007. The relationship of community biopsy-diagnosed cervical intraepithelial neoplasia grade 2 to the quality control pathology-reviewed diagnoses: an ALTS report, *Am.J.Clin.Pathol.*, 127 (5), 805–815.

Cecchini, S, Visioli, CB et al. 2002. Recurrence after treatment by loop electrosurgical excision procedure (LEEP) of high-grade cervical intraepithelial neoplasia, *Tumori*, 88 (6), 478–480.

Celik, C, Gezginc, K et al. 2008. A comparison of liquid-based cytology with conventional cytology, *International Journal of Gynecology and Obstetrics*, 100 (2), 163–166.

Centre for Reviews and Dissemination, 2001. *Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews*, NHS Centre for Reviews and Dissemination, York.

Cibas ES,et al. The MonoPrep Pap Test for the detection of cervical cancer and its precursors. Part 1: results of a multicenter clinical trial. Am J Clin Pathol 2008 Feb;129(2):193-201

Confortini, M, Bonardi, L et al. 2003. A feasibility study of the use of the AutoPap screening system as a primary screening and location-guided rescreening device, *Cancer*, 99 (3), 129–134.

Coste J. 2003. Cross sectional study of conventional cervical smear, monolayer cytology, and human papillomavirus DNA testing for cervical cancer screening, *BMJ.*, 326(7392) (733),

Cunningham, J, Rumbold, AR et al. 2008. Incidence, aetiology, and outcomes of cancer in Indigenous peoples in Australia, *Lancet Oncology.9(6):585–95,*

Davey, E, Barratt, A et al. 2006. Effect of study design and quality on unsatisfactory rates, cytology classifications, and accuracy in liquid-based versus conventional cervical cytology: a systematic review, *Lancet*, 367 (9505), 122–132.

Davey, E, D'Assuncao, J et al. 2007a. Accuracy of reading liquid based cytology slides using the ThinPrep Imager compared with conventional cytology: prospective study, *British Medical Journal*, 335 (7609), 31–35.

Davey, E, Irwig, L et al. 2007b. Cervical cytology reading times: a comparison between ThinPrep Imager and conventional methods, *Diagnostic Cytopathology*, 35 (9), 550–554.

de Cremoux, P, Coste, J et al. 2003. Efficiency of the Hybrid Capture 2 HPV DNA test in cervical cancer screening. A study by the French Society of Clinical Cytology, *Am J Clin Pathol*, 120 (4), 492–499.

Dowie, R, Stoykova, B et al. 2006a. Liquid-based cytology can improve efficiency of cervical smear readers: evidence from timing surveys in two NHS cytology laboratories, *Cytopathology.*, 17 (2), 65–72.

Dowie, R, Stoykova, B et al. 2006b. Assessing the wellbeing of cytoscreeners: experience in two NHS cytology laboratories, *Cytopathology.*, 17 (6), 366–373.

Ferenczy, A, Robitaille, J et al. 1996. Conventional cervical cytologic smears vs. ThinPrep smears. A paired comparison study on cervical cytology, *Acta Cytol.*, 40 (6), 1136–1142.

Flannelly, G, Langhan, H et al. 1997. A study of treatment failures following large loop excision of the transformation zone for the treatment of cervical intraepithelial neoplasia, *Br J Obstet Gynaecol JID - 7503752*, 104 (6), 718–722.

Giorgi-Rossi, P, Segnan, N et al. 2007. The impact of new technologies in cervical cancer screening: results of the recruitment phase of a large randomised controlled trial from a public health perspective, *International Journal of Cancer*, 121 (12), 2729–2734.

Gray, NM, Sharp, L et al. 2006. Psychological effects of a low-grade abnormal cervical smear test result: anxiety and associated factors, *British Journal of Cancer*, 94 (9), 1253–1262.,

Herzog, TJ and Wright, JD 2007. The impact of cervical cancer on quality of life—the components and means for management, Gynecol. Oncol., 107 (3), 572–577.

Hussein T, Desai M, et al (2005). The comparative diagnostic accuracy of conventional and liquid-based cytology in a colposcopic setting. BJOG., 112(11):1542-6.

Hutchinson, ML, Berger, BM et al. 2000. Clinical and cost implications of new technologies for cervical cancer screening: the impact of test sensitivity, *Am. J. Manag. Care*, 6 (7), 766–780.

International Collaboration of Epidemiological Studies of Cervical Cancer, Appleby, P et al. 2007. Cervical cancer and hormonal contraceptives: collaborative reanalysis of individual data for 16,573 women with cervical cancer and 35,509 women without cervical cancer from 24 epidemiological studies, *Lancet*, 370 (9599), 1609–1621.

Irwig, L, Macaskill, P et al. 2004. A randomized crossover trial of PAPNET for primary cervical screening, *J. Clin. Epidemiol.*, 57 (1), 75–81.

Karnon, J, Peters, J et al. 2004. Liquid-based cytology in cervical screening: an updated rapid and systematic review and economic analysis, *Health Technology Assessment*, 8 (1–90.

Krahn, M, McLachlin, M et al. 2008. *Liquid-based techniques for cervical cancer screening: systematic review and cost-effectiveness analysis.* Technology report no. 103. Ottawa: Canadian Agency for Drugs and Technologies in Health.

Kulmala, SM, Syrjanen, S et al. 2004. Human papillomavirus testing with the Hybrid Capture 2 assay and PCR as screening tools, *J. Clin. Microbiol.*, 42 (6), 2470–2475.

Kyrgiou, M, Koliopoulos, G et al. 2006. Obstetric outcomes after conservative treatment for intraepithelial or early invasive cervical lesions: systematic review and meta-analysis, *Lancet*, 367 (9509), 489–498.

Lerma, E, Quintana, MJ et al. 2007. Effectiveness of liquid-based cytology and Papanicolaou tests in a low risk population, *Acta Cytologica*, 51 (3), 399–406.

Lier, D and Jacobs, P 2005, *An economic analysis of the introduction of liquid-based cytology and Human Papillomavirus (HPV) testing in Alberta.* Calgary: Alberta Cervical Cancer Screening Program.

Longatto Filho A, Miranda Pereira SMM, et al (2005). A. DCS liquid-based system is more effective than conventional smears to diagnosis of cervical lesions: Study in high-risk population with biopsy-based confirmation. *Gynecol Oncol*, 97:497–500

Maxwell, GL, Carlson, JW et al. 2002. Costs and effectiveness of alternative strategies for cervical cancer screening in military beneficiaries, *Obstet.Gynecol.*, 100 (4), 740–748.

McCredie, MR, Sharples, KJ et al. 2008. Natural history of cervical neoplasia and risk of invasive cancer in women with cervical intraepithelial neoplasia 3: a retrospective cohort study.[see comment], *Lancet Oncology*.9(5):425–34,

McCrory, DC, Matchar, DB et al. 1999, *Evaluation of cervical cytology*. Evidence report, technology assessment no.5.Rockville (MD): Agency for Health Care Quality and Research.

Medicare Australia 2008, Medicare Australia Statistics [Internet]. Available from: https://www.medicareaustralia.gov.au/statistics/mbs_item.shtml Accessed July 2008

Medical Services Advisory Committee 2002a, *Human papillomavirus testing in women with cytological prediction of low-grade abnormality*. Reference 12b.Canberra: Commonwealth of Australia.

Medical Services Advisory Committee 2002b, *Liquid based cytology for cervical screening*. Reference 12a.Canberra: Commonwealth of Australia.

Medical Services Advisory Committee 2003. *Computer-assisted image analysis for cervical screening*. Reference 12c. Canberra: Commonwealth of Australia.

Medical Services Advisory Committee 2005. *The use of human papillomavirus testing to monitor effectiveness of treatment of high-grade intraepithelial abnormalities of the cervix*. Canberra: MSAC.

Moher, D, Cook, DJ et al. 1999. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses, *Lancet*, 354 (9193), 1896–1900.

Montz, FJ, Farber, FL et al. 2001. Impact of increasing Papanicolaou test sensitivity and compliance: a modeled cost and outcomes analysis, *Obstet. Gynecol.*, 97 (5 Pt 1), 781–788.

Moss, S. M., Gary, Y., Marteau, T. 2004. *Evaluation of HPV/LBC cervical cancer screening studies* [Internet]. NHS Cancer Screening Programmes, Sheffield, UK, Available from: http://www.cancerscreening.nhs.uk/cervical/lbc.html#wales [Accessed 19 October 7 2007].

Myers, ER, McCrory, DC et al. 2000. Mathematical model for the natural history of human papillomavirus infection and cervical carcinogenesis, *Am. J. Epidemiol.*, 151 (12), 1158–1171.

Nanda, K, McCrory, DC et al. 2000. Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review, Ann. Intern. Med., 132 (10), 810–819.

Noorani HZ, Brown A, et al 2003.Liquid-based cytology and human papillomavirus testing in cervical cancer screening. Technology Report Issue 40. Ottawa, ON: Canadian Coordinating Office for Health Technology Assessment (CCOHTA);

National Institute for Clinical Excellence 2003, *Guidance on the use of liquid-based cytology for cervical screening*. Technology Appraisal 69.London, UK: National Institute for Clinical Excellence, NHS.

Neville, AM and Quinn, MA 2005. An alternative cost effectiveness analysis of ThinPrep in the Australian setting, *Australian and New Zealand Journal of Obstetrics and Gynaecology*, 45 (4), 289–294.

NHMRC 1994. *Screening to prevent cervical cancer: Guidelines for the management of women with screen detected abnormalities*. Canberra: National Health and Medical Research Council: Australian Government Publishing Service.

NHMRC 2000. *How to use the evidence: assessment and application of scientific evidence.* Canberra: Biotext.

NHMRC 2005, *Screening to prevent cervical cancer: Guidelines for the management of asymptomatic women with screen detected abnormalities.* Canberra, ACT: National Health and Medical Research Council, Commonwealth of Australia.

NHMRC 2008. *NHMRC additional levels of evidence and grades for recommendations for developers of guidelines.* Canberra: National Health and Medical Research Council (NHMRC).

Nieminen, P, Kotaniemi-Talonen, L et al. 2007. Randomized evaluation trial on automation-assisted screening for cervical cancer: results after 777,000 invitations, *Journal of Medical Screening*, 14 (1), 23–28.

Obwegeser, J.H. & Brack, S. (2001). 'Does liquid-based technology really improve detection of cervical neoplasia? A prospective, randomized trial comparing the ThinPrep Pap Test with the conventional Pap Test, including follow-up of HSIL cases', *Acta Cytologica*, 45, 709-714.

Oyesanya, OA, Amerasinghe, C et al. 1993. A comparison between loop diathermy conization and cold-knife conization for management of cervical dysplasia associated with unsatisfactory colposcopy, *Gynecologic Oncology*, 50 (1), 84–88.

Paraskevaidis, E, Lolis, ED et al. 2000. Cervical intraepithelial neoplasia outcomes after large loop excision with clear margins, *Obstetrics & Gynecology*, 95 (6:Pt 1), t-31.

Park, I.A., Lee, S.N. et al, (2001). 'Comparing the accuracy of ThinPrep Pap tests and conventional Papanicolaou smears on the basis of the histologic diagnosis - A clinical study of women with cervical abnormalities', *Acta Cytologica*, 45, 525-531.

Parkin DM, Whelan SL et al. eds 2002. *Cancer incidence in five continents* Vol. VIII, Lyon, France: IARC Scientific Publications.

Peto, J, Gilham, C et al. 17–7–2004. The cervical cancer epidemic that screening has prevented in the UK, *Lancet.*, 364 (9430), 249–256.

Powell, B 1996. The use of large loop excision of the transformation zone (LLETZ) in an outpatient setting, *Australian & New Zealand Journal of Obstetrics & Gynaecology*, 36 (3), 338–346.

Regan, DG, Philp, DJ et al. 2007. Modelling the population-level impact of vaccination on the transmission of human papillomavirus type 16 in Australia, *Sexual Health.4(3):147–63,*

Rema, P., Suchetha, S. et al. 2007. *Effectiveness, acceptability and safety of LEEP Effectiveness, acceptability and safety of LEEP in Kerala, India: beginners' experience.* Available from International Agency for Research on Cancer.

Riethmuller, D, Gay, C et al. 1999. Genital human papillomavirus infection among women recruited for routine cervical cancer screening or for colposcopy determined by Hybrid Capture II and polymerase chain reaction, *Diagnostic Molecular Pathology*, 8 (3), 157–164.

Roberts, JM, Thurloe, JK et al. 2007. A three-armed trial of the ThinPrep imaging system, *Diagnostic Cytopathology*, 35 (2), 96–102.

Rogstad, KE 2002. The psychological impact of abnormal cytology and colposcopy. BJOG: An International Journal of Obstetrics & Gynaecology.109(4):364–8,

Ronco, G, Cuzick, J et al. 2007a. Accuracy of liquid based versus conventional cytology: overall results of new technologies for cervical cancer screening: randomised controlled trial, *British Medical Journal*, 335 (7609), 28–31.

Ronco, G, Cuzick, J et al. 2007b. HPV triage for low-grade (L-SIL) cytology is appropriate for women over 35 in mass cervical cancer screening using liquid based cytology, *EJC*, 43 (476–480.

Ronco, G, Segnan, N et al. 2006. Human papillomavirus testing and liquid-based cytology: results at recruitment from the new technologies for cervical cancer randomized controlled trial, *Journal of the National Cancer Institute*, 98 (11), 765–774.

Royal College of Pathologists Australasia Cytopathology Quality Assurance Programs 2008. *Performance Standards for Australian Laboratories Reporting Cervical Cytology 2008*.

Safaeian, M, Herrero, R et al. 2007. Comparison of the SPF10-LiPA system to the Hybrid Capture 2 Assay for detection of carcinogenic human papillomavirus genotypes among 5,683 young women in Guanacaste, Costa Rica, *Journal of Clinical Microbiology*, 45 (5), 1447–1454.

Sandri, MT, Lentati, P et al. 2006. Comparison of the Digene HC2 assay and the Roche Amplicor human papillomavirus (HPV) test for detection of high-risk HPV genotypes in cervical samples, *Journal of Clinical Microbiology*, 44 (6), 2141–2146.

Schiffman, M 2007. Integration of human papillomavirus vaccination, cytology, and human papillomavirus testing, *Cancer.*, 111 (3), 145–153.

Schiffman, M and Adrianza, ME 2000. ASCUS-LSIL Triage Study. Design, methods and characteristics of trial participants, *Acta Cytol JID - 0370307*, 44 (5), 726–742.

Schiffman M, Castle PE. 2005 The promise of global cervical-cancer prevention. New England Journal of Medicine 353(20):2101–4.

Schiffman, M and Rodriguez, AC 2008. Heterogeneity in CIN 3 diagnosis., Lancet Oncology.9(5):404–6,

Schiffman, M and Solomon, D 2003. Findings to date from the ASCUS-LSIL Triage Study (ALTS), *Arch Pathol Lab Med JID - 7607091*, 127 (8), 946–949.

Sherlaw-Johnson, C and Philips, Z 2004. An evaluation of liquid-based cytology and human papillomavirus testing within the UK cervical cancer screening programme, *Br.J.Cancer.*, 91 (1), 84–91.

Sherman, ME, Castle, PE et al. 2006. Cervical cytology of atypical squamous cells-cannot exclude high-grade squamous intraepithelial lesion (ASC-H): characteristics and histologic outcomes, *Cancer*, 108 (5), 298–305.

Sherman, ME, Schiffman, M et al. 2002. Effects of age and human papilloma viral load on colposcopy triage: data from the randomized Atypical Squamous Cells of Undetermined Significance/Low-Grade Squamous Intraepithelial Lesion Triage Study (ALTS), *J Natl Cancer Inst*, 94 (2), 102–7.

Smith, MA, Canfell, K et al. 2008. The predicted impact of vaccination on human papillomavirus infections in Australia, *Int J Cancer*, 123 (8), 1854–1863.

Soderlund-Strand, A, Rymark, P et al. 2005. Comparison between the Hybrid Capture II test and a PCR-based human papillomavirus detection method for diagnosis and posttreatment follow-up of cervical intraepithelial neoplasia, *Journal of Clinical Microbiology*, 43 (7), 3260–3266.

Solomon D, Schiffman M et al. 2001. Comparison of three management strategies for patients with atypical squamous cells of undetermined significance: baseline results from a randomized trial, *Journal of the National Cancer Institute*, 93 (4), 293–299.

Stevens, MW, Milne, AJ et al. 2004. Effectiveness of AutoPap system location-guided screening in the evaluation of cervical cytology smears, *Diagnostic Cytopathology*, 31 (2), 94–99.

Sulik, S.M., Kroeger, K. et al, (2001). 'Are fluid-based cytologies superior to the conventional Papanicolaou test? A systematic review', Journal of Family Practice, 50, 1040-1046.

Taylor S, Kuhn L, et al (2006). Direct comparison of liquid-based and conventional cytology in a in a South African screening trial. Int J Cancer;957–62.

Tracey, E, Baker, D et al. 2007, *Cancer in New South Wales: Incidence, Mortality and Prevalence, 2005*. Sydney: Cancer Institute NSW.

Venturoli, S, Cricca, M et al. 2002. Human papillomavirus DNA testing by PCR-ELISA and hybrid capture II from a single cytological specimen: concordance and correlation with cytological results, *Journal of Clinical Virology*, 25 (2), 177–185.

Victorian Cervical Cytology Registry 2006. *Statistical Report 2006*.

Wahlstrom, C, Iftner, T et al. 2007. Population-based study of screening test performance indices of three human papillomavirus DNA tests, *Journal of Medical Virology*, 79 (8), 1169–1175.

Walker, JL, Wang, SS et al. 2006. Predicting absolute risk of CIN 3 during post-colposcopic follow-up: results from the ASCUS-LSIL Triage Study (ALTS), *Am J Obstet.Gynecol.*, 195 (2), 341–348.

Whiting, P, Rutjes, A et al. 2003. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews, *BMC Medical Research Methodology*, 3 (1), 25.

Wilbur, DC and Norton, MK 2002. The primary screening clinical trials of the TriPath AutoPap(registered trademark) System, *Epidemiology*, 13 (3 Suppl.), S30-S33.

Wilbur, DC, Parker, EM et al. 2002. Location-guided screening of liquid-based cervical cytology specimens: a potential improvement in accuracy and productivity is demonstrated in a preclinical feasibility trial, *American Journal of Clinical Pathology*, 118 (3), 399–407.

Willis, BH, Barton, P et al. 2005. Cervical screening programmes: Can automation help? Evidence from systematic reviews, an economic analysis and a simulation modelling exercise applied to the UK, *Health Technology Assessment*, 9 (13), iii–158.

Wordsworth, S, Ryan, M et al. 2006. Women's preferences for cervical cancer screening: a study using a discrete choice experiment, *International Journal of Technology Assessment in Health Care*, 22 (3), 344–350.

Wright, TC, Jr., Gagnon, S et al. 1992. Treatment of cervical intraepithelial neoplasia using the loop electrosurgical excision procedure.[see comment], *Obstetrics & Gynecology*, 79 (2), 173–178.

Yamazaki, H, Sasagawa, T et al. 2001. Hybrid capture-II and LCR-E7 PCR assays for HPV typing in cervical cytologic samples, *International Journal of Cancer*, 94 (2), 222–227.

Zuna, RE, Wang, SS et al. 2005. Determinants of human papillomavirus-negative, low-grade squamous intraepithelial lesions in the atypical squamous cells of undetermined significance/low-grade squamous intraepithelial lesions triage study (ALTS), *Cancer*, 105 (5), 253–262.