

MAVARIC – a comparison of automation-assisted and manual cervical screening: a randomised controlled trial

HC Kitchener, R Blanks, H Cubie, M Desai, G Dunn, R Legood, A Gray, Z Sadique and S Moss, on behalf of the MAVARIC Trial Study Group



January 2011
10.3310/hta15030

Health Technology Assessment
NIHR HTA programme
www.hta.ac.uk





How to obtain copies of this and other HTA programme reports

An electronic version of this title, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (www.hta.ac.uk). A fully searchable DVD is also available (see below).

Printed copies of HTA journal series issues cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our despatch agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per issue and for the rest of the world £3 per issue.

How to order:

- fax (with **credit card details**)
- post (with **credit card details** or **cheque**)
- phone during office hours (**credit card** only).

Additionally the HTA website allows you to either print out your order or download a blank order form.

Contact details are as follows:

Synergie UK (HTA Department)
Digital House, The Loddon Centre
Wade Road
Basingstoke
Hants RG24 8QW

Email: orders@hta.ac.uk

Tel: 0845 812 4000 – ask for ‘HTA Payment Services’
(out-of-hours answer-phone service)

Fax: 0845 812 4001 – put ‘HTA Order’ on the fax header

Payment methods

Paying by cheque

If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *University of Southampton* and drawn on a bank with a UK address.

Paying by credit card

You can order using your credit card by phone, fax or post.

Subscriptions

NHS libraries can subscribe free of charge. Public libraries can subscribe at a reduced cost of £100 for each volume (normally comprising 40–50 titles). The commercial subscription rate is £400 per volume (addresses within the UK) and £600 per volume (addresses outside the UK). Please see our website for details. Subscriptions can be purchased only for the current or forthcoming volume.

How do I get a copy of HTA on DVD?

Please use the form on the HTA website (www.hta.ac.uk/htacd/index.shtml). *HTA on DVD* is currently free of charge worldwide.

The website also provides information about the HTA programme and lists the membership of the various committees.

MAVARIC – a comparison of automation-assisted and manual cervical screening: a randomised controlled trial

HC Kitchener,^{1*} R Blanks,² H Cubie,³ M Desai,⁴ G Dunn,⁵ R Legood,^{6,7} A Gray,⁷ Z Sadique⁶ and S Moss,² on behalf of the MAVARIC Trial Study Group

¹School of Cancer and Enabling Sciences, University of Manchester, St Mary's Hospital, Manchester, UK

²Cancer Screening Evaluation Unit, The Institute of Cancer Research, Sutton, UK

³Specialist Virology Centre, Royal Infirmary of Edinburgh, Edinburgh, UK

⁴Manchester Cytology Centre, Central Manchester University Hospitals NHS Foundation Trust, Manchester, UK

⁵Health Sciences Research Group, School of Community Based Medicine, University of Manchester, Manchester, UK

⁶Health Services Research Unit, London School of Hygiene and Tropical Medicine, London, UK

⁷Health Economics Research Centre, University of Oxford, Oxford, UK

*Corresponding author

Declared competing interests of authors: none

Published January 2011

DOI: 10.3310/hta15030

This report should be referenced as follows:

Kitchener HC, Blanks R, Cubie H, Desai M, Dunn G, Legood R, *et al.* MAVARIC – a comparison of automation-assisted and manual cervical screening: a randomised controlled trial. *Health Technol Assess* 2011;**15**(3).

Health Technology Assessment is indexed and abstracted in *Index Medicus/MEDLINE*, *Excerpta Medica/EMBASE*, *Science Citation Index Expanded (SciSearch®)* and *Current Contents®/Clinical Medicine*.

The Health Technology Assessment (HTA) programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The research findings from the HTA programme directly influence decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC). HTA findings also help to improve the quality of clinical practice in the NHS indirectly in that they form a key component of the 'National Knowledge Service'.

The HTA programme is needs led in that it fills gaps in the evidence needed by the NHS. There are three routes to the start of projects.

First is the commissioned route. Suggestions for research are actively sought from people working in the NHS, from the public and consumer groups and from professional bodies such as royal colleges and NHS trusts. These suggestions are carefully prioritised by panels of independent experts (including NHS service users). The HTA programme then commissions the research by competitive tender.

Second, the HTA programme provides grants for clinical trials for researchers who identify research questions. These are assessed for importance to patients and the NHS, and scientific rigour.

Third, through its Technology Assessment Report (TAR) call-off contract, the HTA programme commissions bespoke reports, principally for NICE, but also for other policy-makers. TARs bring together evidence on the value of specific technologies.

Some HTA research projects, including TARs, may take only months, others need several years. They can cost from as little as £40,000 to over £1 million, and may involve synthesising existing evidence, undertaking a trial, or other research collecting new data to answer a research problem.

The final reports from HTA projects are peer reviewed by a number of independent expert referees before publication in the widely read journal series *Health Technology Assessment*.

Criteria for inclusion in the HTA journal series

Reports are published in the HTA journal series if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

The research reported in this issue of the journal was commissioned by the HTA programme as project number 03/04/02. The contractual start date was in August 2005. The draft report began editorial review in February 2010 and was accepted for publication in June 2010. As the funder, by devising a commissioning brief, the HTA programme specified the research question and study design. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the referees for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

The views expressed in this publication are those of the authors and not necessarily those of the HTA programme or the Department of Health.

Editor-in-Chief: Professor Tom Walley CBE
 Series Editors: Dr Martin Ashton-Key, Professor Aileen Clarke, Dr Peter Davidson,
 Professor Chris Hyde, Dr Tom Marshall, Professor John Powell, Dr Rob Riemsma and
 Professor Ken Stein
 Editorial Contact: edit@southampton.ac.uk

ISSN 1366-5278

© 2011 Queen's Printer and Controller of HMSO

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (<http://www.publicationethics.org/>).

This journal may be freely reproduced for the purposes of private research and study and may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NETSCC, Health Technology Assessment, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk), on behalf of NETSCC, HTA.

Printed on acid-free paper in the UK by the Charlesworth Group.

Abstract

MAVARIC – a comparison of automation-assisted and manual cervical screening: a randomised controlled trial

HC Kitchener,^{1*} R Blanks,² H Cubie,³ M Desai,⁴ G Dunn,⁵
R Legood,^{6,7} A Gray,⁷ Z Sadique⁶ and S Moss,² on behalf of the
MAVARIC Trial Study Group

¹School of Cancer and Enabling Sciences, University of Manchester, St Mary's Hospital, Manchester, UK

²Cancer Screening Evaluation Unit, The Institute of Cancer Research, Sutton, UK

³Specialist Virology Centre, Royal Infirmary of Edinburgh, Edinburgh, UK

⁴Manchester Cytology Centre, Central Manchester University Hospitals NHS Foundation Trust, Manchester, UK

⁵Health Sciences Research Group, School of Community Based Medicine, University of Manchester, Manchester, UK

⁶Health Services Research Unit, London School of Hygiene and Tropical Medicine, London, UK

⁷Health Economics Research Centre, University of Oxford, Oxford, UK

*Corresponding author henry.c.kitchener@manchester.ac.uk

Objectives: The principal objective was to compare automation-assisted reading of cervical cytology with manual reading using the histological end point of cervical intraepithelial neoplasia grade II (CIN2) or worse (CIN2+). Secondary objectives included (i) an assessment of the slide ranking facility of the Becton Dickinson (BD) FocalPoint™ Slide Profiler (Becton Dickinson, Franklin Lakes, NJ, USA), especially 'No Further Review', (ii) a comparison of the two approved automated systems, the ThinPrep® Imaging System (Hologic, Bedford, MA, USA) and the BD FocalPoint Guided Screener Imaging System, and (iii) automated versus manual in terms of productivity and cost-effectiveness.

Design: A 1:2 randomised allocation of slides to either manual reading or automation-assisted paired with manual reading. Cytoscreeners were blinded to whether samples would be read only manually or manually paired with automated. Slide reading procedures followed real-life laboratory protocol to produce a final result and, for paired readings, the worse result determined the management. Costs per event were estimated and combined with productivity to produce a cost per slide, per woman and per CIN2+ and cervical intraepithelial neoplasia grade III (CIN3) or worse (CIN3+) lesion detected. Cost-effectiveness was estimated using cost per CIN2+ detected. Lifetime cost-effectiveness in terms of life-years and quality-adjusted life-years was estimated using a mathematical model.

Setting: Liquid-based cytology samples were obtained in primary care, and a small number of abnormal samples were obtained from local colposcopy clinics, from different women, in order to enrich the proportion of abnormal. All of the samples were read in a single large service laboratory. Liquid residues used for human papillomavirus (HPV) triage were tested (with Hybrid Capture 2, Qiagen, Crawley, UK) in a specialist virology laboratory

in Edinburgh, UK. Histopathology was read by a specialist gynaecological pathology team blinded to HPV results and type of reading.

Participants: Samples were obtained from women aged 25–64 years undergoing primary cervical screening in Greater Manchester, UK, with small proportions from women outside this age range and from women undergoing colposcopy.

Interventions: The principal intervention was automation-assisted reading of cervical cytology slides which was paired with a manual reading of the same slide. Low-grade cytological abnormalities (borderline and mild dyskaryosis) were triaged with HPV testing to direct colposcopy referral. Women with high-grade cytology were referred for colposcopy and those with negative cytology were returned to recall.

Main outcome measures: The principal outcome measure was the sensitivity of automation-assisted reading relative to manual for the detection of CIN2+. A secondary outcome measure was cost-effectiveness of each type of reading to detect CIN2+. The study was powered to detect a relative sensitivity difference equivalent to an absolute difference of 5%.

Results: The principal finding was that automated reading was 8% less sensitive relative to manual, 6.3% in absolute terms. 'No further review' was very reliable and, if restricted to routine screening samples, <1% of CIN2+ would have been missed. Automated and manual were very similar in terms of cost-effectiveness despite a 60%–80% increase in productivity for automation-assisted reading.

Conclusions: The significantly reduced sensitivity of automated reading, combined with uncertainty over cost-effectiveness, suggests no justification at present to recommend its introduction. The reliability of 'no further review' warrants further consideration as a means of saving staff time.

Trial registration: Current Controlled Trials ISRCTN66377374.

Funding: This project was funded by the NIHR Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 15, No. 3. See the HTA programme website for further project information.

Contents

List of abbreviations	vii
Executive summary	ix
1. Introduction	1
The English cervical screening programme	1
Liquid-based cytology	2
Automated technologies	3
Other current trials of automated screening	8
Human papillomavirus testing	9
2. Study design and methods	13
Aims and objectives of the MAVARIC study	13
Trial design	14
Inclusion and exclusion criteria	15
Human papillomavirus triage	16
Settings and ethics approval	16
Logistical considerations	17
Summary of significant changes to the protocol during the course of the study	18
Automated cytology methods	18
Clinical management	21
Data collection	22
Statistical analysis, including statistical considerations	25
Economic analysis	28
Measuring costs	31
Selection of end points	37
3. Results	39
Summary of randomisation	39
The consolidated standards of reporting trials diagram	42
Clinical results	42
Primary outcome	54
Secondary outcomes	57
Economics and organisational outcomes	66
4. Discussion	87
Cytology outcomes	88
Colposcopy referral	88
Summary cytology data in the paired arm	88
Primary outcome	89
Secondary outcomes	91
Implications for the NHS Cervical Screening Programme	94
Research recommendations	95
Acknowledgements	97
References	101

Appendix 1 Time-and-motion survey questionnaire for loading and unloading of automated sample	107
Appendix 2 Time-and-motion survey questionnaires	109
Appendix 3 Primary screener worksheet	113
Appendix 4 Staff satisfaction questionnaire	115
Appendix 5 Roche Amplicor human papillomavirus testing	119
Appendix 6 Automated cytology training	123
Appendix 7 Proforma for the review of discordant pairs	125
Appendix 8 Human papillomavirus genotyping	127
Appendix 9 National Screening Committee's criteria for appraising the viability, effectiveness and appropriateness of a screening programme	131
Appendix 10 Tables restricted to routine samples from women aged 25–64 years	133
Appendix 11 Staff satisfaction survey results	135
Appendix 12 Results of model fitting and additional parameters used in sensitivity analyses	141
Appendix 13 Additional tables relating to the comparison of results between manual and automated readings in the paired arm	147
Appendix 14 Comparison of results between manual-only and paired arm	149
Appendix 15 Final trial protocol	151
Appendix 16 The standards for the reporting of diagnostic accuracy studies checklist	167
Appendix 17 The consolidated standards of reporting trials 2010 checklist of information to include when reporting a randomised trial	169
Health Technology Assessment programme	171

List of abbreviations

ACCS	Advisory Committee on Cervical Screening
AP	advanced biomedical scientist practitioner
ARF	automated read failure
AR1	automated result 1
AR2	automated result 2
ARTISTIC	A Randomised Trial In Screening To Improve Cytology
ASCUS	abnormal squamous cells of undetermined significance
BD	Becton Dickinson
BMS	biomedical scientist
CI	confidence interval
CIN	cervical intraepithelial neoplasia
CIN1	cervical intraepithelial neoplasia grade I
CIN1–	any lesion of CIN grade I or less – cases not requiring treatment
CIN2	cervical intraepithelial neoplasia grade II
CIN2+	any lesion of CIN2 or worse
CIN2–	any lesion of CIN grade II or less
CIN3	cervical intraepithelial neoplasia grade III
CIN3+	any lesion of CIN3 or worse
CONSORT	consolidated standards of reporting trials
CSEU	Cancer Screening Evaluation Unit, Surrey
DNA	deoxyribonucleic acid
FDA	the US Food and Drug Administration
FOV	field of view
FAR	final automated result
FMR	final manual result
GP	general practitioner
GS	guided screener
HCHS	Hospital and Community Health Service
HC2	Digene high-risk HPV Hybrid Capture® 2
HPV	human papillomavirus
HTA	Health Technology Assessment
LBC	liquid-based cytology
LREC	Local Research Ethics Committee
LSIL	low-grade squamous intraepithelial lesion
LSIL+	LSIL or worse
MAVARIC	Manual Assessment Versus Automated Reading In Cytology
MLA	medical laboratory assistant
MR	management result
MR1	manual result 1
MR2	manual result 2
MWP	microwell plate
NFR	no further review
NHSCSP	NHS Cervical Screening Programme
NICE	National Institute for Health and Clinical Excellence
NPV	negative predictive value
Pap	Papanicolaou
PCR	polymerase chain reaction
PCT	primary care trust

PPV	positive predictive value
QALY	quality-adjusted life-year
QC	quality control
RLU	relative light unit
RLU/CO	relative light unit/cut-off
<i>rt</i> HPV	real-time high-risk HPV
TPR	true-positive rate
WHO	World Health Organization

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices, in which case the abbreviation is defined in the figure legend or in the notes at the end of the table.

Executive summary

Objectives

Cervical screening currently relies on manually read slides in which the cytoscreener scans the entire slide looking for abnormal cells. This study evaluated technology that assists reading cytology by automatically detecting abnormal fields of view on a slide and presenting these to a cytoscreener on an automated microscope. This could potentially achieve greater sensitivity and productivity, thus saving lives and achieving a more efficient use of the cytology workforce. This study had the following objectives:

- To determine the sensitivity of automation-assisted reading relative to manual reading.
- To determine any added productivity of automated reading.
- To estimate the comparative cost-effectiveness of automated and manual reading.
- To determine the reliability of 'no further review' (NFR) without any reading.

Design

Samples were randomised to a paired arm reported by both automated and manual reading and an arm with manual reading only. All of the cytology was liquid based, and the study incorporated randomisation of both widely used liquid-based cytology systems and their corresponding automated imaging technology, one of which ranks slides in terms of abnormality and will select around one-fifth as requiring NFR.

Setting

The samples were obtained from women undergoing cervical screening in the NHS programme, principally in general practices, in Greater Manchester, UK.

Samples

Samples from 73,266 women were obtained between March 2006 and February 2009; 72,837 were included in the study. Almost all of the women were aged 25–64 years (69,218). Randomisation resulted in 24,566 (33.7%) slides in the manual arm and 48,271 (66.3%) in the paired arm.

Intervention

In the paired arm, automation-assisted reading of slides was performed in addition to manual reading and management determined by the worse result. Low-grade cytological abnormalities were triaged by a human papillomavirus (HPV) test (Hybrid Capture 2[®]; Qiagen, Crawley, UK) to select women for colposcopy referral. All women with high-grade abnormalities were referred for colposcopy. If cervical intraepithelial neoplasia grade II (CIN2) or worse (CIN2+) was detected, the woman was treated. Additionally, a detailed economic analysis of the cytology reading was undertaken.

Main outcome measures

The primary outcome was the sensitivity of the final automated result relative to that of the final manual result in the paired arm. Secondary outcome measures included an assessment of productivity and estimates of cost-effectiveness, and an evaluation of the reliability of the NFR facility in the Becton Dickinson (BD) FocalPoint™ Guided Screener (GS) Imaging System (BD, Franklin Lakes, NJ, USA).

Results

The proportion of abnormal cytology management results by grade were: borderline, 3.6%; mild dyskaryosis, 2.4%; and moderate and severe dyskaryosis combined, 1.22%. These were very similar to England as a whole. The non-negative cytology amounted to 5.47% in the paired arm and 5.52% in the manual-only arm. Within the paired arm the proportion of discordant pairs on final result was 3.8% (1850/48,271); for 1.3% (625/48,271), the discordance was between inadequate and negative. Discordant pairs occurred in both directions with respect to manual and automated reading. There were 192 additional low-grade/HPV-positive abnormalities detected by manual reading only (manual positive/auto negative) and 47 additional high-grade abnormalities detected by manual reading only in the paired arm. The overall referral rate to colposcopy was 4.7%. The proportion with CIN2+ was 1.6% (398/24,566) and 1.5% (707/48,271) for the manual and paired arms respectively ($p=0.10$). The primary outcome of the relative sensitivity for CIN2+ of automated reading compared with manual reading in the paired arm was 0.92 [95% confidence interval (CI) 0.85 to 0.95]. The relative specificity was 1.006 (95% CI 1.005 to 1.007).

Productivity in terms of the number of slides read per day by primary screeners was estimated to be 60%–80% higher for automated reading than for manual reading. The overall costs per case of CIN2+ detected were almost identical between automated and manual reading (£2892, 95% CI £2720 to £3098; and £2838, 95% CI £2676 to £3030 respectively). The overall costs per case of cervical intraepithelial neoplasia grade III (CIN3) or worse (CIN3+) detected are also very similar between automated and manual reading (£4762, 95% CI £4378 to £5245; and £4775, 95% CI £4400 to £5244 respectively). Manual screening is therefore slightly more expensive and effective, and could be considered cost-effective compared with automated reading if decision-makers were willing to pay at least £5000 each additional case of CIN2+ detected. NFR in the BD FocalPoint GS Imaging System was reported in 22% of slides and was a very reliable indicator of the absence of underlying disease, with only 3.1% of detected CIN2+ being missed by NFR, and even more so if NFR was restricted to routine screening slides. When both savings in staff time to read slides and the additional equipment costs were taken into account, utilising the NFR option generated cost savings. Based on all slides included in the MAVARIC (Manual Assessment Versus Automated Reading In Cytology) study, assessment of the incremental cost per case detected revealed that decision-makers would need to be willing to pay £2500 per additional case of CIN2+ detected for it to be more cost-effective to read slides manually instead.

Results of the lifetime modelling indicated that when life-years were used as an outcome measure, manual reading was within the £20,000–30,000 per life-year saved range in which the National Institute for Health and Clinical Excellence would neither accept nor reject this technology on cost-effectiveness grounds alone. Modelled results were also estimated for quality-adjusted life-years gained, but these are highly uncertain given the absence of trial evidence on utility values.

Conclusions

The principal finding was that automation-assisted reading was 8% less sensitive than manual in the detection of CIN2+ and 5% less sensitive for CIN3+. To a large extent, this was due to automation-assisted reading failing to detect cases of low-grade abnormalities that were detected in manual reading. The majority of missed cases were due to failure to detect abnormalities presented rather than location-guided errors. Despite the undoubted productivity gains that could be achieved in terms of slide throughput, there do not appear to be sufficient grounds to recommend automation. The slight gain in specificity is not of clinical importance; the positive predictive value (CIN2+) of additional manually read abnormal cytology leading to colposcopy referral would be in line with that of HPV-positive/mild abnormalities currently triaged to colposcopy. Secondly, given the pricing obtained from the companies and used in this study, the cost-effectiveness of automation-assisted reading is marginal at best, compared with manual reading. Thirdly, there was a general view among the cytoscreeners that they find the automation-assisted reading more monotonous and prefer manual reading.

Although automation-assisted reading did not compare favourably with manual reading, the robust evaluation of the NFR mode of the BD FocalPoint GS Imaging System showed it to be very reliable and able to achieve cost savings in staff time, even if some methods of manual rapid review were maintained for quality control purposes. A significant reduction in the number of slides needing full screening would enhance efficiency and turnaround times.

Were, however, conclusive evidence to emerge in the future that the sensitivity concerns had been resolved and the cost-effectiveness of automation significantly improved, then the recommendation against automation would warrant reconsideration.

Trial registration

This trial is registered as ISRCTN66377374.

Funding

The National Institute for Health Research Health Technology Assessment programme.

Chapter 1

Introduction

The English cervical screening programme

Evidence of the effectiveness of cervical screening

The NHS Cervical Screening Programme (NHSCSP) began a managed programme of call and recall in 1988 and is estimated to save as many as 5000 lives per year in the UK.¹ It has become recognised as one of the world's leading cervical cancer prevention programmes.

Harnessing new technology to improve service efficiency is a key strategy of the NHSCSP. Desirable advances in cytology include improving sensitivity and specificity, and reducing human workload. The number of tests processed by the screening programme has dropped significantly in recent years owing to service improvement. The roll-out of liquid-based cytology (LBC), completed in 2008, has seen the number of inadequate samples (and the associated repeat testing) drop from 9% in 2004–5 to 2.9% in 2007–8.² The implementation of six sentinel sites for human papillomavirus (HPV) triage and test of cure around England has reduced the number of repeat tests taken by triaging women on the basis of their HPV results. Women attending for routine tests who are found to have a low-grade abnormality and a positive HPV result are referred directly to colposcopy without repeat cytology testing, and those who are HPV negative are returned to routine recall without cytological follow-up.³ National roll-out of HPV triage and 'test of cure' would further reduce the amount of cytology, and allow women either to be diagnosed and, if necessary, treated more quickly, or to be returned to routine recall.

Current manual screening practice

Current programme guidelines recommend that all cytology is primary screened; slides reported as negative or inadequate receive a rapid review, and slides that are suspected to be abnormal are reviewed and reported by senior laboratory staff.⁴ It is recommended that cytoscreeners do a maximum of 5 hours of microscopy work in a 24-hour period, with a complete break from the microscope at least every 2 hours.⁵ With the introduction of LBC, rapid screening is carried out by screening staff performing a rapid review of the whole slide in 90 seconds. Current screening techniques are time-consuming and require a large and committed laboratory workforce. Despite the effectiveness of the screening programme, cytoscreeners have often felt under pressure, particularly when failures receive media attention.

Screening schedule and coverage

Currently, women aged 25–49 years are invited every 3 years, and women aged 50–64 years are invited every 5 years.² Of the 3.6 million women aged 25–64 years who were screened in 2008–9, around 6.7% received an abnormal result.⁶ In the same period there were 134,000 referrals to colposcopy prompted by an abnormal screening result, 28.9% of which were for results of moderate dyskaryosis or worse,⁶ the remainder resulting from low-grade cytological abnormalities.

Despite the efficiency of the call–recall system, coverage for the year 2007–8 fell below 80% for the first time, at 78.6%.⁷ There has been particular concern in recent years over the fall in attendance in the under-30s, although this trend was bucked during 2009 following the occurrence of cervical cancer in a media celebrity. A total of 3.6 million women aged 25–64 years

were screened in 2008–9 compared with 3.2 million in 2007–8 – an increase of 11.9% with an increase in coverage to 78.9% [with a range of 65.8%–85.8% between primary care trusts (PCTs)].⁶ The durability of this increase will not be confirmed until the publication of screening statistics for tests taken in 2009–10.

Future programme considerations

Alongside the question of whether or not to implement automated screening, there are several organisational challenges that face the NHSCSP. In 2007 the Department of Health published the *Cancer Reform Strategy*.⁸ This document recommended that in order to achieve the Government's target of a 14-day turnaround time (from cervical sample being taken to the result being received by the woman), laboratories and screening offices should be reconfigured to make them larger and more efficient. Some laboratories currently operate as 'hub and spoke' with larger central laboratories processing the LBC samples and returning them to the smaller laboratories for screening. Amalgamation of smaller laboratories will see further changes to this service configuration. NHS pathology services as a whole across England are also under review by the Department of Health as part of the NHS Pathology Improvement Programme,⁹ which may have further implications for the NHSCSP's laboratory infrastructure.

The HPV vaccination programme will also have an impact on screening once vaccinated girls enter the screening programme. (Girls are vaccinated at ages 12–13 years. This began in September 2008 when a 3-year catch-up campaign began to vaccinate older girls aged 14–17 years.) Screening intervals and follow-up protocols will need to be reviewed once the evidence base regarding screening in a vaccinated population becomes clearer. The importance of following up the screening outcomes of recently vaccinated girls was stressed by the Advisory Committee on Cervical Screening (ACCS) during the review of current screening policy in women aged 20–24 years.¹⁰ Following recommendations from the ACCS, the Department of Health decided against making any changes to current policy regarding screening in women aged 20–24 years. Instead, further education of general practice staff will ensure that symptomatic women aged < 25 years are assessed appropriately.¹¹

Liquid-based cytology

The conventional method of producing cervical cells on a glass slide involved a sample being obtained from the cervix using a spatula which was smeared onto a glass slide and then fixed. Fifty years on, this method is still widely used worldwide. The quality of the slide material is variable, with blood cells and mucus capable of obscuring the cervical cells, as well as cells being unevenly spread. This has led to a large number of slides being designated as 'inadequate' for reporting.

With LBC, the cervical sample is dissipated in a fluid medium which contains fixative. The liquid sample is then subjected to either a process which filters the cells onto a slide (ThinPrep[®] LBC, Hologic, Bedford, MA, USA) or cell enrichment [Becton Dickinson (BD) SurePath[™] LBC, BD, Franklin lakes, NJ, USA] producing a cleaner, more homogeneous preparation which facilitates examination of the cervical cells. In 2001–3 an NHSCSP pilot study was performed in England in order to evaluate LBC in comparison with conventional cytology in a historical population. The findings were that inadequate samples were reduced from around 7%–8% to around 1%, that LBC was certainly not less sensitive than conventional cytology and possibly more so, that laboratory throughput was more efficient, and that laboratory staff preferred LBC.¹² LBC was determined to be cost-effective and meant that far fewer women were recalled because of an 'inadequate' smear. The National Institute for Health and Clinical Excellence (NICE)

recommended its adoption¹³ and between 2003 and 2008 LBC was rolled out nationally across the entire UK.

Two of the critical differences between LBC and conventional cytology are (1) reading of LBC slides can be automated using the technology being evaluated in the Manual Assessment Versus Automated Reading In Cytology (MAVARIC) study and (2) the LBC residue can be used for real-time reflex testing such as HPV testing to triage low-grade cytological abnormalities. The adoption of LBC provided the means for a more efficient cytology service, enabling both triage and the potential to move to automated technology if that were shown to be cost-effective.

Automated technologies

Development of technologies

Two US Food and Drug Administration (FDA)-approved automated machines were developed in the 1990s, the AutoPap[®] 300 QC (NeoPath, Redmond, WA, USA) and the PapNet[®] (Neuromedical Systems Inc., Suffern, NY, USA), both systems being designed to work with conventional cytology slides. AutoCyte had also developed a machine known as the AutoCyte-Screen which was able to read AutoCyte-Prep slides (now BD SurePath LBC). Despite the initial promise of the technology none of these machines is now available. AutoCyte and NeoPath merged to form TriPath Imaging Inc. (Burlington, NC, USA) and discontinued both the AutoCyte and the AutoPap 300 QC, replacing the systems with the AutoPap Primary Screening System, which is now known as the BD FocalPoint[™] GS Imaging System (BD Diagnostics, Franklin Lakes, NJ, USA).

There are currently two commercially available FDA-approved automated screening systems – the BD FocalPoint GS Imaging System and the ThinPrep Imaging System (Hologic[™], Bedford, MA, USA). The BD FocalPoint Slide Profiler scans the slides and assigns each one a rank according to the likelihood of there being abnormal cells present. The slides are assigned to quintiles, with quintile 1 containing the highest ranking slides. The machine also categorises slides into one of four of categories: review (comprising quintiles 1–5), no further review (NFR; up to 25% of slides), process review (indicating a technical problem) and quality control review (requiring a full screen). NFR designates the 25% of slides least likely to contain an abnormality which could be reported as negative and archived without human reading. Slides that are flagged for review by the system are examined by screening staff using the BD FocalPoint Guided Screener Workstation (previously known as TriPath Slide Wizard[®]). This comprises a standard screening microscope fitted with an electronic stage linked to a desktop computer. The Workstation directs screening staff towards 10 electronically marked fields of view (FOVs) on the slide. If abnormal cells are seen in any of the FOVs the entire slide is screened and appropriate action taken in line with laboratory protocols. The BD FocalPoint Guided Screener (GS) Imaging System has received FDA approval to scan both conventional and BD SurePath LBC slides.

In contrast, the ThinPrep Imaging System is designed to work with ThinPrep LBC slides (stained with the Hologic Imager stain) alone. The ThinPrep Imaging System scans all of the slides and selects 22 FOVs which are presented to screening staff on the review scope. The review scope comprises a Hologic automated screening microscope with a motorised stage to guide screeners to each of the 22 FOVs. If an abnormality is suspected in any of the 22 FOVs then a full screen of the slide is undertaken. Unlike the BD FocalPoint GS Imaging System, the ThinPrep Imaging System does not assign scores to slides and is therefore unable to rank and select slides for archiving without further intervention, or to select slides for quality control (QC) reviewing.

Capability of automated cytology

Two systematic reviews have been published on the potential of automated screening technologies.^{14,15} A review commissioned by the Health Technology Assessment (HTA) programme and published in 2005 concluded that there was a need for rigorous, unbiased public sector research into the effectiveness of automated screening technologies.¹⁴ One drawback of this review was that the majority of the papers included relate to the now obsolete PapNet and AutoPap 300 QC systems. An earlier review by the New Zealand HTA programme reached a similar conclusion and recommended large-scale prospective trials to be conducted under normal laboratory conditions with reliable gold standards for diagnostic verification.¹⁵ This review also focused on technologies that are no longer commercially available. As yet there have been no systematic reviews that focus on the two currently available technologies which are under appraisal in the MAVARIC study.

Table 1 summarises previous 'controlled' studies, in which there was a general pattern of increased rates of abnormality detection in the automated arm. The studies are, however, characterised by methodological weaknesses including the use of outdated systems, using split samples, the use of manually read conventional (as opposed to liquid-based) cytology, using the same slide set for retrospective comparative readings and not reporting histological outcomes.

TABLE 1 Recent automated cytology publications

Study and design	Comparison groups	CIN detection rates	Sensitivity/Specificity	PPV/NPV
Halford et al.¹⁶				
Prospective two-armed masked study. Histology taken within 6 months of the Pap smear was used as the reference standard	87,284 split sample conventional slides read manually and ThinPrep LBC slides read with the ThinPrep Imaging System. Biopsy data were available for 1083 HSIL lesions	Automated-LBC reading showed a 3.2% increase in possible high-grade and HSIL reports compared with manually reading convention slides	For ASCUS+ the sensitivity of automated was 96.0% and manual 91.6% ($p=0.001$)	For 1083 biopsy confirmed HSIL cases automated was correct in 61% of cases and 59.4% on manual ($p=0.05$)
Wilbur et al.¹⁷				
Prospective two-armed masked study. Truth adjudication used as the gold standard	12,313 slides screened using both the BD FocalPoint GS Imaging System's FOV and QC and manually with manual QC	Not given	HSIL+ sensitivity 85.3% in automated arm and 65.7% in manual ($p<0.0001$) with a 2.6% decline ($p<0.0001$) in specificity. LSIL+ sensitivity 86.1% automated and 76.4% in manual ($p<0.0001$) with a 1.9% ($p=0.0032$) in specificity ASCUS+ sensitivity and specificity were not significantly different between the two arms	NPV of a not HSIL+ slide in the automated arm was 99.7% and 99.4% in the manual arm
Pacheco et al.¹⁸				
Retrospective analysis comparing samples taken during the first 6 months of both 2004 and 2005. Final and initial diagnoses on the same slide were compared for the analysis	79,791 manually screened ThinPrep slides and 76,887 slides screened with the ThinPrep Imaging System	Number of diagnosed HSIL cases increased from 0.46% to 0.78% with use of the ThinPrep Imaging System ($p<0.01$)	Not given	Not given

TABLE 1 Recent automated cytology publications (*continued*)

Study and design	Comparison groups	CIN detection rates	Sensitivity/Specificity	PPV/NPV
Papillo et al.¹⁹				
Retrospective comparison study with biopsy data collected for 64% of HSIL cases	55,547 ThinPrep Imaging System slides and 54,565 manually read LBC slides	LSIL cytology significantly increased by 29%, HSIL by 54%	Not given	Not given
Passamonti et al.²⁰				
Routine consecutive conventional Pap slides prospectively processed on the BD FocalPoint GS Imaging System. Histology was obtained for 67% of slides showing abnormalities	37,306 conventional Pap slides processed and screened using the BD FocalPoint GS Imaging System. All slides then received a manual rapid screen before the results were compared	91% of CIN2+ cases were ranked in high-risk quintiles along with 93% of CIN1. 97% of HSIL+ and 98% of LSIL slides were triaged for a full manual review by screening the FOVs	Not given	Not given
Lozano²¹				
Retrospective comparison with biopsy data collected for all HSIL+ samples	39,717 ThinPrep Imaging System slides and 87,262 manually read LBC slides	HSIL+ cytology significantly increased by 38% and LSIL by 46%	Not given	PPV of HSIL for CIN2+ = 83% for automated and 84% for manual. HSIL for CIN1+ = 98% for automated and 96% for manual
Troni et al.²²				
Concurrent cohorts retrospectively identified with a negative screen at baseline. Screening modality at repeat smear was independent of the baseline screen. All subjects with CIN2+ at repeat screening were identified	AutoPap Primary Screening System 300 using conventional slides compared with manually read conventional slides. 33,646 women at baseline, 30,658 of whom returned for repeat screening. 30% randomised to manual reading	No significant difference in CIN2+ detection at repeat screening when comparing baseline automated and manual cohorts	Not given	Not given
Miller et al.²³				
Two consecutive cohorts. Biopsy data were used as the reference standard for ASCH+	82,063 manually read ThinPrep slides, 84,473 slides read with the ThinPrep Imaging System	Significant decrease in ASCUS (15.56%) in the automated cohort along with a significant increase in LSIL (37.62%) and HSIL (42.42%)	Not given	Not given
Davey et al.²⁴				
Prospective study using split sample pairs. Histology results were obtained for discordant pairs	55,164 split samples – ThinPrep Imaging System compared with manually read conventional slides	Significantly fewer inadequates in the automated arm (1.8% vs 3.1%). Automated detected 1.29 more cases of histologically confirmed high-grade disease per 1000 women and classified 8.6 more slides as low grade per 1000 women	Not given	Not given

continued

TABLE 1 Recent automated cytology publications (*continued*)

Study and design	Comparison groups	CIN detection rates	Sensitivity/Specificity	PPV/NPV
Schledermann et al.²⁵				
Comparative study with three distinct phases: manual screening, automated screening training and routine automated screening. All abnormal slides discussed with senior pathologists	11,354 slides in total to compare ThinPrep Imaging System read slides during training and routine use with manually read LBC slides	Not given	During routine use the sensitivity of the ThinPrep Imaging System was 93.3% and the specificity 97.6%	Not given
Roberts et al.²⁶				
Three-armed trial. The worst histopathology result within 9 months of the end of the trial was collected	11,416 split sample ThinPrep and conventional slides. ThinPrep slides read both manually and with the ThinPrep Imaging System. Conventional slides read manually	14 false-negatives in the ThinPrep Imaging System arm, nine in the ThinPrep manual arm and 28 in the conventional arm	Sensitivity for reporting high-grade disease = 86.8% in the ThinPrep manual arm and 81.1% in the ThinPrep Imaging System arm	No significant difference between the PPV of the ThinPrep Imaging System arm and both the ThinPrep and conventional manual arms for high-grade reports
Dziura et al.²⁷				
Two consecutive cohorts. All available biopsy data collected for ASC-H and HSIL	27,525 manually screened ThinPrep slides and 27,725 ThinPrep Imaging System read slides	29% increase in ASCUS detection, 50% increase in ASC-H detection, 30.7% increase in LSIL detection and 20% increase in HSIL detection in ThinPrep Imaging System arm (all significant). Also an increase in ASC-H (11.7%) and HSIL (8.9%) samples showing HSIL on biopsy in ThinPrep Imaging System arm (not significant)	Not given	Not given
Bulgaresi et al.²⁸				
An evaluation of rapid review of slides designated NFR as a QC procedure. ASCUS–SIL+ samples were reviewed before referral. Negative colposcopy or biopsy used as the gold standard	24,503 slides classified as NFR by the AutoPap Primary Screening System 300	98.6% of slides reviewed as negative, 0.4% as inadequate, 0.4% as ASCUS-R and 0.12% (31 cases) as ASCUS–SIL+	Not given	Estimate of 99.99% NPV for NFR based 51.6% compliance rate with repeat cytology and 83.3% with colposcopy referral
Biscotti et al.²⁹ sponsored by Cytyc				
Two-armed comparison. Slides received an automated read by the same member of staff 48 days after the manual read. Screeners blinded to the manual read results. Cytological truth adjudication on all non-negative and 5% of negative slides	9550 slides included in the analysis that had been read both manually and by the ThinPrep Imaging System	Not given	Sensitivity for LSIL+ = 79.7% for manual and 79.2% for automated, for HSIL+ = 74.1% for manual and 79.9% for automated Specificity for LSIL+ = 99.0% for manual and 99.1% for automated, for HSIL+ = 99.4% for manual and 99.6% for automated	Not given

TABLE 1 Recent automated cytology publications (*continued*)

Study and design	Comparison groups	CIN detection rates	Sensitivity/Specificity	PPV/NPV
Parker et al.³⁰ sponsored by TriPath Imaging				
Two-armed retrospective masked study. Discrepant results screened by a single cytopathologist	1275 SurePath slides seeded with abnormal. Screened manually with 10% QC and with BD FocalPoint GS Imaging System with NFR slides classed as WNL and review slides screened and triaged to WNL or requiring full screen	58% of HSIL+ slides ranked in Q1 and 83% in Q1 and Q2. All HSIL slides were ranked as review	Not given	Not given
Stevens et al.³¹				
Two-armed retrospective study. Truth was taken as a concordant diagnosis. Discrepant pairs reviewed by a discrepancy panel	6000 conventional slides screened manually and with the AutoPap Primary Screening System using PapMaps	AutoPap identified 35 additional abnormal slides, but missed 92 (94.5% of which were low grade). The difference between low-grade detection in the two arms was significant. AutoPap was equivalent to manual for the detection of high-grade abnormalities. NFR correctly identified 975/986 slides as normal	Not given	Not given
Ronco et al.³²				
Retrospective comparison, with the result of the manual read taken as the gold standard	481 conventional slides read manually then reviewed several months later by the same cytotechnologist using PapMaps	Not given	Sensitivity of PapMaps for selecting abnormal slides = 100% for SIL and 80% for ASCUS	Not given
Confortini et al.³³				
Retrospective comparison with histology obtained from punch and loop biopsies. The worst result was taken used as the gold standard	14,145 conventional slides read manually then rescreened (unless classified as NFR) 3–4 days later by the same cytotechnologist using PapMaps with the AutoPap Primary Screening System	Not given	AutoPap and manual reading are equivalent in terms of sensitivity. The AutoPap had a slightly higher specificity than manual reading	Not given
Wilbur et al.³⁴ supported by TriPath Imaging				
Two-armed retrospective, masked study. Cytological truth adjudication taken as the gold standard	1275 AutoCyte PREP slides (seeded with known abnormal) read manually and with the AutoPap system using the Slide Wizard 2	False-positive rate was 3.8% for AutoPap and 4.4% for manual	Sensitivity of AutoPap for truth determined HSIL+ = 98.4% and manual 91.1%. Specificity of AutoPap = 96.1% and manual 95%	Not given

continued

TABLE 1 Recent automated cytology publications (*continued*)

Study and design	Comparison groups	CIN detection rates	Sensitivity/Specificity	PPV/NPV
Vassilakos et al.³⁵				
Two-armed comparison study using the manual reading as the gold standard	8688 AutoCyte PREP slides read manually and compared with the AutoPap Primary Screening System's review rankings	47.4% of LSIL slides were in Q1, 20.8% in Q2, 10.6% in Q3, 10.1% in Q4, 5.3% in Q5 and 5.8% in NFR 85.2% of HSIL slides were in Q1, 12.7% in Q2, 2.1% in Q3. 0% were in Q4, Q5 and NFR. 84% of all abnormalities were in the highest scoring group along with 100% of HSIL	Not given	Not given

ASC-H, abnormal squamous cells, cannot exclude high grade; ASCH+, ASC-H or greater; ASCUS, abnormal squamous cells of undetermined significance; ASCUS+, ASCUS or worse; ASCUS-R, ASCUS – favouring reactive; ASCUS-SIL+, ASCUS – favouring SIL or worse; CIN, cervical intraepithelial neoplasia; CIN1, cervical intraepithelial neoplasia grade I; CIN1+, cervical intraepithelial neoplasia grade I or worse; HSIL, high-grade squamous intraepithelial lesion; HSIL+, HSIL or worse; LSIL, low-grade squamous intraepithelial lesion; LSIL+, LSIL or worse; NPV, negative predictive value; Pap, Papanicolaou; PPV, positive predictive value; Q, quarter; SIL, squamous intraepithelial lesion; WNL, within normal limits.

There has not been a single rigorous prospective randomised comparison of manual and automated reading which has been specifically powered to show superiority or non-inferiority, in terms of detection of any lesion of cervical intraepithelial neoplasia grade II (CIN2) or worse (CIN2+).

Productivity and cost-effectiveness

Automation has productivity implications for staff time reviewing slides in the laboratory with potential for cost savings in staff time. There are also additional costs associated with the automated equipment. The HTA programme's systematic review concluded that there were productivity gains associated with automation when compared with manual reading with conventional cytology.¹⁴ Studies published since, which have evaluated the cost and productivity implications associated with using the ThinPrep Imaging System and BD FocalPoint GS Imaging system, have suggested that automation results in both increased productivity and increased costs. In all studies the authors found that automation resulted in at least a 50%^{25,26,29} increase in productivity, with the biggest increase reported being 56%.³² A study based in Italy which estimated the costs associated with automated screening concluded that similar costs to manual screening could be achieved only if 60,000 samples per year were processed by the AutoPap Primary Screening System (now BD FocalPoint GS Imaging System) with a 30% NFR rate.³³

There is also a lack of rigorously evaluated data relating to the incremental cost-effectiveness of automated screening compared with manual reading. The HTA programme's systematic review concluded that there were insufficient data to draw any conclusions regarding the cost-effectiveness of automated screening and acknowledged that the papers included in the review did not consider the effect of combining LBC with the technologies.¹⁴

Other current trials of automated screening

Becton Dickinson FocalPoint Guided Screener Imaging System

Currently, there are two ongoing evaluations in the UK involving the BD FocalPoint GS Imaging System. Cervical Screening Wales began an evaluation in 2006 to assess the utility of the BD

FocalPoint GS Imaging System for QC by comparing the 10 FOVs with the current manual QC method. The technology has been used as an additional QC tool. All slides were then manually primary screened. This evaluation has since been extended to include four laboratories across Wales, and was due to be completed by March 2010. A similar evaluation was also undertaken at Derby City Hospital. This study was completed in early November 2009, over 40,000 slides were included. In both studies the slides were sent to Source Bioscience's (formerly Medical Solutions) laboratory in Nottingham for scanning, with the images being read remotely at the trial sites (Wilma Anderson, Source Bioscience Plc., 2010, personal communication).

ThinPrep Imaging System

The Scottish Government Health Department has commissioned a feasibility study of the ThinPrep Imaging System which began in 2008 and aims to compare 40,000 manually read ThinPrep LBC slides with 40,000 ThinPrep Imaging System read slides. The trial has been running in six laboratories – two laboratories processing and reviewing ThinPrep Imaging System slides plus four remote reviewing laboratories.³⁶ The analysis of the first two phases of the study showed that the ThinPrep Imaging System performed as well as manual screening.³⁷ The results of phase 3 of the study involving the Review Scope Plus are described in *Chapter 4*. There are three further feasibility studies taking place in England: one based in Ashford and a second based in Taunton; a QC evaluation study is also taking place in Northampton General Hospital (Glenn Weatherley, Hologic, 2009, personal communication).

The characteristics of further studies involving the ThinPrep Imaging System that are ongoing worldwide are summarised in *Table 2*.

Human papillomavirus testing

Epidemiology of human papillomavirus

It is now universally accepted that HPV infection by so-called 'high risk' types is essential for the process of cervical carcinogenesis.³⁸ There are > 100 different HPV types based on differences in genetic sequences. Of these, > 20 oncogenic types are associated with cervical cancer and, of these, type 16 alone is thought to be responsible for up to two-thirds of all cases.³⁹ Types 16, 18, 31, 33 and 45 are probably responsible for almost 90% of cervical cancers.⁴⁰ HPV including all high-risk types is considered to be responsible for virtually 100% of cervical cancer.³⁸ There are

TABLE 2 Ongoing studies of automated cytology systems

Site	Sample size	Type of study	Control	Intervention
Rheinland Pfalz and Saarland, Germany	20,000	Clinical trial	Manually screened ThinPrep LBC slides	ThinPrep Imaging System
Cologne, Germany	984,509	Retrospective study	890,090 conventional Pap tests	94,419 ThinPrep LBC slides read with the ThinPrep Imaging System
Cerba Laboratories, France	Not known	Internal evaluation	Not known	ThinPrep Imaging System
Leper, Belgium	c.18,000 in first year of study	Evaluation study	Manually screened ThinPrep LBC slides	ThinPrep Imaging System
Abruzzo, Italy	Not known	Clinical trial	Conventional Pap tests	ThinPrep Imaging System and BD FocalPoint Imaging System

Pap, Papanicolaou.

Information provided by Glenn Weatherley, Hologic, 2009, personal communication.

two crucial implications from this. The first is that prevention of high-risk HPV infection will prevent the chain of events that leads to cervical cancer, which has resulted in the production of prophylactic vaccines based on virus-like particles.^{41,42} Beginning in 2008, a prophylactic vaccination programme directed against types 16 and 18 was established across the UK, directed at girls aged 12–13 years with a one-off catch-up programme over 3 years to vaccinate girls aged 14–18 years. The second has been the development of HPV tests which can be used diagnostically. The rationale of these is that women who test HPV-negative are not at risk of cervical neoplasia, and so HPV testing can be used to distinguish HPV-positive women who are at risk from the HPV-negative women who are not.

Current technologies

The first HPV deoxyribonucleic acid (DNA) test to receive FDA approval was the so-called Digene high-risk HPV Hybrid Capture[®] 2 (HC2) (Qiagen, Crawley, UK) test in which a cocktail of 13 high-risk types are tested, which can be used with the liquid cytology medium and which does not require the step of polymerase chain reaction (PCR) to amplify the viral DNA. This test has become the current standard by which emerging tests need to be compared with, in terms of sensitivity and specificity. It is the test currently used in the NHSCSP sentinel sites protocol both for triage and for test of cure for cervical intraepithelial neoplasia (CIN)-treated women, and was adopted for the MAVARIC study (see *Triage and test of cure*). New tests have been developed and others are under development. These tests rely on PCR and most test for DNA, but two test for ribonucleic acid, believed by the manufacturers to achieve greater clinical specificity. Another feature of several new tests is the ability to genotype HPV, with the intention of adding specificity to clinical testing by identifying types such as type 16 which are most strongly associated with high-grade CIN. Some testing kits will combine generic testing for a mixture of high-risk types, with restricted genotyping. Others will rely solely on genotyping. The full potential of HPV testing for cervical screening has yet to be realised.

Triage and test of cure

Triage was employed to achieve maximal detection of underlying CIN2+ in the MAVARIC trial. The use of HPV testing to triage women with low-grade cervical cytology has already been referred to above. Various studies have demonstrated the value of HPV triage in terms of avoiding the need for colposcopy for HPV-negative women as well as increasing the relative sensitivity for detecting CIN2+ compared with repeated cytology.^{43–45} These benefits of HPV triage were demonstrated in the NHSCSP pilot study, although it did result in an increase in rates of colposcopy referral.⁴⁶ These benefits included immediate colposcopy referral, avoiding failure by women to comply with repeat cytology, and increased rates of CIN2+ suggesting that either CIN was being diagnosed more rapidly or triage was more sensitive than repeat cytology, or indeed an element of both.

Test of cure is a term coined for HPV testing following treatment of CIN. A process of long-term cytological surveillance has evolved which has resulted in 10-year annual cytological follow-up in England for treated women found to have CIN2+. Test of cure using HPV testing exploits its high negative predictive value (NPV), to identify the large majority of women who are HPV negative following treatment (who are therefore at very low risk) and allowing them to be returned to routine recall. An assessment of HPV testing as test of cure in the NHS system was undertaken in a recently published study of 900 treated women.⁴⁷ The incidence of cytological abnormality over 2 years among women who were cytology negative and HPV negative at 6 months was sufficiently low to recommend return to routine recall. This would save many thousands of women multiple annual follow-up cytology and this approach has been incorporated into the current sentinel sites protocol. Some samples in MAVARIC underwent HPV test of cure as part of this protocol.

Primary screening

Primary screening using HPV testing is not relevant to the MAVARIC study, which is based on primary screening by cytology. Nonetheless, there is a strong rationale for considering a move to HPV testing in the future based on three considerations:

1. greater sensitivity than cytology
2. the potential for increased screening intervals
3. greater throughput efficiency than cytology.

It should be recognised that the NHSCSP is extremely effective, based as it currently is on cytology. In the future, however, the majority of screened women will have been vaccinated, strengthening the rationale for HPV as the initial test. Published randomised trials indicate that HPV and cytology combined do not increase the overall detection of CIN2+ and CIN grade III (CIN3) or worse (CIN3+) over two successive rounds of screening,⁴⁸⁻⁵⁰ but HPV as a single initial test could be a cost-effective means of screening if suitable strategies can be developed to manage HPV-positive women. Such strategies could combine reflex cytology, HPV genotyping and biomarkers.

Chapter 2

Study design and methods

Aims and objectives of the MAVARIC study

The principal aim of MAVARIC was to compare ‘automation-assisted’ reading with manual reading in cervical screening in terms of effectiveness and cost-effectiveness in the detection of CIN2+, which defines lesions which are treated in the prevention of cervical cancer. This necessitated a randomised design in order to achieve an unbiased comparison and to allow all primary cytology to be read manually as this is the current standard. The first objective required cytology staff to be unaware of whether they were reading a slide which would be read only manually or by automation-assisted backed up by manual reading. The second objective was therefore to create a framework for initial reporting by one method blinded to the result of the other method. The third objective was to accommodate both LBC platforms being used in the NHSCSP: ThinPrep and SurePath. Each of these uses different automated technology – ThinPrep LBC uses the ThinPrep Imaging System and BD SurePath LBC uses the BD FocalPoint GS Imaging System. The fourth objective was to ensure that cytology randomised between manual and automation, and that assessment by the BD FocalPoint GS Imaging System and the ThinPrep Imaging System was comparable in terms of abnormality rates; to achieve this the general practices generating the cytology were stratified by the Townsend Index of Deprivation. A fifth objective was to be able to achieve as rapid and complete a confirmation of clinical outcomes as possible. HPV triage was used to select women with low-grade cytology for colposcopy referral in order to avoid the delays and failure to comply associated with repeat cytology which could lead to non-detection of underlying CIN.

The primary outcome was the relative sensitivity of screening by automated or manually read cytology to detect CIN2+. The relative sensitivity to detect CIN3+ was also determined.

Other outcomes – clinical:

1. The detection rates of CIN2+ and CIN3+ in the manual-only and paired arms.
2. The detection rates [positive predictive values (PPVs)] for each category of cytology including the threshold of borderline or greater and mild dyskaryosis or greater following HPV triage.
3. Relative specificity of screening by automated and manual reading.
4. All of the above comparing the BD FocalPoint GS Imaging System with the ThinPrep Imaging System using BD SurePath LBC and ThinPrep LBC respectively.
5. The reliability of NFR in the BD FocalPoint GS Imaging System in terms of NPV using manual reading in the paired reading as the reference standard.
6. To determine the inadequate rates with both technologies.
7. To determine how automated reading compares with manual reading when used in conjunction with HPV triage of low-grade abnormalities.

Other outcomes – economics and organisational:

1. Comparative throughput and reporting times (for each stage of screening).

2. Detailed cost estimates of the total cost of processing samples at the laboratory and total cost per sample including consideration of inadequate rates and using NFR at different cut-off levels.
3. Estimate of the comparative cost-effectiveness of automated versus manually read cytology using trial data and modelled lifetime costs and effects.
4. Assessment of cytoscreeners' experience and satisfaction with automated systems and the organisational changes that automation would require in implementation.

Trial design

Randomisation of technologies

Initial cluster randomisation between technologies was performed at the general practice level (Figure 1) because it was not feasible for both cytology systems to be used within a single practice. The overall aim of this randomisation was to ensure, as far as possible, that sources allocated to the two systems should have similar population numbers and include women with similar underlying risk. Randomisation was stratified by PCT to take account primarily of variation in Townsend Deprivation Score, but also in ethnic minority composition and screening interval. Sources within each PCT were assumed to have closer levels of these risk indicators. Community clinics were included as a separate stratum. There were a total of nine PCT strata; seven consisted of one PCT where the PCT was expected to contribute large numbers and two strata consisted of more than one PCT (grouped by high or low deprivation) for PCTs where only a small number of women were expected to contribute [i.e. contributing fewer general practitioner (GP) practices]. Owing to the variation in population size, the sources were ordered by decreasing size (number of women) within each PCT and block randomisation of four sources at a time ensured that similar numbers from each PCT were allocated to each of the two techniques.

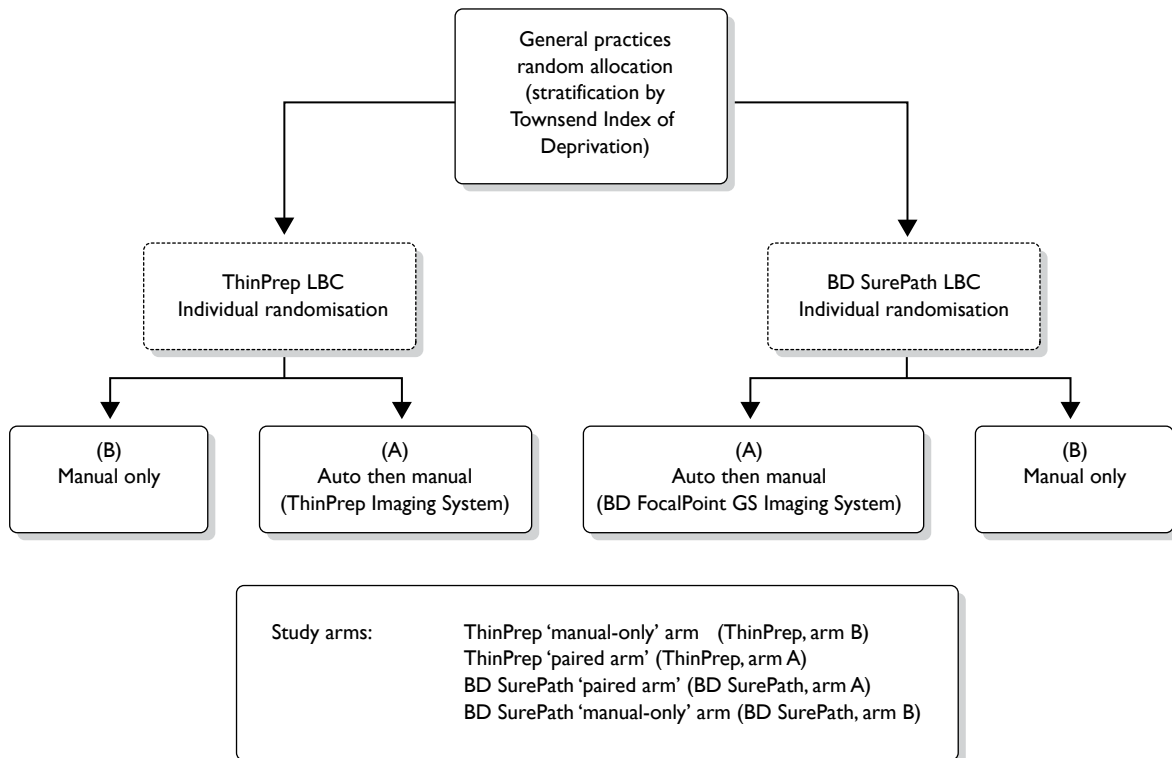


FIGURE 1 Randomisation flow chart.

The numbers 1–6 in *Table 3* show the various possible allocations of two As and two Bs within a block of four where A coded for ThinPrep LBC and B for BD SurePath LBC.

For each PCT stratum the sources were therefore ordered by decreasing size and a series of random digits were generated, with each digit giving the randomisation for a block of four of the sources. For example, in a series of random digits such as 21234, the first number, 2, allocated the first four sources in the PCT to ABAB, the number 1 the next four to AABB and so on until all the sources in the PCT had been allocated to A (ThinPrep LBC) or B (BD SurePath LBC).

Randomisation of slides

The Cancer Screening Evaluation Unit (CSEU) provided two spreadsheets, one for each system, with unique numbers allocated to either the manual-only arm or the paired comparison arm. In the laboratory a query was set up to run on the CliniSys information technology system to pick up all samples eligible for the study and populate the appropriate randomisation spreadsheet. Laboratory randomisation lists were prepared by the laboratory trial co-ordinator and placed alongside the appropriate slides ready for screening.

Inclusion and exclusion criteria

All samples from women attending for screening within the randomised general practices, family planning clinics and colposcopy clinics were initially eligible for randomisation. The inclusion criteria for HPV triage changed part-way through the trial to include only samples from women aged >25 years who were on routine recall to bring the triage protocol into line with the NHSCSP's sentinel sites project.

Analysis of the trial was by intention to treat; however, there were some instances where slides were randomised in error and had to be excluded from the analysis. Slides were excluded for the following reasons:

1. Vault samples taken from hysterectomised women who were no longer part of the cervical screening programme.
2. Subsequent slides randomised from the same woman on early repeat screening.
3. Slides that had to be removed from the automated reading arm because the results were required urgently.

In some instances slides were reported as an automated read failure (ARF) by the imaging systems. When this occurred the final manual result (FMR) was taken as the final automated result (FAR) (see definitions on page 24) for analysis purposes as this reflects what would happen to slides failing an automatic read in real-life practice.

TABLE 3 Blocks used in randomisation

Block number	Source allocation
1	AABB
2	ABAB
3	ABBA
4	BBAA
5	BABA
6	BAAB

Human papillomavirus triage

It had originally been intended to use the Amplicor™ HPV microwell plate (MWP) test (Roche, Basel, Switzerland) because of certain theoretical advantages including increased sensitivity of HPV detection.^{49,51,52} Early comparison with the Amplicor HPV MWP test revealed a number of problems, particularly a higher proportion of HPV-positive tests with ThinPrep LBC compared with BD SurePath LBC. In addition, a significant proportion of BD SurePath LBC samples gave inadequate results (see *Appendix 5*). It was therefore decided to revert to the HC2 DNA test which had been validated by the company for use with both ThinPrep LBC and BD SurePath LBC samples.

Settings and ethics approval

Ethical approval was initially received from Central Manchester Local Research Ethics Committee (LREC) in December 2005 – project reference number 04/Q1407/318 – based on a need for individual signed consent, which was required for HPV triage, which was not at that time part of NHSCSP standard practice.

Research and development approval was received from Central Manchester and Manchester Children's University Hospital NHS Trust, Ashton, Leigh and Wigan PCT, Bury PCT, Heywood, Middleton and Rochdale PCT, Manchester PCT, Oldham PCT, Salford PCT, Tameside and Glossop PCT, Trafford PCT, St Helens PCT, Salford Royal Hospitals NHS Trust and NHS Lothian.

In August 2005, information was sent to randomised general practices and family planning clinics to introduce the trial. Two study sessions were held in 2006 for general practice and family planning staff where they were given the opportunity to put questions to the chief investigator. The trial opened to recruitment on 1 March 2006 in Salford and Trafford, Tameside and Glossop, Oldham and Manchester PCTs. Ashton, Leigh and Wigan PCTs began recruitment in 2007. Women were sent copies of the patient information sheet with their invitation for screening by the local call/recall agencies and surgeries were supplied with copies to give to women who presented opportunistically.

Initial recruitment was slow. Many GPs were unable to recruit women into the trial and gain their consent owing to time constraints within their surgeries and the lack of financial reimbursement. Nurses also reported finding the opt-out system of consenting difficult to work with. Patients were asked to sign an opt-out form to decline either participation in the trial as a whole or to decline a reflex HPV test in the event of a low-grade cytological abnormality. This decision was communicated to the cytology laboratory on the cervical cytology request form which accompanied the sample. Signed opt-out forms were also returned to the laboratory.

Incorporation of trial into the NHS Cervical Screening Programme sentinel sites protocol

In September 2006 the Manchester Cytology Centre agreed to become one of the NHSCSP's sentinel sites for HPV triage, making reflex HPV testing (triage) of low-grade cytological abnormalities routine for all NHS cervical screening samples received at the laboratory. This removed the need for the option to opt out of HPV testing and the LREC agreed that women need no longer be given the opportunity to opt out of the trial. Randomised practices began working to the sentinel site protocol from mid-2007 after consultation with the local Cervical Screening Steering Groups.

Monitoring

The trial was monitored by the HTA programme in July 2007 and by Central Manchester University Hospitals NHS Foundation Trust R&D Office in March 2009, receiving a satisfactory report on both occasions.

Logistical considerations

Processing of samples for cytology testing

The cytology samples were received in the Manchester Cytology Centre in either BD SurePath or ThinPrep LBC vials depending on the system to which the surgery had been randomised. On receipt in the cytology laboratory all samples were allocated a unique identifying number. The ThinPrep samples were processed using the ThinPrep 3000 Processor to produce slides with a printed 14-digit number including the unique identifying number which, after staining with ThinPrep Imaging System stain, were ready to be read on the ThinPrep Imaging System. The use of acetic acid to remove blood from heavily blood-stained ThinPrep LBC samples had to be discontinued as this procedure could affect the validity of the HPV result.

The BD SurePath LBC samples were processed using the BD PrepStain™ Slide Processor to produce slides ready to be read by the BD FocalPoint GS Imaging System. Prior to processing the samples on the BD PrepStain Slide Processor a paper label containing a barcode with the unique identifying number was placed onto the appropriate slide.

All slides were left overnight to dry before being placed into the appropriate imaging system. Both systems produced a print-out of the number of samples processed with any errors incurred during processing; however, the print-out from the BD FocalPoint GS Imaging System could be run only after 120 slides had been processed. The print-outs from both systems were passed to the laboratory co-ordinator to check for errors.

Transporting samples for human papillomavirus testing

The vials from the LBC samples showing low-grade abnormalities were collated at the Manchester Cytology Centre for dispatch to the Specialist Virology Centre in Edinburgh. The samples were anonymised prior to sending by removing the woman's name, date of birth and NHS number. The identifier used for subsequent interaction between Manchester and Edinburgh was the sample number assigned by the Manchester laboratory.

The transfer of samples was performed according to the United Nations' (UN's) regulations governing the packaging of diagnostic and infectious samples UN3373 (packing instruction 650). CitySprint (www.citysprint.co.uk) was the designated courier. The samples were sent on Monday to arrive in Edinburgh on Tuesday and the results of the test sent back to the Manchester Cytology Centre within 4 days. An electronic sheet was sent to Edinburgh with the unique identifying number, date and type of sample.

Processing of samples for human papillomavirus testing

In Edinburgh, samples were accorded an internal sample number for HPV testing. A MAVARIC trial sample identification worksheet and laboratory checklist were completed in the laboratory throughout the testing process. Sample information was entered into a password-protected, bespoke Microsoft ACCESS (Microsoft Corporation, Redmond, WA, USA) database.

For the Amplicor HPV MWP test, nucleic acids were extracted from a 1-ml aliquot using a Qiagen BioRobot 9604 in conjunction with the QIAamp 96 DNA Swab BioRobot Kit and a protocol validated in Edinburgh for use with ThinPrep LBC medium.⁵³ Where weekly sample

numbers were small (<22), nucleic acids were extracted manually using the Roche Diagnostics AmpliLute™ Liquid Media Extraction Kit.

For the HC2 test both ThinPrep LBC and BD SurePath LBC samples were processed according to the manufacturer's instructions. Initial sample preparation involved denaturation with sodium hydroxide rather than nucleic acid extraction. HC2 is a solution hybridisation assay for the qualitative detection of high-risk HPV DNA (types 16/18/31/33/35/39/45/51/52/56/58/59/68) in cervical samples. It uses an oligonucleotide probe cocktail of 13 probes. Hybrids are captured on the wells of a microtitre plate and detected with an amplified chemiluminescent signal. This assay is FDA approved and CE marked.

A positive sample, i.e. indicating the presence of high-risk HPV DNA sequences, was reported, where a relative light unit/cut-off (RLU/CO) measurement was ≥ 3.0 . From 19 February 2008, the protocol was changed to report a positive sample with an RLU/CO ratio ≥ 2.0 to be in line with the NHSCSP sentinel sites protocol. Both these cut-off values deviate from the manufacturer's recommendation of 1.0 RLU/CO, values below which indicate that the HPV DNA levels were below the detection limit of the assay or absent. The reason for the higher cut-off was to achieve additional specificity without significant loss of sensitivity based on data from the ARTISTIC (A Randomised Trial In Screening To Improve Cytology) trial.⁵⁴ From 2 March 2009 any remaining HPV testing was performed in the Manchester virology department along with triage samples from the NHSCSP sentinel sites.

Test data were entered into the local database and results returned to the Manchester Cytology Centre electronically as a Microsoft EXCEL (Microsoft Corporation) password-protected file after each batch run.

Summary of significant changes to the protocol during the course of the study

Significant changes that were made to the protocol throughout the course of the trial are summarised in *Table 4*. The major changes have been described fully in *Statistical analysis, including statistical considerations and Processing of samples for human papillomavirus testing*. The original trial protocol has been included as an appendix (see *Appendix 15*).

Automated cytology methods

Machine set-up

Both companies, Hologic and BD Diagnostics, assessed the site prior to installing the imaging machines. Several changes to the layout of the preparation laboratory and the screening room had to be made to accommodate the installation of the machines.

Training

Staff with varying levels of LBC experience were selected to receive automated screening training. Both companies performed the training, further details of which are provided in *Appendix 6*. Eight medical laboratory assistants (MLAs) were trained in the handling and maintenance of the imaging systems. Eight cytoscreeners and one chief biomedical scientist (BMS) were trained in the use of the automated microscopes and cell morphology recognition. The laboratory trial co-ordinator and two cytopathologists were trained in the handling and maintenance of the imaging systems, the use of the automated microscopes and cell morphology recognition for both systems.

TABLE 4 Significant changes to the protocol during the course of the study

Change to protocol	Months into study	Impact
Two colposcopy clinics (with similar number of referrals) were allocated either ThinPrep or SurePath LBC and were invited to have their samples processed as part of the study	2	Increased the amount of abnormal cytology (and underlying CIN2+) in both arms
Recruitment methods changed to allow staff at GP surgeries to hand women a patient information sheet if they had not received one with their invitation	5	More women were informed about the trial and were able to participate
Manchester Cytology Centre becomes one of the NHSCSP's sentinel sites for HPV triage	16	HPV triage protocol is aligned with the NHSCSP's protocol (i.e. only first borderline and milds triaged). This allowed the need for an opt-out system of consent to be removed as HPV triage had become standard practice and resulted in a more rapid accrual of samples
HPV testing changed to HC2	18	Resolved initial problems with the Roche Amplicor which were resulting in a number of invalid tests on BD SurePath samples
Sample size reduced to 75,000	18	The number of samples in the manual-only arm was reduced to allow the study to finish on time while still achieving the pre-specified number of samples in the paired arm
HC2-positive cut-off changed from ≥ 3.0 RLU/CO to ≥ 2.0 RLU/CO to align the HPV triage protocol with the NHSCSP's sentinel sites protocol	24	This was not thought to have any significant impact on the trial as only 1% of triage samples had an RLU/CO value between 2 and 3
Randomisation ratio changed from 1:1 to 3:1	24	The randomisation ratio was changed in favour of the paired arm to ensure the number of samples specified in the power calculation was achieved. The reduced number of samples entering the manual-only arm remained sufficient to blind the cytoscreeners to the randomised allocation of the samples

Staining

The Becton Dickinson SurePath staining parameters were changed slightly for the study (an additional water wash was added to the process to comply with company recommendations). For the ThinPrep LBC slides, the routine laboratory Papanicolaou (Pap) stain had to be changed to the ThinPrep Imaging System formulation and the Hologic staining schedule had to be followed. The ThinPrep Imaging System formulation stains the cells darker than conventional formulations. The initial proposal was to stain only the trial slides; however, it was recognised that this could cause bias by (a) indicating to the screeners which slides were being read by the automated systems and (b) one of the stains being advantageous in terms of detection of abnormalities. It was therefore necessary to stain all ThinPrep LBC slides received in the laboratory with the ThinPrep Imaging System stain to prevent such bias occurring.

ThinPrep Imaging System stain validation process

In order to validate the ThinPrep Imaging System stain, 100 slides stained with the department's routine Pap stain (of which 25% were abnormal) were screened. A second slide from each of the 100 samples was made and stained using the ThinPrep Imaging System stain. The slides were then processed by the ThinPrep Imaging System and the 22 FOVs were reviewed using the Hologic automated microscopes. The results of the slides read by the ThinPrep Imaging System were compared with the original diagnoses. The reviewers were blinded to the original diagnoses throughout the validation process. Both Hologic and the departmental validators (two cytopathologists and the laboratory trial co-ordinator) classed the ThinPrep Imaging System stain as not significantly different from the routine Pap stain.

All levels of screening staff manually screened 100 ThinPrep Imaging System stained slides to ensure that they had become accustomed to the new staining process. Slides stained with the

ThinPrep Imaging System had to pass the Regional Technical External Quality Assurance. Slides were fed into the first available round and achieved an acceptable result on assessment.

Screening of cytology samples

The slides for automated screening were screened using the review scopes; no marks were made on the slides to indicate any abnormal cells, and the results were entered onto the randomisation list. The list and slides were then passed to another screener for rapid review. The list was removed and passed to the laboratory co-ordinator prior to the slides being placed back into the routine screening in numerical order, thus ensuring that the manual screener was blinded to the result of the automated read. Manual screening (in both arms of the trial) was carried out according to routine laboratory protocols, including the practice of marking areas of interest on the slide. In the paired arm the automated reading was undertaken first, followed by the manual read, and the woman's management was based on whichever reading was the greater in terms of abnormality.

Blinding procedures

One of the principal reasons for the manual reading-only arm was to blind the screener to whether or not slides had received an automated read. The other main reason for the manual-only arm was to provide a comparison with manual reading in the paired arm in order to be able to demonstrate that manual reading in the paired arm was neither superior nor inferior in terms of sensitivity to that in the manual-only arm. Manual screening was performed in the routine laboratory flow of work by a mixture of auto-trained and non-auto-trained cytoscreeners. This created the potential for the same screener to read the slide both manually and on the automated system; however, owing to the large pool of cytoscreeners performing manual screening the chance of this happening was low.

In order to blind the manual screener to knowledge of which slides had been screened using the automated review scopes, no marks were made on the slides during the automated screen. Routinely in the cytology department any abnormal cells found are highlighted by marking the slide above and below the abnormal cells with a coloured marker pen. The Hologic ThinPrep Imaging system utilises a marker pen on the review scope to mark the FOV after the automated screen has been performed, this pen was removed so no marks could be made. The automated screener could add electronic marks as these could be viewed only when using the automated review scopes.

Once the automated read had been performed the result was added to the randomisation sheet, the sheet and the slides were then passed to another screener to perform a rapid screen. The rapid screener then passed the randomisation sheet and slides to the laboratory co-ordinator, the co-ordinator removed the sheets and placed the slides back into the routine screening in numerical order, again helping to blind the manual screeners to which slides had been screened on the automated system.

Review of discordant pairs

Discordant pairs are defined in *Table 5*. A list of eligible discordant pairs was produced by the CSEU for the cytology laboratory. A review of the discordant pairs with a known clinical outcome of CIN2+ was undertaken to assess whether or not the discrepant results were due to a location error by either of the imaging systems or an interpretation error by the cytoscreener assessing the FOVs. Two cytopathologists and the laboratory trial co-ordinator reviewed the FOVs (blinded to both the automated and manual results) and recorded their findings on a mismatch proforma (see *Appendix 7*). A random sample of 10 known CIN2+ concordant

TABLE 5 Discordant pair review criteria

Automation result	Manual result
High grade (moderate /severe dyskaryosis or worse)	Negative/Inadequate
Negative/Inadequate	High grade (moderate/severe dyskaryosis or worse)
Low-grade (borderline or mild dyskaryosis) HPV positive	Negative/Inadequate
Negative/Inadequate	Low-grade (borderline or mild dyskaryosis) HPV positive
NFR	Inadequate, borderline or worse

pairs was added as a control in order to provide blinding as to whether or not slides were from discordant pairs.

When the results of the review had been recorded on the proforma, the results of the initial and final automated and manual reads plus any histology were entered on to the form. A majority view determined the outcome of each discordant pair. The review of the discordant pairs was to determine whether the FOVs were showing the significant cells. The cytological consensus resolved the discordant results and was used to determine whether or not the slide had been interpreted incorrectly on either the automated or the manual reading. In cases where the reviewers agreed with the negative automated reading it was agreed that the machine had not presented any abnormal cells in the FOVs.

Clinical management

Cytology management

All samples were initially reported as per the departmental/NHSCSP protocols for manual reading, but not authorised. The laboratory co-ordinator then recorded the results of the automated screening (which were recorded on separate proformas to blind the manual screening process) onto the laboratory computer system. In the event of a discordant result the samples were taken to peer review meetings for discussion after being reviewed on the automated system by a checker/BMS and a consensus report produced. All results were reported using the British Society for Clinical Cytology 1986 classification (*Table 6*).⁵⁵ Final reports were issued as described in *Table 7*.

Colposcopy management

The management of abnormal cytology is shown in *Figure 2*. Colposcopy was undertaken according to national Cervical Screening Programme clinical practice guidelines. Women with high-grade cytology (moderate dyskaryosis or worse) underwent either a targeted biopsy with subsequent treatment for CIN2+ or an immediate 'see and treat' loop excision. Women with borderline or mild dyskaryosis were referred for colposcopy if they were HPV positive. If they were HPV negative they were returned to routine recall. In triaged cases, a biopsy was not mandated in the presence of normal satisfactory colposcopy. CIN2+ was treated by excision, usually loop excision, and CIN grade I (CIN1) would usually be managed conservatively. The study biopsy result was the higher grade in the event of both a targeted biopsy and subsequent loop excision. All histology was read with the pathologist unaware of the trial arm or LBC type and was reported using the World Health Organization (WHO) and the International Society of Gynecological Pathologists CIN classification system. The pathologist was aware of the grade of cytology. The definitions applied to the colposcopy and histology outcomes for the analysis are given in *Table 8*.

TABLE 6 Cytology classifications

BSCC 1986 ⁵⁵	Bethesda System 2001 ⁵⁶	Definition
Negative	Negative for intraepithelial lesion or malignancy	Normal cytology
Inadequate	Unsatisfactory for evaluation	Low-grade cytology (PPV for CIN2+ generally in the range of 15%–20%)
Borderline nuclear change (includes koilocytosis)	1. Atypical squamous cells ASCUS (undetermined significance), ASC-H (cannot exclude HSIL) 2. Atypical endocervical/endometrial/glandular cells: NOS or favour neoplastic	
Mild dyskaryosis	LSIL	
Moderate dyskaryosis	HSIL	High-grade cytology (PPV of CIN2+ generally in the range of 69%–85%)
Severe dyskaryosis	HSIL	
Severe dyskaryosis query invasive	Squamous cell carcinoma	
Query glandular neoplasia	1. Endocervical carcinoma in situ 2. Adenocarcinoma Endocervical Endometrial Extrauterine NOS	

ASC-H, abnormal squamous cells, cannot exclude high grade; BSCC, British Society for Clinical Cytology; HSIL, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion; NOS, not otherwise specified.
Taken from Kitchener *et al.*⁴⁹

TABLE 7 Results reporting procedures

Manual	Automatic	Reported by
Negative	Negative	Screener/Checker/Senior BMS/Chief BMS
Negative	Abnormal	Medic/Advanced BMS practitioner
Abnormal	Negative	Medic/Advanced BMS practitioner
Abnormal	Abnormal	Medic/Advanced BMS practitioner

Data collection

Transferring data

Data were transferred to the CSEU from a number of sources. Cytological and histological data stored in the Manchester Cytology Centre database (CliniSys Labcentre Laboratory Information System, Chertsey, UK) were downloaded to either a plain text file or Microsoft EXCEL spreadsheet. The file was compressed and encrypted to AES 256 standard using WINZIP version 11 (WinZip Computing, Mansfield, CT, USA). Finally, the encrypted file was sent to the CSEU by secure file transfer protocol (FTP) data transfer. Randomisation data were also sent from Manchester by the same method.

Human papillomavirus results were sent from Edinburgh by secure FTP, but without encryption. Data on exact ranking and quintile for each slide relating to the BD FocalPoint GS Imaging System were stored on hard disk and also backed up on tape. The hard disk was accessed via the internet by BD Diagnostics and archived. The data on tape were also sent to the company by post. From Erembodegem, the unprocessed data files were passed on to CSEU by e-mail, again without encryption. Encryption was thought unnecessary in the latter stages as they did not contain personal identifiers.

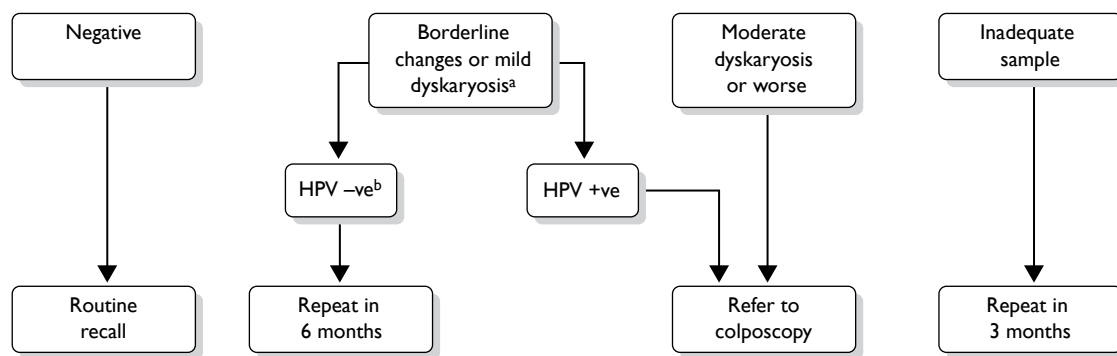


FIGURE 2 Final management cytology protocol. a, After adoption of the Sentinel Sites protocol only women who were on routine recall were eligible for triage. Women on early repeats and follow-up were managed according to national guidelines. b, After adoption of the Sentinel Sites protocol women who had a negative HPV test were returned to routine recall as amendments had been made to the national screening database to permit this management recommendation.

TABLE 8 Colposcopy and histology definitions

Colposcopy and histology outcomes	Definitions
Other cancer	A non-cervical cancer found during further investigations
Adenocarcinoma/squamous cell carcinoma stage 1a+	Invasive cervical squamous cell carcinoma or adenocarcinoma reported as stage 1a or greater according to the FIGO system
CIN3 (squamous cell carcinoma in situ) and CGIN	High-grade pre-cancerous squamous or glandular cell changes on colposcopically directed biopsy
CIN2	
CIN1	Low-grade pre-cancerous squamous cell changes on colposcopically directed biopsy
No CIN/HPV only	No pre-cancerous abnormalities detected on colposcopically directed biopsy
Colposcopy NAD	No abnormalities seen during colposcopic examination

CGIN, cervical glandular intraepithelial neoplasia; FIGO, International Federation of Gynecology and Obstetrics; NAD, no abnormality detected.

Database development

At the CSEU all the data were stored and processed on a secure Microsoft ACCESS database. The database was under the control of the investigators and there was no involvement by either BD Diagnostics or Hologic in the conduct of the study or analysis of the results.

Recording cytology and human papillomavirus results

The data received from the cytology laboratory consisted of the manual reading results, the automated reading results and the final management result (MR). The final MR was the result that determined clinical management (routine recall, triage by HPV test or direct colposcopy referral). The results of the manual readings included up to five readings [the first reading, the rapid review for negative and inadequate first readings, the second reading if required, and further readings by a checker or pathologist/advanced BMS practitioner (AP) for samples with positive cytology]. For each reading, the data received included the test cytology result, whether the screen was full or rapid and the cytoscreener classification (cytoscreener, trainee cytoscreener/BMS, checker, BMS, medic or AP). The data related to the automated readings included the results of the first automated (auto) reading and of the auto rapid review if the first auto reading was negative or inadequate. The data also indicated whether the auto result was used to help determine the final result.

The protocol for determining the FMR and FAR is shown in *Figure 3*.

The following definitions are used:

- The first manual result pre (MR1) and post (MR2) rapid review.
- The FMR, defined as the result of the first reading by a medic or AP, or the result of the last reading that led to the report being signed off if the slide was not seen by a medic or AP (usually a negative finding signed off by a checker or screener as part of the manual process).
- The auto result pre (AR1) and post (AR2) rapid review.
- The FAR, defined as the first medic or AP result from a slide considered as abnormal after the first auto reading or abnormal from the rapid review (post checker), or any negative or inadequate result after the rapid review of negative and inadequate samples that was signed off without being seen by a medic or AP. The three possible pathways are shown in *Figure 3*. Algorithm A was where the automatic read AR1 was negative and the subsequent manual rapid review was negative; the FAR was therefore also negative. Algorithm B occurred where

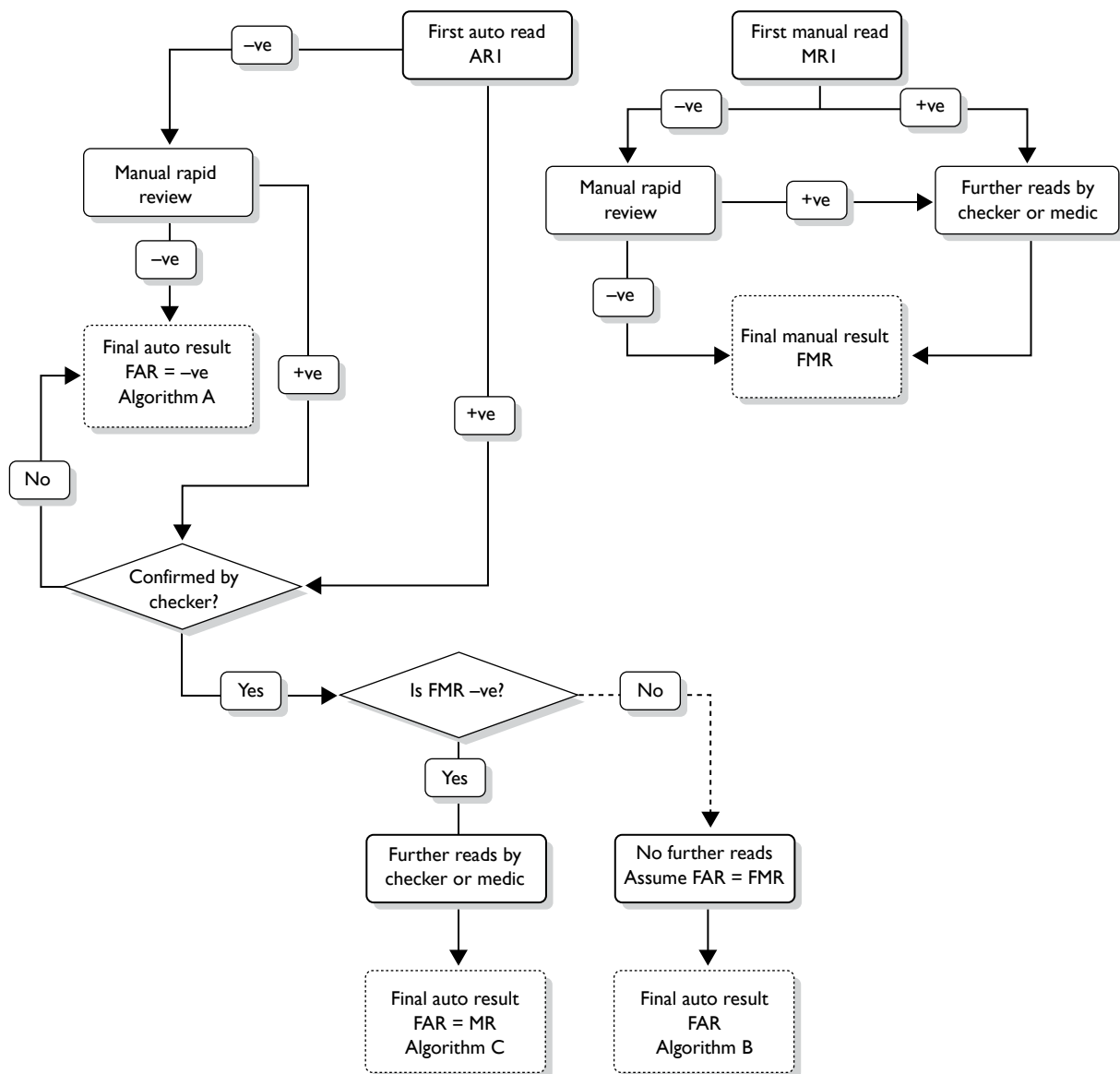


FIGURE 3 Pathways based on actual laboratory application of study.

the auto result after confirmation by a checker was positive, but the FMR was also positive and the auto result was therefore not considered further. Under such circumstances it is assumed that in the real-life situation the slide would have proceeded to be seen by a checker and/or medic and the best estimate of the FAR was therefore assumed to be same as the FMR. Algorithm C was where the auto result (after confirmation by a checker) was positive, but the FMR was negative. Under these circumstances the slide was reviewed again and the FAR was that recorded after this further review, which was also the MR as recorded on the Manchester system.

- The MR was based on the manual result and/or the auto result, whichever was worse, and this determined the woman's management.

In the paired arm, the further reads by a checker or a medic applied to both the manual and auto; hence, the discordant pairs could arise only from a negative or inadequate read or from NFR.

Collecting histology data

Histology results were linked to the cytology results using patient identifiers from the Manchester Cytology Centre database and dates. The histology result was considered to be related to the cytology if the histology date was between 3 weeks and 12 months after the cytology date. In the case of more than one histology result being recorded during that time period, the highest grade abnormality was used. For samples taken at colposcopy clinic visits, the histology result from that visit was used unless superseded by a further result.

Missing data

Data were missing or unobtainable for the reasons given in *Table 9*.

Statistical analysis, including statistical considerations

The sample size calculations were based on a test of non-inferiority of the automated technology in terms of its sensitivity (relative to that of the manual reading) based only on data from the paired observations. Inclusion of the unpaired data increases statistical power, but we chose

TABLE 9 Reasons for missing data

Data type	Reasons why data were unavailable
Cytology	Apart from those samples excluded for technical or clinical reasons (as detailed in <i>Figure 6</i>), all cytology results were obtained
Randomisation	None missing – all cytology samples were associated with a valid randomisation code
Colposcopy/histology ^a	Inadequate biopsy Failed to attend colposcopy Woman left GP or practice area Colposcopy delayed for known reason
HPV data	Follow-up search inconclusive Sample was spoiled before assay HPV test failed No HPV test performed on samples taken at colposcopy clinic No HPV test performed on samples from subjects aged ≤ 24 years
BD FocalPoint GS Imaging System ranking data (quintile information)	Data could not be retrieved by BD Diagnostics from either the BD FocalPoint GS Imaging System via the internet or the backup tapes

^a Results were expected only where the woman was referred to colposcopy, or where the cytology sample was taken at colposcopy clinic. Where no colposcopy outcome was known after 12 months, follow-up was undertaken to attempt to establish the reason for this.

a conservative approach based solely on the paired comparisons. Sample sizes for the paired comparison were determined by the numbers of CIN2+ outcomes (see *Table 10*) needed to evaluate relative true-positive rates (TPRs). When the number of CIN2+ outcomes is about 630, a paired test with a 0.025 one-sided significance level has an 80% power to reject the null hypothesis that the sensitivities are not equivalent [the difference in sensitivities (TPRs) is 0.050 or further from 0 in the same direction] when the expected difference in proportions is 0, assuming that the proportion of discordant pairs is 0.200 (NQUERY ADVISOR, Version 3, Statistical Solutions, Saugus, MA, USA). The sample size estimation is sensitive to the assumed value for the proportion of discordant pairs. It was thought that 0.2 was likely to be the upper limit. The power would increase to about 95% if the proportion of discordant pairs were actually 0.1; in this case the study would have about 70% power to exclude a difference in the TPRs of 0.03 or further from 0 in the same direction. If the proportion of women who are CIN2+ in the population is about 3% we needed to obtain a total of about 46,000 participants in the paired arm to have a probability of 0.975 that it contained at least 630 CIN2+ outcomes. We chose a conservative estimate of 50,000 samples for the paired comparison, and an equal number of unpaired samples (hence a total of $2 \times 50,000 = 100,000$ samples in the trial overall). The above absolute difference of 5% in sensitivity defining non-equivalence between manual and automated reading would require a relative difference in sensitivity of at least 6.5%, assuming a sensitivity for LBC (to detect CIN2+) of 79%.⁵⁷

Owing to accrual problems in the early part of the study the study design was later changed to increase the proportion of samples allocated to the paired arm, in order to ensure that the primary analysis was adequately powered. In June 2007 the sample size for the manual-only arm was reduced from 50,000 to 25,000, reducing the total requirement from 100,000 to 75,000 samples to complete the study. The original design based on the accrual of 100,000 samples required 1 : 1 randomisation, but the later design where only 75,000 samples were required to accrue changed the randomisation to 3 : 1 to achieve the required numbers, with a final paired–manual ratio of 2 : 1. This change retained equal numbers of ThinPrep and SurePath in each arm. The purpose of the manual arm was to ensure that manual reading was reported as it would be if no automated reading was taking place. The distribution of manual reading cytology grades in the manual and paired arms was compared for the two periods before and after the change in randomisation in order to determine whether this change had any impact.

The analysis compares the FMR with the FAR including the results of HPV triage. A ‘positive’ test was one that led to the woman being referred directly to colposcopy (moderate or worse or a result of borderline/mild dyskaryosis accompanied by a positive HPV test). A ‘negative’ test was a result of negative or borderline/mild dyskaryosis with a negative HPV test. The FAR was defined as positive if the cytology result was moderate or severe, or if the cytology result was borderline or mild with a positive HPV test. An FAR of borderline or mild with negative HPV was considered as negative. For borderline/mild samples where the HPV status was not known, the result was taken as positive if the woman was referred to colposcopy. The same applied to the FMR. The main analysis was conducted for each of the ThinPrep Imaging System and BD FocalPoint GS Imaging System arms, based on cytological and histological findings. *Tables 10* and *11* show the final analysis of the paired data. *Table 10* analyses the disease-positive outcomes (defined as CIN2+, essentially all cases requiring treatment). *Table 11* includes histological outcomes that are CIN1 or less (CIN1–), essentially all cases not requiring treatment. The outcome of colposcopy was taken to be the gold standard, available only for those women who were referred to colposcopy. Note that in *Tables 10* and *11*, numbers in enclosed brackets ([D] and [H]) are those that, from the nature of the design, cannot be directly observed, because women who were negative on both the manual and automated reading were not referred to colposcopy.

We estimated the relative sensitivity of automated screening against manually read cytology outcomes to detect both CIN2+ and CIN3+. CIN2+ represents the threshold for treatment and was used to determine true-positives. However, detection of CIN3+ was also used as a clinical outcome in the analysis.

Estimating the relative sensitivity using CIN2+ as disease positive

The sensitivity of the FAR from Table 10 = $(A + C)/(A + B + C + [D])$.

The sensitivity of the FMR from Table 10 = $(A + B)/(A + B + C + [D])$.

Although D is unknown and the absolute sensitivity cannot be calculated, the relative sensitivity can be calculated as $R = (A + C)/(A + B)$.

The 95% confidence interval (CI) is calculated as $[R/y, R \times y]$, where

$$y = \exp 1.96 \sqrt{(B+C)/((A+C) \times (A+B))}$$

A calculation for the relative sensitivity was undertaken for both the BD FocalPoint GS Imaging System and the ThinPrep Imaging System in the paired arm.

Relative specificity rates of screening by automated and manual reading

The relative specificity was calculated in a similar manner to that for the relative sensitivity using Table 11.

The specificity of the FAR from Table 11 = $(F + [H])/(E + F + G + [H])$

The specificity of the FMR from Table 11 = $(G + [H])/(E + F + G + [H])$

The relative specificity = $(F + [H])/(G + [H])$

H is unknown – but a very close estimate can be achieved by assuming that D (CIN2+ not detected by either manual or auto) is 0 so that $H = N - [E + F + G + A + B + C]$, where N is the total number of samples.

The calculations of relative sensitivity and specificity were undertaken for both the BD FocalPoint GS Imaging System and the ThinPrep Imaging System separately.

TABLE 10 Disease positive (CIN2+)

	FAR positive	FAR negative
FMR positive	A	B
FMR negative	C	$[D]$

TABLE 11 Disease negative (CIN1-)

	FAR positive	FAR negative
FMR positive	E	F
FMR negative	G	$[H]$

Further analysis for Becton Dickinson FocalPoint Guided Screener Imaging System and ThinPrep Imaging System involving unpaired arms data and specific data for the Becton Dickinson FocalPoint Guided Screener Imaging System

Additional analyses of secondary outcomes based on the BD FocalPoint GS Imaging System and ThinPrep Imaging System have been performed. A further comparison of the two systems was undertaken using the unpaired data and the combined data. Finally, a further analysis of the BD FocalPoint Imaging System was undertaken to determine the performance of the system with regard to the classification of slides for NFR.

Analysis of manual arm data

The detection rates and PPVs were estimated for the manual-only arm for both the BD SurePath and the ThinPrep LBC systems.

Comparison of paired and manual arms combined

Data from both the paired and unpaired arms were also compared for the two automated tests. Owing to potential confounding factors due to different distribution of the source samples between technologies, these comparisons are restricted to routine samples from women aged 25–64 years. Inadequate rates were examined for both LBC systems. These were also calculated after adjustment for age and reason for test.

Analysis of Becton Dickinson FocalPoint Guided Screener Imaging System ‘no further review’ and quintiles

The results of the ranking of samples by the BD FocalPoint GS Imaging System were also compared with the cytology MR.

The ranking categories are:

- NFR – slides have the highest probability of being normal and may be archived by the laboratory as within normal limits. In total $100 \times A / \text{total slides}$ are classified as NFR. The BD FocalPoint GS Imaging System classifies up to 25% of slides as NFR.
- Review – slides are divided into quintiles of which quintile 1 slides have the highest probability of cytological abnormality. The proportion of slides for each quintile with a final histology of CIN1+ or CIN2+ was analysed. CIN2+ was the most important outcome as this was regarded as disease positive in this study. The study examined the relative sensitivity using the NFR category, but also using the cut-off of quintiles 5, 4, 3, 2 and 1 respectively.
- Process review – indicates a problem such as stain out of limits or slide not scanned.
- Rerun – occurs if tray is rejected.

A comparison of the colposcopy outcomes from quintiles 1–5 was also undertaken to examine the CIN2+ rate in each quintile.

Economic analysis

Introduction

The aim of the economic analysis and organisational assessment was to compare the productivity and cost-effectiveness implications of automated screening technologies with manually read cytology. Automated cytology has a number of implications for the cytology laboratory, and in particular has productivity implications for cytoscreeners due to changes in slide reading practice. A large element of the economic evaluation related to detailed field work in the laboratory to assess the productivity implications of automated cytology versus manual reading

and to assess the broader organisational impact of automated cytology. Changes in laboratory productivity and workload have potential implications for the cost of cytology. In addition, changes in cytology referral rates could affect the total cost per woman screened. To assess the comparative cost-effectiveness of the technologies, a mathematical model was used to assess the long-term cost-effectiveness using cost per quality-adjusted life-year (QALY) gained as an outcome.

Specific objectives of the economic analysis and organisational assessment were as follows:

- to assess the productivity and organisational impact
- to measure costs per slide and per woman screened
- to estimate the cost effectiveness of the alternative technologies.

Measuring productivity and organisational impact

A detailed assessment was made of the productivity implications and broader organisational impact of each automated screening system compared with manual screening. Productivity of laboratory staff, including both cytoscreeners and laboratory assistants, was measured using a number of different approaches throughout the trial. Technical differences between the technologies have productivity implications for both the duration of each activity in the screening pathway and the necessity/probability of undertaking different activities.

In addition to the preparation required for reading slides manually, automated cytology requires 'loading' and 'unloading' of slides onto either the BD FocalPoint GS Imaging System or the ThinPrep Imaging System. The differences between the two automated systems as described earlier (see *Introduction*) have implications for the time taken to undertake primary screening. In summary, a number of different factors that could potentially affect productivity were measured during the trial:

- staff time to load and unload automated equipment
- average time for primary screening (time and motion)
- average number of slides screened per day (daily record sheet)
- average workload per year
- average total time per slide for reading [including checking/medic (or AP) review]
- other organisational factors potentially influencing productivity.

Loading and unloading time of equipment in the automated arm

In addition to the preparation time for manual reading, automated cytology slides also need to be loaded onto the automated machines and then unloaded. To determine the additional time involved, record sheets were developed to measure staff workload (see *Appendix 1*). These record sheets were completed by laboratory staff over a series of batch runs, to estimate the additional time involved by staff loading and unloading samples for both of the automated technologies.

Average primary slide reading time (time and motion)

Automated cytology changes the way in which cytoscreeners read slides. Instead of reviewing the whole slide, cytoscreeners review only specific marked fields on the slide. To compare the staff time involved across the technologies, following initial piloting work, a time-and-motion study was designed. Cytoscreeners recorded timings for reading consecutive slides on a paper form (see *Appendix 2*). Timings were undertaken by each cytoscreener and measured using stop watches at his or her workstation. Initially, timings were recorded after staff had been reading slides for approximately 6 months. A further, much larger time-and-motion survey was conducted near the end of the trial, when staff had been screening with automated cytology for about

3 years. Timings included the time for reviewing the slide. Within the time-and-motion study, administration times were recorded only in the manual arm. These costs were assumed to be the same across each of the LBC systems.

Average number screened per day (daily record sheet)

While the time-and-motion studies give a valuable insight into the average time taken to read individual slides, an important consideration for cytology laboratories is how this might translate to the number of cytology staff required. Cytoscreeners undertake a number of activities during their working day, and primary screen slides for only up to 4 hours (5 hours including rapid reviews). Within the screening period there are also natural breaks between reading individual slides. Hence, in addition to the time-and-motion studies, a questionnaire (see *Appendix 3*) was devised to record the cytoscreeners' overall workload and to help estimate the overall implications of automated technologies for productivity in terms of the actual number of slides screened per day. This survey was undertaken after cytoscreeners had been reading automated slides for over 3 years. The questionnaire recorded the number of hours cytoscreeners work on different activities and number of slides processed over a 5- to 6-week period.

Average total reading time per slide

The total time for slide reading is dependent on a number of factors. Firstly, in the automated arm some slides may not be available for automated reading owing to an ARF and therefore have to be read manually. Secondly, with the BD FocalPoint GS Imaging System, up to 25% of slides are classified by the automated equipment as not requiring review by a cytoscreener – NFR. Furthermore, as well as the time per slide for the primary screening detailed above, automated technologies could potentially affect the rates of referral of slides for 'checking' or review by pathologists/APs with time-related implications for these staff.

To allow for these factors, average total reading time per slide for was estimated by adjusting the average time duration of different stages of slide reading activities, with the probabilities of ARE, NFR, checking and onwards referral which were obtained from the clinical trial database held at the CSEU. The average time required for primary screening and rapid review was obtained from both the workload and time-and-motion surveys. Time duration for checking and secondary screening was taken from an earlier study in the same laboratory.⁴⁹

Average workload per year (daily record sheet)

To assess the overall workload for cytoscreeners per year and the potential implications for the number of cytology staff, we also estimated annual cytoscreener workloads based on the daily record sheet data. Using this weekly information, and assuming that cytoscreeners work 43 weeks a year, the annual workload was estimated for each technology.

Other organisational factors potentially influencing productivity

Throughout the trial a detailed record was made of any other factors that could potentially influence productivity, such as days of machine downtime or other organisational factors. Utilising the BD FocalPoint Slide Profiler as a stand-alone piece of equipment making use of the 'quality control review' and 'no further review' options combined with manual screening also has potential productivity implications. We evaluated the time implications of utilising the BD FocalPoint Slide Profiler as a stand-alone piece of equipment combined with manual reading. We estimated the time savings associated with the NFR option and for slide reading by quintile.

The QC method used in the paired arm of the trial was the same as manual reading, which is a rapid review on all negative and inadequate samples. We assessed a potential further option when utilising the BD FocalPoint Slide Profiler as a stand-alone device of dropping the rapid review for routine samples where slides are determined as requiring NFR. We also explored the time

implications of not primary screening or rapid reviewing the slides in the lowest quintiles which were least likely to have abnormalities.

Staff satisfaction and preferences could also affect productivity. Following an initial focus group discussion with staff, a questionnaire was developed to assess staff satisfaction with using the different types of automated equipment compared with manual reading (see *Appendix 4*). Staff preferences between the two technologies were also obtained for different aspects of screening. In addition, staff were asked (1) if they found it easier to concentrate using the automated system compared with manual reading; (2) if work was more challenging using the automated reading system; and (3) if work was more monotonous using the automated reading system than with manual reading.

Measuring costs

The cost analysis was carried out from the NHS perspective. It is unlikely that the technology would have significant cost implications for social services or patients. All costs refer to 2007, adjusted when required to that year using the Hospital and Community Health Service (HCHS) pay and price index.⁵⁸ Within the trial it was possible to observe detailed differences in screening costs both between automated and manual reading and across the different technologies. However, as this was a diagnostic accuracy trial and the same woman was screened with both types of reading, it was not possible to observe directly the downstream costs of individual patients as they were screened both ways and events could be triggered by either technology.

Unit costs were estimated for each cost-generating event and were combined with data from the productivity assessment and epidemiological data in order to estimate the total costs per slide and per woman screened in each arm. Unit costs were derived from observational studies (mostly undertaken specifically for the MAVARIC trial), existing tariffs and contracts, as well as from published sources.

Cytology laboratory costs

Total costs per slide were calculated by combining the cost of preparation and slide reading equipment with the costs of slide reading.

Cost of preparation and slide reading equipment

The unit cost of LBC cytology test preparation and slide reading equipment in the manual arm covered the following: costs of the LBC slide preparation system (BD PrepStain Slide Processor and ThinPrep 3000 Processor), maintenance, LBC consumables, cost of staff processing time and microscope costs. The number of microscopes required was identified via consultation with laboratory staff and the purchase cost was written off over 5 years. Costing of LBC equipment/consumables was based on 5-year contract lease prices and the assumption that equipment would be used at the recommended annual capacity. The contract prices for manual equipment and consumables were provided by the corresponding manufacturers and based on existing contracts between manufacturers and the NHS Purchasing and Supply Agency. To maintain confidentiality over the contract prices, we have presented the costs in combination with the staff costs of slide preparation.

The cost of preparation with the automated technologies included the same preparation costs as outlined above for manual reading, plus the additional cost of equipment, maintenance and staff time associated with the automated technologies. Equipment costs for both automated technologies were indicative, and as with manual equipment were based on 5-year rental contracts. In the cost analysis, we present costs on the basis of a laboratory processing at the

maximum capacity per year for each technology. The indicative prices of the BD FocalPoint GS Imaging System were based on rental of one BD FocalPoint GS Imaging system with five BD FocalPoint GS Review Stations plus full maintenance contracts for 5 years. This system uses existing microscopes and was costed as above.

The indicative price of the ThinPrep Imaging System was also obtained from the manufacturer. This price is based on rental of one ThinPrep Imaging System plus three guided screener workstation microscopes over 5 years based on their recommended annual capacity. With both automated technologies it is necessary to load and unload slides onto the automated machines. The additional staff time for loading and unloading slides was estimated using record sheets as outlined in the productivity section. The cost of staff time was valued using the unit cost of staff time described above.

Costs of slide reading

Data from the productivity surveys were used to determine the grades of staff undertaking the different screening tasks: primary screening, rapid review, checking and secondary screening. To attribute salary cost for each activity to the different grades of laboratory staff, the mid-scale point for the corresponding band in the Agenda for Change salary structure was applied.⁵⁸ Salary costs included qualifications and NHS employers' costs (that is, the employer's national insurance contribution plus 14% of salary for employer's contribution to superannuation).

Average staff costs per slide were determined by combining the data on the staff cost associated with each screening activity with the probability of each event in the screening pathway. Primary screening costs in the BD FocalPoint GS Imaging System arm were adjusted to incorporate the fact that some slides would not require primary screening because of the NFR option. Where there was an ARF it was assumed that these slides would be read manually. Further analyses were also undertaken to assess the staff costs associated with NFR and by quintile in the BD FocalPoint GS Imaging System arm. These data were used to estimate the costs of utilising the BD FocalPoint Slide Profiler as a stand-alone device combined with manual reading with or without rapid review for slides identified as requiring NFR. We also explored the cost implications of not primary screening or rapid reviewing the slides in the quintiles which were the least likely to contain abnormalities.

Primary care, human papillomavirus testing and colposcopy costs

Average total cost per woman screened included primary care, cytology costs, HPV testing and colposcopy costs. Probabilities of different screening events and related care were combined with appropriate unit costs. The clinical database was used to determine the final result in each arm to model the costs of downstream events. It was assumed that where there was an inadequate screening result these women would have a further sample taken by their GP. Where the final result was borderline or mild, HPV testing costs were included. Colposcopy was costed according to grade of CIN diagnosed at the colposcopy clinic. Unit costs of primary care, HPV testing and colposcopy costs were determined as follows:

- *General practice/community clinic unit costs* The unit cost for obtaining a cervical sample using the LBC technique included the time for taking the sample by a doctor or nurse, the cost of the materials and the cost of transportation of the vial containing the sample to a cytology laboratory. As both manual and automated cytology involve the same methods for collecting samples there was no reason why automated technology would change the unit costs in primary care, and so these costs were obtained by reviewing and updating earlier studies.
- *HPV testing unit costs* The cost of HPV testing includes equipment, consumables and staff costs. Costs were based on the HC2 assaying technique. Equipment costs were based on

a manual preparation system as used in the Specialist Virology Centre, Edinburgh. The costing of equipment was based on a 5-year lease cost. Indicative prices for leasing HC2 systems were provided by Digene according to a range of assumptions over volume of sales. However, as the prices were provided in confidence, the unit costs are presented inclusive of consumable and staff costs. For costing purposes it was assumed that each assay run would be at full capacity and all the wells would be used. The amount of time spent by technical staff in operating the system was derived from observational field work at the Edinburgh virology laboratory. Staff costs were then estimated based on the mid-point of the BMS pay rate band.⁵⁸ Given the distance between the cytology and virology laboratory, transport costs were also estimated.

- *Colposcopy and histology unit costs* Unit costs of colposcopy were derived from NHS *Payment by results* average tariffs.⁵⁹ The costs of biopsy, histology outcome and related treatment were obtained from a recent large costing study⁶⁰ which reported costs of treatment by cytology and histology grades for 600 women with first abnormal cervical screening result who had been recruited from six specialised gynaecology/colposcopy clinics in England and Wales.

Estimating cost-effectiveness

Cost-effectiveness was estimated utilising the within-trial results on the cost per case detected. We also estimated the lifetime cost-effectiveness of alternative technologies utilising a mathematical model.

Within-trial cost-effectiveness:

We combined data on the total cost per woman and clinical outcomes to estimate the incremental cost per case of CIN2+ and CIN3+ detected on automated compared with manual reading. To assess the uncertainty in the estimates we utilised a non-parametric bootstrapping procedure: we randomly sampled 5000 slides (with replacement) 5000 times from the trial data and for each sample estimated the mean costs and effects. Results were then plotted on a cost-effectiveness plane. Cost-effectiveness acceptability curves were also generated, reflecting the probability that options were cost-effective given different willingness-to-pay thresholds for CIN2+ and CIN3+ cases detected.

Analyses were conducted both to assess the cost-effectiveness of automated reading versus manual reading, and to estimate the cost effectiveness of alternative options for utilising the BD FocalPoint Slide Profiler as a stand-alone device.

Modelling beyond the study end points

The aim of the analysis was to compare the lifetime effects, costs and cost-effectiveness (using life-years saved as the primary outcome measure) of using LBC alone compared with using automated cytology screening. The evaluation used the final results of the MAVARIC trial, including both the clinical results and the cost data. As there are no long-term follow-up data on cancer outcomes from the trial, a mathematical model was used to estimate lifetime effects, costs and cost-effectiveness. We used a model adapted from previously published models.^{61–64} The model was a Markov simulation model with two components, the first dealing with HPV natural history, progression to CIN and cervical cancer, and the second dealing with screening and treatment. The model provides a comprehensive map of the current screening pathways for managing cytology results and treatment following referral to colposcopy.

The simulation follows a cohort of women, with transitions between states occurring annually according to age-dependent probabilities. The model predicted the lifetime costs and effects of alternative strategies from age 10 years to age 84 years (inclusive). For this analysis the model was adapted to run using 6-month rather than 12-month cycles. Full details of the assumptions and

parameter sources are given below and in *Appendix 12*. The analysis was conducted from a health service perspective, excluding any costs or savings that might be incurred by patients or their families, in line with current UK recommendations. Inclusion of any such costs would be unlikely to materially affect the results or conclusions.

Screening strategies

Current UK screening protocols using LBC (the standard technology in the UK) were compared with a strategy of LBC in conjunction with automated cytology. In both strategies, women with moderate or worse cytology results were referred directly to colposcopy; inadequate cytology samples were retested (this is assumed to occur immediately for modelling purposes, and to result in an adequate sample); and women with normal results returned to routine screening. In line with the trial, it was assumed that women with low-grade abnormalities would be tested for HPV using HC2. Women were referred to colposcopy if the reflex HPV test was positive, otherwise they were returned to routine screening.

Natural history

The probability of transitions between pre-invasive health states (well, HPV only, CIN grades I–III) and invasive cancer states (stages I–IV) and the probability of symptoms in an unscreened population were based on a previous natural history model, updated to reflect a 6-month time frame as shown in *Figure 4*.^{62,63}

All transition probabilities were calculated for a 6-month time frame (reflecting screening protocol time frames). The model used data from the West Midlands cancer registry on invasive cancer survival and mortality from other causes.⁶⁵ The following assumptions were made: (1) all cases of pre-invasive and invasive cervical cancer begin with a HPV infection; (2) annual cervical cancer-specific mortality 6–10 years after diagnosis is assumed to be the same in the fifth year after diagnosis; and (3) women who survive for 10 years after diagnosis and treatment for cancer are assumed to have the same life expectancy as women in the general population.

Attendance

We used registry data from Oxfordshire to estimate the cumulative rescreened proportion at various times after a negative smear for women who appeared on the register.⁶⁴ We used age-specific data on the percentage of eligible women who attended at least once in a 5-year period in England (2007–8),⁷ to adjust these data and derive an age- and interval-specific probabilities of women attending for routine screening. This allows the model to take into account non-attendance, early rescreening, late rescreening, and screening in ages outside the target age range for screening. Attendance rates for screening and colposcopy were based on an earlier study⁴⁶ and on routinely collected screening data for England (2007–8).⁷ It was assumed that if women did not attend for colposcopy then they would only be recalled for screening at the next round.⁶³

Effectiveness of screening and colposcopy

Data from the trial inform only on the relative sensitivity between manual and automated reading. For the cost-effectiveness modelling it is necessary to have estimates of the true sensitivity and specificity. True sensitivity equals the probability of testing positive given true underlying disease and true specificity is the probability of testing negative given that there is no underlying disease. Sensitivity and specificity are defined for a given disease threshold and, in the case of cytology, a given test positive threshold.

As management of cytology results varies based on the cytology result, our cost-effectiveness model required the probability of a given cytology test result, given a true underlying health state. These probabilities were derived from the outcome data from women who attended colposcopy in the trial. Within the trial, there are colposcopy outcomes data only on women whose cytology

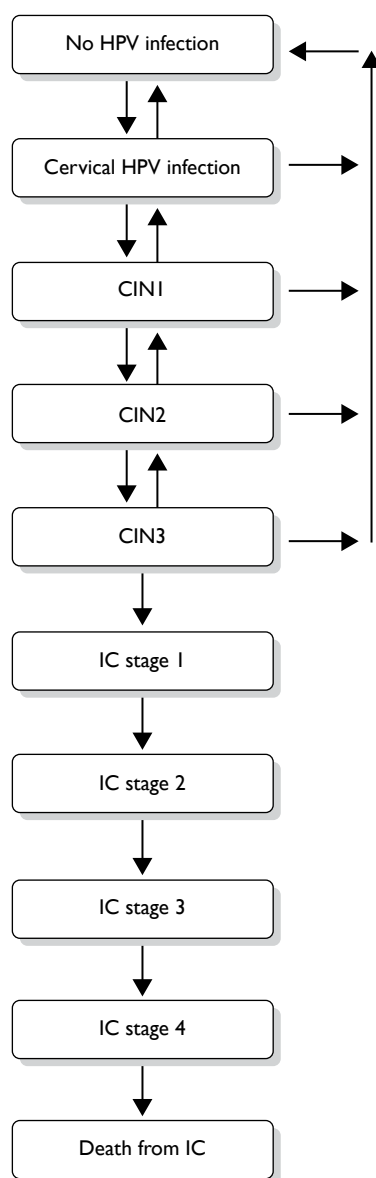


FIGURE 4 Health states defined by the natural history model and the potential transitions between states. IC, invasive cancer.

results (with either manual or automated reading) were moderate or above or who were borderline or mild and then had positive HPV test results, and attended colposcopy. The true underlying disease status of other women (those with negative cytology or a negative HC2 result, or who did not attend colposcopy) is unknown from the trial data. To inform the probability of underlying disease given negative LBC results, we utilised data from the ARTISTIC trial. This large trial was chosen as it was also undertaken in Manchester and reflects current screening practice using LBC.⁴⁹ The probability of disease in women with negative HPV test results was informed by estimates for HC2 positivity rates for each underlying health state. The probability of disease in women who did not attend for colposcopy was assumed to be the same as in those women with the same set of test results who did attend. As the ARTISTIC data related to manually read LBC, we made assumptions for automated LBC which maintained the relative test sensitivity and specificity observed in MAVARIC (see *Tables 35 and 36*). During the sensitivity analysis we investigated two alternative sets of test characteristics, where (1) automated LBC

had the worst performance relative to manual LBC consistent with the MAVARIC trial findings, and (2) automated LBC had the best performance relative to manual LBC consistent with the MAVARIC trial findings (see *Appendix 12, Table 97*).

A previous review of the international literature was used to define a feasible range for HC2 positivity rates for each underlying health state (i.e. true disease state) in the model.⁶⁶ We have assumed that for all screening and diagnostic tests the sensitivity for cancer was the same as that for CIN3. Estimates of the sensitivity and specificity of colposcopy and of CIN treatment pathways and recurrence were undertaken and have been described in detail in a separate report.¹² It was assumed that treatment for CIN is 96% effective by 6 months, and that in women whose treatment was successful, 84% return to a well state with no HPV infection, based on the findings of a systematic review.⁶⁶

Costs and utilities

Data from the MAVARIC trial were used for the unit costs of LBC, automated cytology and HPV testing. As costs varied by manufacturer of the test technology, the prices for manual and automated LBC were set to be the average cost of the two available technologies. Similarly, HPV test cost varied depending on whether the original LBC preparation was ThinPrep or BD SurePath, therefore an average cost was used. Data on utilities and treatment cost were obtained from the literature.^{67–69} To convert from these 12-month values to 6-month values, we assumed that the full 12-month disutility occurred in the first 6 months, with no disutility in the second 6 months. The following assumptions were made:

1. Disutility associated with a false-positive was applied to women with LBC results of moderate or greater, or with a borderline/mild and a positive HC2 result with no histological confirmation of CIN.
2. Costs for no CIN were applied to women who attended colposcopy, but had no histological confirmation of CIN.
3. In both cases women with no histological confirmation of CIN includes women in whom no biopsy was taken owing to negative/unsatisfactory colposcopy.
4. Women who are referred to colposcopy but do not attend have no additional cost or disutility applied in that cycle (only costs associated with cytology/HC2).

Model fitting/calibration

The natural history model was adapted from previously published models.^{61,62,64} Predictions from this model for age-specific and age-standardised rates of cancer incidence in an unscreened population closely matched rates seen in 25 developing countries without significant levels of cervical screening (data published in the International Agency for Research on Cancer's *Cancer incidence in five continents*⁷⁰).

The output of the combined natural history and screening model was compared with:

1. the age-specific and age-standardised cervical cancer incidence in England (2006)⁷¹
2. the age-specific and age-standardised cervical cancer mortality in England and Wales (average 2001–5)⁷²
3. the age-specific prevalence of high-risk HPV by HC2 in the ARTISTIC study population⁷³
4. distribution of cancer stage at time of diagnosis in the West Midlands (2006).

The natural history model was adapted for a 6-month model cycle and adjusted to be consistent with UK data for cancer incidence and HPV prevalence. The results of the model fitting are presented in *Appendix 12*.

Analysis

Following current UK recommendations, future costs and future benefits were discounted at 3.5% for the first 30 years, commencing after age 10 years, and 3% thereafter.⁷⁴ To estimate the comparative cost-effectiveness between the strategies, the strategies were first ranked in ascending order of effectiveness. Options that were dominated (that is, less effective and more costly than an alternative) and strategies that were extended dominated (that is, inside the cost-effectiveness frontier) were excluded. The incremental costs, effects and resulting cost-effectiveness ratios (incremental costs divided by incremental effects) were then calculated for the remaining strategies. To test the effect of parameter uncertainty, one-way and probabilistic sensitivity analyses were conducted. In the one-way sensitivity analysis each parameter was varied in turn using the minimum and maximum parameter estimates (see *Appendix 12, Tables 96–98*).

Selection of end points

Two clinical outcomes were chosen as study end points – detection of both CIN2+ and CIN3+. CIN2+ represents the threshold for treatment within the NHSCSP and can therefore be used to determine true-positives. CIN3+ is another valid outcome in terms of protection against invasive cancer and death from the disease.

Chapter 3

Results

Summary of randomisation

There were two randomisation processes, involving firstly the allocation of sources to each LBC preparation and secondly the randomisation of samples from these sources to the manual or paired arms of the study. Between 1 March 2006 and 28 February 2009 73,266 samples were obtained. The accrual curve is shown in *Figure 5*. All of the samples were randomised, initially in a ratio of 1 : 1, to either manual or paired reading. In January 2008 the randomisation ratio changed to 1 : 3 as described in the methods. Following 429 exclusions (see *Figure 6*), there were 24,566 (33.7%) samples in the manual arm and 48,271 (66.3%) in the paired arm.

There were initially 212 GP/community clinic sources eligible for randomisation. However, sources that were identified as being linked (e.g. two GPs within the same practice or a clinic operating on two sites) and had to be allocated to the same preparation were combined into a single unit, giving 174 randomisation units, of which 89 were randomised to ThinPrep and 85 to BD SurePath LBC. Of these, 124 (71%) contributed samples to the study as randomised and 22 (13%) sent samples using the alternative preparation, owing to the contractual arrangements of

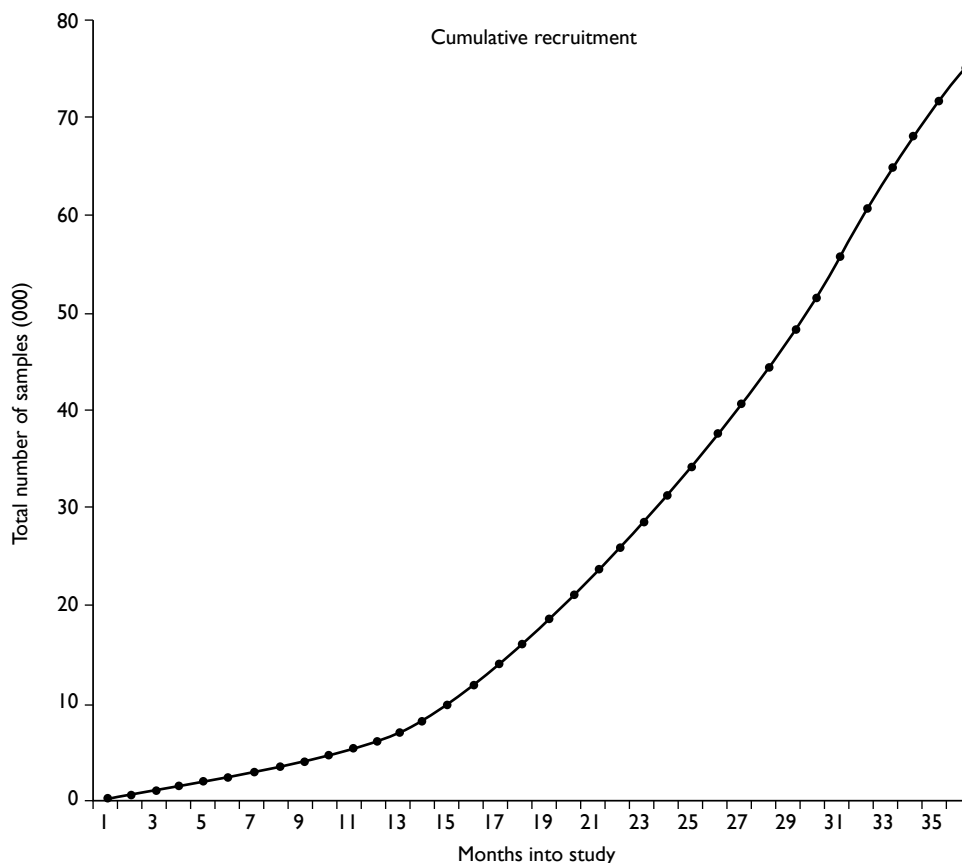


FIGURE 5 Accrual curve for samples randomised into the study.

the PCT. Of the remainder, it is possible that some contributed samples using a different source code to that originally supplied and are therefore included as 'non-randomised' sources. The non-randomised sources also include two colposcopy clinics that were included in order to increase the numbers of high-grade cytology. In addition, some further GP/community clinic sources were added after the randomisation had been completed. Many of the non-randomised sources contributed only a small number of samples. There were therefore a number of limitations to the randomisation process, and the success of the randomisation can be measured only by the mean Townsend Deprivation Score within the arms of the trial. The numbers of samples received are summarised in *Table 12*.

Table 13 shows that for all ages the mean Townsend Deprivation Score was similar for both arms and both LBC systems with an overall value of 3.8. The largest difference was between BD SurePath in the paired arm with a mean value of 3.64 and ThinPrep in the manual arm with a mean value of 3.99. The mean age was also similar for all groups. Despite the constraints imposed on the study, in practice, the randomisation was successful. The data restricted to ages 25–64 years only, also shown in *Table 13*, are almost identical.

The allocation of cytology slides following randomisation is shown in *Figure 6*. There were 429 slides (0.58%) excluded, the majority because they were 'vault' cytology, i.e. vaginal samples in the absence of a cervix, post hysterectomy.

Most of the samples (82.5%) were derived from routine cervical screening, 10.6% were repeat samples requested following a low-grade cytological abnormality and 6.2% were taken at a colposcopy clinic where there had not been a prior study sample from that woman (*Table 14*). This source was only feasible initially as colposcopy samples subsequently came from women

TABLE 12 Number of samples received

Source randomised to	Sending	Number of sources	Number of samples	%	Cumulative %
ThinPrep	ThinPrep	64 ^a	19,411	26.65	26.65
BD SurePath	BD SurePath	60 ^a	22,656	31.11	57.76
ThinPrep	BD SurePath	12 ^a	4,799	6.59	64.35
BD SurePath	ThinPrep	10 ^b	4,676	6.42	70.77
Not randomised ^c	BD SurePath	34 ^d	8,144	11.18	81.95
Not randomised ^c	ThinPrep	82 ^d	13,151	18.06	100
Total			72,837	100	100

a Randomised units.

b Includes two units each with two sources sending some BD SurePath LBC samples.

c Source code not identified as being randomised.

d Sources.

TABLE 13 Mean and standard deviation (SD) for Townsend Deprivation Scores (all ages)

Arm	Preparation	Number		Mean Townsend Score (SD)		Mean age, years (SD)	
		All ages	25–64 years	All ages	25–64 years	All ages	25–64 years
Manual	SP	12,195	11,502	3.84 (3.08)	3.81 (3.09)	39.3 (10.8)	39.8 (10.0)
Manual	TP	12,371	11,717	3.99 (3.26)	3.97 (3.26)	38.8 (10.6)	39.4 (10.0)
Paired	SP	23,404	22,282	3.64 (3.13)	3.61 (3.14)	39.3 (10.5)	39.7 (9.9)
Paired	TP	24,867	23,717	3.85 (3.27)	3.83 (3.28)	38.8 (10.5)	39.2 (9.9)

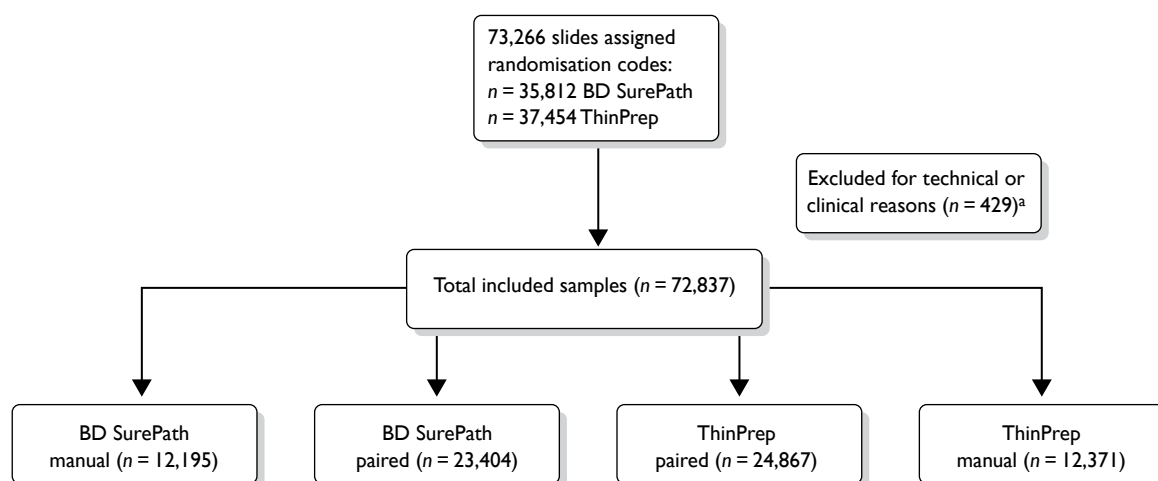


FIGURE 6 Randomisation of cytology slides (March 2006 to February 2009 inclusive). a, 242 vault cytology samples; 117 routine samples taken at colposcopy clinic, but without explicit consent for the samples to be involved in the study; 21 slides were removed from the automated process because the results were required urgently; and 49 slides were excluded owing to procedural errors, i.e. the slide was not auto screened or the result not recorded. Total samples in the paired arm: $n = 48,271$. Total samples in the manual-only arm: $n = 24,566$.

TABLE 14 Source of the randomised samples

Source of samples	BD SurePath		ThinPrep		Total	%
	Manual	Paired	Manual	Paired		
Routine ^a	9765	19,331	10,207	20,799	60,102	82.5
Other/colposcopy clinic ^b	988	1576	657	1320	4541	6.2
Other ^c	1363	2327	1440	2556	7686	10.6
Missing	79	170	67	192	508	0.7
Total	12,195	23,404	12,371	24,867	72,837	100.0

a Defined as the following codes (assigned by sample takers): 1 – routine call, 2 – routine recall, 7 – previous inadequate, 8 – opportunistic.

b Defined as the following codes (assigned by sample takers): 5 – previous biopsy/treatment, 6 – annual tests.

c Defined as the following codes (assigned by sample takers): 3 – clinically indicated, 4 – previously abnormal, 9 – other.

whose initial screening sample, prior to colposcopy referral, had already been included in the MAVARIC study.

Comparisons between results in the manual-only arm and those from the manual reading in the paired arm were restricted to routine screening samples as there was a larger proportion of non-routine samples in the manual-only arm. This arose because of the change in randomisation ratio. At the beginning of the trial women were recruited from two colposcopy clinics when samples were being randomised in a 1 : 1 ratio. Further into the trial when randomisation to the manual-only arm was 1 : 3, women attending colposcopy had already had samples taken in primary care included in the trial and such samples became ineligible. In addition, a higher proportion of BD SurePath LBC samples were taken at colposcopy clinics owing to one clinic recruiting a larger number of patients. As a result of these factors the manual-only arm contained disproportionately large numbers of colposcopy clinic and BD SurePath LBC samples. Therefore, comparison of cytology results between the two technologies was also restricted to routine samples.

The consolidated standards of reporting trials diagram

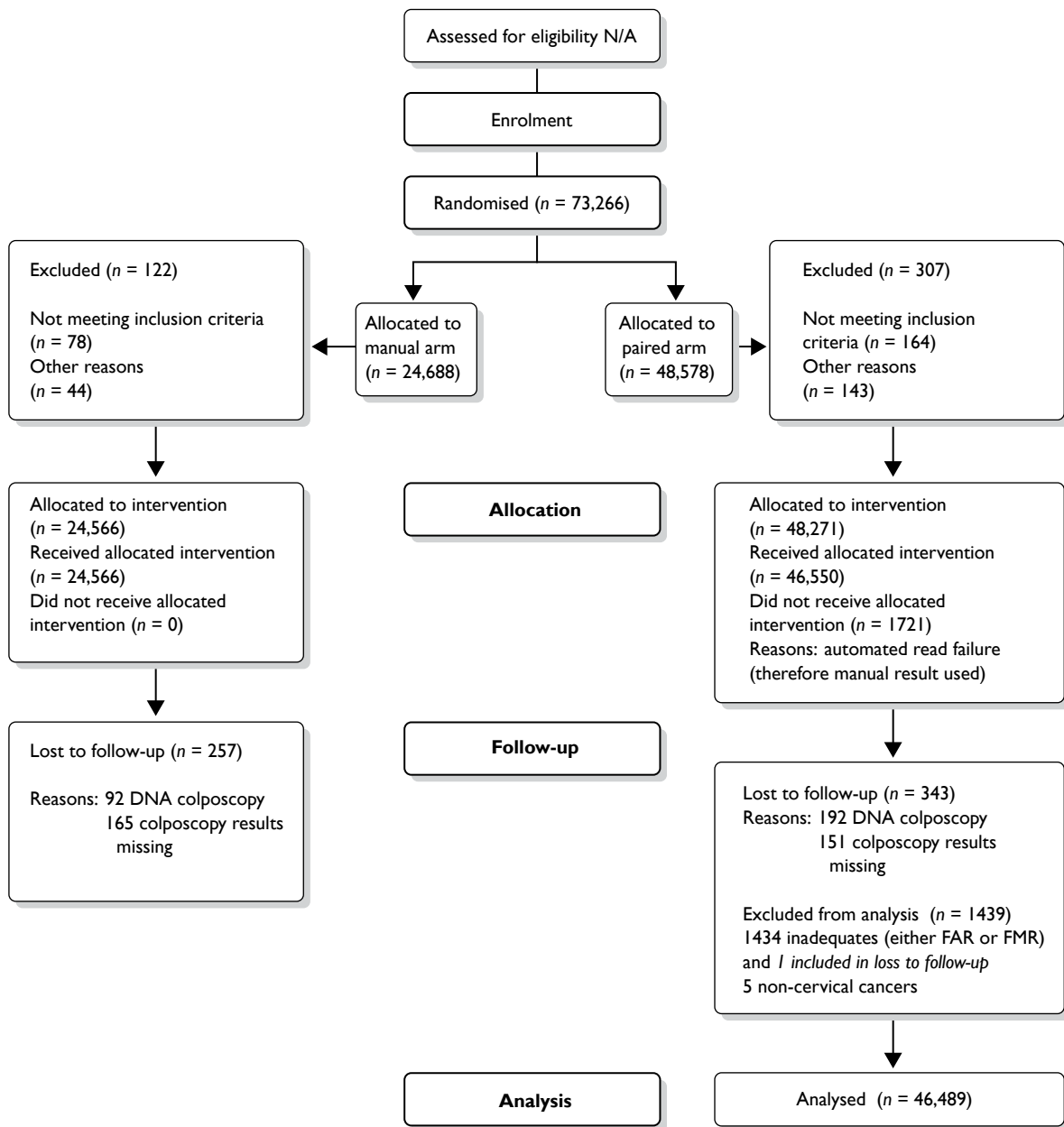


FIGURE 7 The consolidated standards of reporting trials (CONSORT) diagram. N/A, not applicable.

Clinical results

Overall cytology results by age

The age range of women who provided the cytology samples is shown in *Table 15* by quinquennia. Relatively fewer samples from women aged ≥ 50 years were obtained because screening takes place every 5 years in this age range compared with every 3 years between ages 25 and 49 years. In total there were 25,053 samples from women aged 25–34 years, 22,934 from women aged 35–44 years and 21,231 from women aged 45–64 years. There were 3619 (5.0%)

TABLE 15 Summary of subject age

Total	Age at date sample taken (years)										Total
	<25	25–29	30–34	35–39	40–44	45–49	50–54	55–59	60–64	65+	
Manual only and paired arm											
<i>n</i>	3013	13,777	11,276	11,713	11,221	9295	5881	3432	2623	606	72,837
%	4.1	18.9	15.5	16.1	15.4	12.8	8.1	4.7	3.6	0.8	100%
Paired arm only											
<i>n</i>	1890	9143	7520	7876	7444	6176	3901	2232	1707	382	48,271
%	3.9	18.9	15.6	16.3	15.4	12.8	8.1	4.6	3.5	0.8	100%

slides from women outside the screening age range in England, 3013 from women <25 years and 606 from women ≥65 years.

Overall cytology results

The breakdown of cytology results for the paired arm in terms of the MR is shown in *Figure 8*. The proportions of abnormal results by grade are very similar to those for England overall (shown in brackets): borderline 3.6% (3.9%), mild dyskaryosis 2.4% (2.3%) and moderate and severe combined 1.22% (1.2%).⁷ The proportions of inadequate results were 2.8% for the manual read and 1.8% for the auto read. The majority of inadequate results were considered so by both methods, but all non-concordant results were negative by the other method.

Automated read failures

The rates of ARFs are shown in *Table 16*. These have been retained in the analysis; it is assumed that had this occurred in service, the slides would have been subjected to full manual reading. Thus the manual results are used for the auto results: specifically AR1 = MR1 and FAR = FMR. ARFs encompass a number of biological and technical reasons (*Table 17*). The most common is inability to read scanty or thick cell preparations. The large proportion of ARFs experienced with the ThinPrep Imaging System during 2006 was due to problems with the review scope, which were resolved by Hologic. There was a similar problem to a lesser extent with the BD FocalPoint GS Imaging System. Overall, the ARF rates for the BD FocalPoint GS Imaging System and the ThinPrep Imaging System were 3.11% and 3.99% respectively. When the data are restricted to July 2007 onwards (after the initial problems had been rectified) the average proportion of ARFs was 2.93% for both the BD FocalPoint GS Imaging System (585/19,950) and for the ThinPrep Imaging System (655/22,390).

Cytology results (paired arm)

Table 18 shows the distribution by grade of the cytology samples according to the read. For automated reading it is clear that the process of AR1 to AR2 and FAR does not affect high-grade rates, but did produce a drop in borderline/mild dyskaryosis from 5.3% (AR1) through 6% (AR2) to 4.2% (FAR). When the FAR for BD SurePath was compared with ThinPrep there was a noticeable difference between borderline/mild dyskaryosis combined: 3.92% (BD SurePath) and 4.5% (ThinPrep). Other cytology grades were similar, but there was a slightly higher inadequate rate for ThinPrep (1.94%) than for BD SurePath (1.70%).

When a similar comparison is made for manual reading in the paired arm, the same pattern is seen, with borderline/mild dyskaryosis rates going from 7.5% in the MR1 through 8.1% in MR2 to 5.5% in FMR. This borderline/mild rate is higher than that for auto (5.5% vs 4.2%). When BD SurePath and ThinPrep are compared for the paired manual read there is very little difference

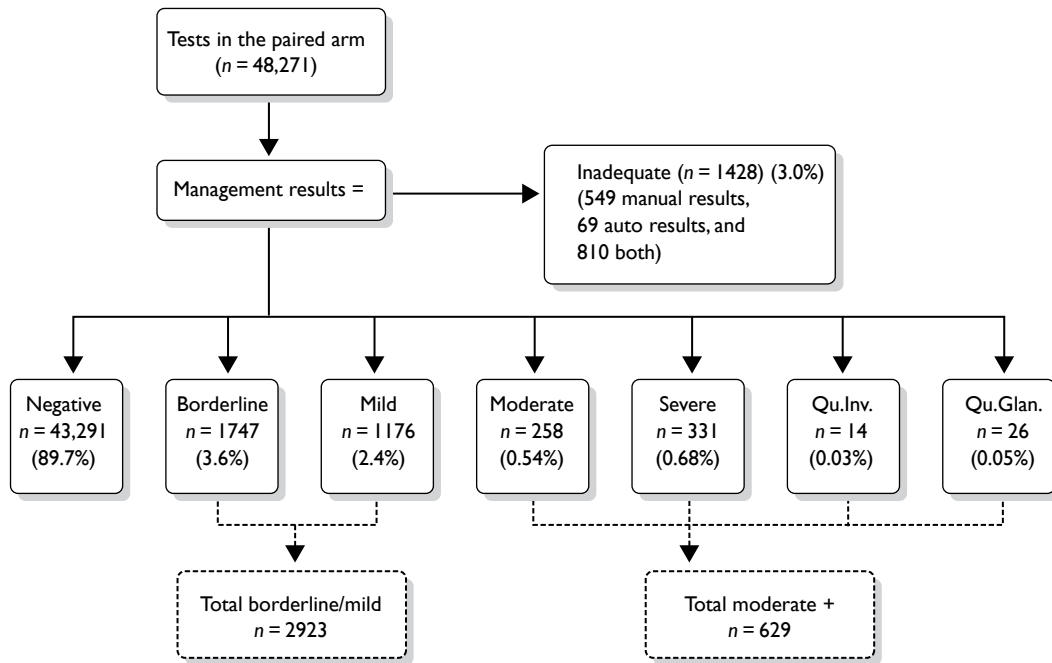


FIGURE 8 Management results of samples in the paired arm. Qu.Glan., query glandular neoplasia; Qu.Inv., query invasive.

TABLE 16 Missing cytology data in paired arm due to ARF

	Total	Missing due to ARF	% missing due to ARF
<i>BD SurePath</i>			
February–June 2006	497	68	13.68%
July–December 2006	1151	38	3.30%
January–June 2007	1806	37	2.05%
July–December 2007	3510	81	2.31%
January–June 2008	6599	168	2.55%
July–December 2008	7370	228	3.09%
January–February 2009	2471	108	4.37%
Total	23,404	728	3.11%
<i>ThinPrep</i>			
February–June 2006	205	81	39.51%
July–December 2006	648	156	24.07%
January–June 2007	1624	101	6.22%
July–December 2007	3629	132	3.64%
January–June 2008	7693	204	2.65%
July–December 2008	9175	267	2.91%
January–February 2009	1893	52	2.75%
Total	24,867	993	3.99%

Total ARF = 1721.

TABLE 17 Reasons for ARF

	Error	Explanation
ThinPrep Imaging System [sample from March 2008, total samples n = 2950 (2.1% ARF rate)]		
<i>Biological</i>		
n = 5 (0.16%)	Sample too scanty	The slide contains an insufficient number of cells to be analysed
n = 14 (0.47%)	Sample too thick	Too many cells present on the slide creating overlapping nuclei and causing problems for the machine in differentiating individual nuclei for imaging
	Sample too clumped	Cytolysis occurs causing clumps of cells to be present on the slide causing problems for the machine in differentiating individual nuclei for imaging
<i>Technical</i>		
n = 7 (0.23%)	Stain too light or dark	Variation in the stain formulation hinders the imaging process
n = 11 (0.37%)	Too many bubbles or mounting media	Refers to bubbles developing in the mounting media underneath the slide cover slip and hindering the imaging process
n = 24 (0.81%)	Too many artefacts on slide	Refers to dirt or small particles of paint from the fiducal marks being present on the slide at a high enough level to hinder the imaging process
n = 2 (0.06%)	OCR read fail	Unable to read the barcode number on the slide
BD FocalPoint GS Imaging System [sample from March 2008, total samples n = 2037 (2.4% ARF rate)]		
<i>Biological</i>		
n = 7 (0.34%)	Sample too scanty or 3D	The slide contains an insufficient number of cells to be analysed, or a 3D effect can be produced by the cell sedimentation process which creates problems for the machine in differentiating individual nuclei for imaging
n = 2 (0.09%)	Sample too thick	Too many cells present on the slide creating overlapping nuclei and causing problems for the machine in differentiating individual nuclei for imaging
	Sample too clumped	Cytolysis occurs causing clumps of cells to be present on the slide, causing problems for the machine in differentiating individual nuclei for imaging
n = 10 (0.49%)	Insufficient reference cells	All the cells in the sample appear similar (e.g. in an atrophic sample) which creates problems for the machine in differentiating enough cells to reference on the slide
<i>Technical</i>		
n = 6 (0.29%)	Stain too light or dark	Variation in the stain formulation hinders the imaging process
	Too many bubbles in mounting media	Refers to bubbles developing in the mounting media underneath the slide cover slip and hindering the imaging process
n = 21 (1.03%)	Unable to read barcode	Unable to read the barcode number on the slide
n = 3 (0.14%)	Unable to analyse slide	Generic error code for failed imaging

3D, three-dimensional; OCR, optical character reader.

across all grades including borderline/mild dyskaryosis (5.5% vs 5.49%), but again there is a slightly higher inadequate rate for ThinPrep (2.98%) than for BD SurePath. The proportion of inadequate results with automated reading is significantly lower than that with manual reading (FAR 1.82%, FMR 2.83%, $p < 0.001$).

Cytology results (manual-only arm)

Table 19 shows the comparative rates of cytological abnormality between manual reads 1 and 2, final manual read and the MR which in the manual-only arm is identical to the final manual read because there is no automated read to alter the management. The same effect of checking is seen, reducing the combined borderline/mild dyskaryosis rate from 8.2% to 6.0%, but there was a slight increase in the combined moderate/severe dyskaryosis rate from 1.2% to 1.4%. This latter rate is slightly higher than the corresponding MR for the paired-reading arm of 1.3% (see Table 18). A

TABLE 18 Distribution by grade of cytology samples according to the category of reading

Read	Cytology result [<i>n</i> (%)]					Total
	Inadequate	Negative	Borderline/ mild	Moderate	Severe+	
AR1 ^a	947 (1.96)	44,151 (91.46)	2543 (5.3)	299 (0.62)	331 (0.7)	48,271 (100)
MR1 ^b	1237 (2.56)	42,826 (88.72)	3636 (7.5)	276 (0.57)	296 (0.6)	48,271 (100)
AR2 ^c	1171 (2.43)	43,543 (90.20)	2913 (6.0)	303 (0.63)	341 (0.7)	48,271 (100)
MR2 ^d	1377 (2.85)	42,416 (87.87)	3892 (8.1)	276 (0.58)	307 (0.63)	48,271 (100)
FAR ^e	879 (1.82)	44,771 (92.75)	2039 (4.22)	238 (0.49)	344 (0.71)	48,271 (100)
FAR BD FocalPoint GS Imaging System	397 (1.70)	21,791 (93.11)	917 (3.92)	118 (0.50)	181 (0.77)	23,404 (100)
FAR ThinPrep Imaging System	482 (1.94)	22,980 (92.41)	1122 (4.5)	120 (0.48)	163 (0.66)	24,867 (100)
FMR ^f	1366 (2.83)	43,647 (90.42)	2641 (5.5)	252 (0.52)	365 (0.75)	48,271 (100)
FMR BD SurePath LBC	626 (2.67)	21,176 (90.48)	1277 (5.5)	130 (0.56)	195 (0.83)	23,404 (100)
FMR ThinPrep LBC	740 (2.98)	22,471 (90.36)	1364 (5.49)	122 (0.49)	170 (0.68)	24,867 (100)
MR	1428 (2.96)	43,291 (89.68)	2923 (6.1)	258 (0.53)	371 (0.77)	48,271 (100)

a AR1 results are the result of the first auto read.

b MR1 results are the result of the first manual read providing this was not by a trainee.

c The AR2 results are defined as the AR1 result if AR1 was positive and the outcome of the auto rapid review for all AR1 results that were negative or inadequate. The AR2 value is prior to checking by a checker.

d The MR2 results are defined as the MR1 result if MR1 was positive or the outcome of the auto rapid review for all MR1 results that were negative or inadequate. The MR2 value is prior to checking by a checker.

e FAR is defined as (a) negative if both AR1 and AR2 are negative; (b) negative if AR1 was negative and the auto rapid review (AR2) was positive, but deemed negative by the checker; (c) the FMR result where AR2 was confirmed as positive by a checker, but the FMR was also positive; and (d) the MR result where AR2 was positive and FMR was negative.

f The FMR is defined as the last manual result before any automated result is taken into account.

TABLE 19 Summary of manual read results in the manual-only arm

Read	Cytology result [<i>n</i> (%)]					Total
	Inadequate	Negative	Borderline/mild	Moderate	Severe+	
MR1 ^a	584 (2.38)	21,680 (88.25)	2000 (8.15)	162 (0.66)	140 (0.58)	24,566 (100)
MR2 ^b	641 (2.61)	21,486 (87.46)	2131 (8.68)	164 (0.67)	164 (0.59)	24,566 (100)
FMR	639 (2.60)	22,118 (90.04)	1476 (6.01)	158 (0.64)	175 (0.72)	24,566 (100)
MR	639 (2.60)	22,118 (90.04)	1476 (6.01)	158 (0.64)	175 (0.72)	24,566 (100)

a MR1 results are the result of the first manual read providing this was not by a trainee.

b The MR2 results are defined as the MR1 result if MR1 was positive and the outcome of the auto rapid review for all MR1 results that were negative or inadequate. The MR2 value is prior to checking by a checker.

more detailed comparison of final manual reading between the manual-only and paired-reading arms is shown in *Tables 22–26*.

Comparison of readings: manual and automated

It is relevant to compare the results of readings at various stages in the process not only between manual and auto, but also between manual readings at different stages and similarly for auto. The key comparison is between the final manual reading and the final auto reading, which generated the difference in relative sensitivity. We also present manual results between the arms, first and final readings and final and MRs. These data are shown in *Tables 20–26*.

TABLE 20 Comparison of the manual read results in the manual and paired arms by reading and cytology result

Read	Cytology result [n (%)]					Total
	Inadequate	Negative	Borderline/mild	Moderate	Severe+	
MR (manual)	639 (2.60)	22,118 (90.04)	1476 (6.01)	158 (0.64)	175 (0.72)	24,566 (100)
MR (paired)	1428 (2.96)	43,291 (89.68)	2923 (6.06)	258 (0.53)	371 (0.77)	48,271 (100)
FMR (manual)	639 (2.60)	22,118 (90.04)	1476 (6.01)	158 (0.64)	175 (0.72)	24,566 (100)
FMR (paired)	1369 (2.84)	43,644 (90.41)	2642 (5.47)	251 (0.52)	365 (0.75)	48,271 (100)

TABLE 21 First manual result versus FMR

FMR	MR1								Total
	Inadequate	Negative	Borderline	Mild	Moderate	Severe	Glan neo	Q invasive	
Inadequate	564	52	20	2	1				639
Negative	18	21,528	542	23	3	3		1	22,118
Borderline	2	72	623	89	13	4	2	3	808
Mild		18	201	413	33	3			668
Moderate		4	18	47	80	9			158
Severe		4	12	7	32	105	1		161
Glan neo						2	1	1	4
Q invasive		2	3			3		2	10
Total	584	21680	1419	581	162	129	4	7	24,566

Glan neo, query glandular neoplasia; Q, query.

Concordant results 23,316 (94.9%); discordant results 1250 (5.1%).

Comparison of manual results (manual arm) versus manual results (paired arm)

The actual MRs were almost identical between the arms, with slightly fewer mild and moderate dyskaryosis and slightly more borderline in the paired arm (*Table 20*). The comparison of FMRs between the arms is important in indicating that the manual reading in the paired arm was similar to 'real-life' manual reading in the manual-only arm which serves as a control. For routine samples, the rates of abnormality are very similar. The non-negative rates of cytology (as a percentage of all adequate samples) are 5.48% (2046/37,369) in the paired arm and 5.52% (1021/18,507) in the manual-only arm. A comparison of these results before and after the change in randomisation showed that the rates in the two arms were similar in both periods (8.32% vs 8.31% and 6.71% vs 6.44% respectively).

Comparison between manual readings in manual-only arm

The association between MR1 and FMR is shown in *Table 21*. There was discordance in 5.1% of cases, half of which were due to borderline/negative mismatches; most were borderline MR1s downgraded to negative in checking. The majority of remaining mismatches were between mild and borderline, most of which were borderline on MR1 being upgraded to mild dyskaryosis. The rates and pattern of these mismatches are similar to those in the paired arm. In the manual-only arm the FMR is by definition equivalent to the MR.

TABLE 22 First automated result versus FMR

FMR	FAR	Inadequate	Negative	Borderline/mild			Moderate+	Total
				HPV positive	HPV negative	HPV not known		
Inadequate		810	556					1366
Negative		69	43,284	125	101	56	12	43,647
Borderline/mild	HPV positive		317	900				1217
	HPV negative		350		334			684
	HPV not known		217			523		740
Moderate+			47				570	617
Total		879	44,771	1025	435	579	582	48,271

Concordant results 46,421 (96.2%); discordant results 1850 (3.8%).

TABLE 23 First manual result versus MR

MR	FMR	Inadequate	Negative	Borderline/mild			Moderate+	Total
				HPV positive	HPV negative	HPV not known		
Inadequate		1359	69					1428
Negative		7	43,284					43,291
Borderline/mild	HPV positive		125	1217				1342
	HPV negative		101		684			785
	HPV not known		56			740		796
Moderate+			12				617	629
Total		1366	43,647	1217	684	740	617	48,271

Concordant results 47,901 (99.2%); discordant results 370 (0.8%).

TABLE 24 Final automated result versus MR

MR	FAR	Inadequate	Negative	Borderline/mild			Moderate+	Total
				HPV positive	HPV negative	HPV not known		
Inadequate		879	549					1428
Negative			43,291					43,291
Borderline/mild	HPV positive		317	1025				1342
	HPV negative		350		435			785
	HPV not known		217			579		796
Moderate+			47				582	629
Total		879	44,771	1025	435	579	582	48,271

Concordant results 46,791 (96.9%); discordant results 1480 (3.1%).

TABLE 25 First automated result versus MR1

MR1	AR1	Inadequate	Negative	Borderline/mild			Moderate+	Total
				HPV positive	HPV negative	HPV not known		
Inadequate		610	608	3		15	1	1237
Negative		307	41,559	118	106	675	61	42,826
Borderline/ mild	HPV positive	7	333	721			58	1119
	HPV negative	1	384		241		13	639
	HPV not known	21	1198			568	91	1878
Moderate+		1	69	24	6	66	406	572
Total		947	44,151	866	353	1324	630	48,271

Concordant results 44105 (91.4%); discordant results 4166 (8.6%).

TABLE 26 First auto result + auto rapid review (AR2) versus first manual result + manual rapid review (MR2)

MR2	AR2	Inadequate	Negative	Borderline/mild			Moderate+	Total
				HPV positive	HPV negative	HPV not known		
Inadequate		765	581	3		27	1	1377
Negative		352	40,989	96	94	831	54	42,416
Borderline/ mild	HPV positive	9	311	798			61	1179
	HPV negative	5	372		304		15	696
	HPV not known	36	1228			655	98	2017
Moderate+		4	62	25	7	73	415	586
Total		1171	43,543	922	405	1586	644	48,271

Concordant results 43,926 (91.0%); discordant results 4345 (9.0%).

Comparison between manual and automated readings in the paired arm

The effect of checking on serial readings is shown in *Tables 20–26*. When the MR1 was compared with the FMR, a large proportion of low grades (29.5%) were downgraded to negative. In addition, nine high grades (1.6%) were downgraded to negative. A similar comparison between AR1 and FAR again shows a large proportion of low grades (29.1%) downgraded to negative, although the overall number of low grades was far fewer in auto than in manual. In comparison, 4.6% of high grades were downgraded to negative (see also *Tables 99 and 100* in *Appendix 13*).

Looking at the data between grades of abnormality is less significant because all abnormal results were acted upon such that all at-risk women (low grade/HPV positive and high grade) were referred to colposcopy.

There were significant numbers of negatives on first reading that were classified as abnormal on the final read. This would have resulted from a rapid review producing an abnormality which was then sent for checking. So, in the FMR, 161 (6.1%) of low grades were classified as negative in the MR1, together with 15 (2.43%) moderate or worse. In the FAR, 211 (10.3%) of low grades were originally classified as negative in the AR1, together with 21 (3.6%) moderate or worse.

When the FARs and FMRs were compared (*Table 22*) there was a discordant rate of 3.8% (1850/48,271), of which half (931/1850) represented abnormal FMRs reported as negative on FAR. This outweighs the discordants (294/1850) where there were abnormalities on FAR reported as negative on FMR. This clearly indicates the potential for greater relative sensitivity by manual than by automated reading. HPV-negative discordants are of little consequence because there is very little risk of disease in these women (350 FMR low-grade/HPV negative were FAR negative), but there were 317 who were low-grade/HPV positive on FMR and reported negative by FAR, with only 125 the other way round. There were therefore 192 more such women referred to colposcopy on the basis of the manual reading process as well as an additional 161 low-grade women in whom the HPV status was not known, and 47 additional FAR negatives who were designated moderate or worse on manual. These additional referrals will have resulted in a number of CIN2+ and CIN3+ (see *Tables 44* and *45*).

The differences between FAR and FMR are reflected in *Tables 23* and *24*, where it is clear that the abnormalities among the MR consist of fewer FMR negatives than FAR negatives.

When first readings were compared (*Table 25*) the non-concordance rate was 8.6% with a far higher proportion of MR1 positive/AR1 negative [1984/44,151 (4.5%)] than MR1 negative/AR1 positive [960/42,826 (2.2%)]. There was a similar rate of discordant results between auto read 2 (AR2) and manual read 2 (MR2) (*Table 26*) with similar patterns: MR2 positive/AR2 negative [1973/43,543 (4.5%)] and MR2 negative/AR2 positive [1075/42,416 (2.5%)] between the discordants. In both of these comparisons there were also similar proportions of discordants between inadequate and satisfactory results, except that more inadequates were identified on second readings (manual 2.85%, auto 2.42%) than on primary reading (manual 2.56%, auto 1.96%).

The comparison between HPV testing results and the MR is shown in *Table 27*, the final manual reading result in *Table 28* and the final auto reading result in *Table 29*. For borderline changes in MRs, the HPV-positive rate among those tested with a valid result was 49.7% (642/1291). For mild dyskaryosis the HPV-positive rate was 83.7% (700/836). The equivalent rates for the FMR were 49.2% (539/1096) and 84.2% (678/805), respectively (see *Table 28*), and for FAR were 56.3% (426/757) and 85.2% (599/703) respectively (see *Table 29*). In a number of cases the HPV testing was performed on the basis of FAR/FMR rather than the MR. This led to some cases

TABLE 27 Comparison of MR and HPV outcome – paired arm

MR	HPV outcome				Total
	Positive	Negative	Invalid	Not tested ^a	
Inadequate		1		1427	1428
Negative	7	55	2	43,227	43,291
Borderline	642	649	24	432	1747
Mild	700	136	7	333	1176
Moderate	16			242	258
Severe	3			328	331
Q Inv				14	14
Q Glan				26	26
Total	1368	841	33	46,029	48,271

Q Glan, query glandular neoplasia; Q Inv, query invasive.

a Commonly, where an HPV test was not carried out when one was expected, the test was a non-routine sample inadvertently labelled as routine by the sample taker and therefore not eligible for triage. In most of the other cases, the subject was aged ≤ 24 years.

TABLE 28 Comparison of FMR and HPV outcome – paired arm

FMR	HPV outcome				Total
	Positive	Negative	Invalid	Not tested ^a	
Inadequate		1		1365	1366
Negative	133	156	4	43,354	43,647
Borderline	539	557	22	384	1502
Mild	678	127	7	327	1139
Moderate	15			237	252
Severe	3			322	325
Q Inv				14	14
Q Glan				26	26
Total	1368	841	33	46,029	48,271

Q Glan, query glandular neoplasia; Q Inv, query invasive.

a Commonly, where an HPV test was not carried out when one was expected, the test was a non-routine sample inadvertently labelled as routine by the sample taker and therefore not eligible for triage. In most of the other cases, the subject was aged ≤ 24 years.

TABLE 29 Comparison of FAR and HPV outcome – paired arm

FAR	HPV outcome				Total
	Positive	Negative	Invalid	Not tested ^a	
Inadequate				879	879
Negative	325	406	10	44,030	44,771
Borderline	426	331	17	269	1043
Mild	599	104	6	287	996
Moderate	16			222	238
Severe	2			308	310
Q Inv				14	14
Q Glan				20	20
Total	1368	841	33	46,029	48,271

Q Glan, query glandular neoplasia; Q Inv, query invasive.

a Commonly where an HPV test was not carried out when one was expected, the test was a non-routine sample inadvertently labelled as routine by the sample taker and therefore not eligible for triage. In most of the other cases, the subject was aged ≤ 24 years.

where the MR was negative and either the FMR or FAR was abnormal. In this context the HPV positive rates for negative cytology were 46.0% (133/289) and 43.3% (325/751) respectively (see *Tables 28* and *29*). This indicates that many samples were wrongly classified, as the expected HPV prevalence for negative cytology in this Manchester population would be around 15%.⁷³ This phenomenon was more marked in the FAR when the HPV-positive rates were similar, but the proportion of HPV triaged slides classified negative on FAR (see *Table 29*) was much higher than for FMR: FAR 33% (731/2209) compared with FMR (see *Table 28*) 13% (289/2209). This indicates that there will be more false-negative cytology tests reported on FAR than on FMR, which will affect the sensitivity of FAR relative to FMR. The final MR corrected almost all of these false-negatives for both FMR and FAR because, of the 62 final management negative samples that were HPV tested, only 11.2% were HPV positive, which is the rate seen in negative cytology in population screening. This demonstrates the validity and thoroughness of the reporting process.

Overall human papillomavirus triage results

Table 30 provides details of the HPV-positive rates (according to the LBC platform used) and grade of cytology for the MR in the manual-only arm (equivalent to the FMR) and for the FMR, FAR and MR in the paired arm. Overall there were no major differences in HPV-positive rates between LBC platforms and between arms for corresponding cytology grades. The significance of the HPV-positive rates for the MAVARIC trial outcome is that for borderline and mild dyskaryosis, these represent the rate of referral for colposcopy among triaged women and therefore possible detection of CIN2+. The cut-off for reporting a sample as positive was changed from 3.0 to 2.0 RLU/CO in February 2008; however, only 1% of samples tested had RLU/CO values between 2.0 and 3.0, so the impact on the proportion referred was minimal.

Overall, it can be seen that for the MR of borderline/mild in the manual-only arm, 66% of samples were HPV positive for both BD SurePath and ThinPrep. Within the paired arm the corresponding figure was 63%. The total proportion of samples with an MR of low-grade cytology (borderline and mild dyskaryosis, including those with unknown HPV status) is almost identical between the arms [manual 1476/24,576 (6.01%); paired 2923/48,271 (6.06%)]. For the FMR, there was a slightly higher rate [1476/24,576 (6.01%)] in the manual-only arm than in the paired arm [2642/48,271 (5.47%)]. When the data were restricted to routine samples in the age range 25–64 years, there was a slightly higher proportion of samples with MRs of low-grade cytology in the paired arm: manual 4.31% (821/19,041) versus paired 4.77% (1839/38,522). For the FMR the rates in the two arms were similar: manual 4.31% (821/19,041) versus paired 4.27% (1643/38,522).

TABLE 30 Human papillomavirus positivity rates by cytology outcome, arm and technology (% positive of those tested)

Arm		Number of samples with borderline changes			Number of samples with mild dyskaryosis			Total			Total
		HPV positive (%)	HPV negative (%)	HPV not known ^a	HPV positive (%)	HPV negative (%)	HPV not known ^a	HPV positive (%)	HPV negative (%)	HPV not known ^a	
FMR manual arm (MR)	BD SurePath	151 (56)	120 (44)	139	148 (82)	33 (18)	129	299 (66)	153 (34)	268	720
	ThinPrep	175 (53)	157 (47)	66	225 (82)	49 (18)	84	400 (66)	206 (34)	150	756
	Total	326 (54)	277 (46)	205	373 (82)	82 (18)	213	699 (66)	359 (34)	418	1476
FMR paired arm	BD SurePath	260 (50)	256 (50)	231	289 (86)	48 (14)	193	549 (64)	304 (36)	424	1277
	ThinPrep	279 (48)	301 (52)	176	389 (83)	79 (17)	141	668 (64)	380 (36)	317	1365
	Total	539 (49)	557 (51)	407	678 (84)	127 (16)	334	1217 (64)	684 (36)	741	2642
FAR paired arm	BD SurePath	195 (59)	133 (41)	151	236 (87)	36 (13)	166	431 (72)	169 (28)	317	917
	ThinPrep	231 (54)	198 (46)	136	363 (84)	68 (16)	127	594 (69)	266 (31)	263	1123
	Total	426 (56)	331 (44)	287	599 (85)	104 (15)	293	1025 (70)	435 (30)	580	2040
MR paired arm	BD SurePath	290 (51)	279 (49)	253	293 (85)	51 (15)	193	583 (64)	330 (36)	446	1359
	ThinPrep	352 (49)	370 (51)	203	407 (83)	85 (17)	147	759 (63)	455 (37)	350	1564
	Total	642 (50)	649 (50)	456	700 (84)	136 (16)	340	1342 (63)	785 (37)	796	2923

^a High proportion of HPV not known because the HPV test was not requested in such cases for one of the following reasons: cytology was non-routine; subject was under 25 years; subject refused HPV test; cytology from colposcopy clinic; surgery failed to gather permission for HPV test; or not first borderline or mild result.

Analysis of cytology management

The analysis of actual management following the cytological MRs is shown in *Table 31*. Most negatives [38,240/43,291 (88.3%)] were returned to routine recall, 4814/43,291 (11.1%) had repeat cytology, presumably as part of a follow-up strategy after an abnormality, and a small number were being seen at colposcopy either because there was clinical suspicion requiring a colposcopic examination or because colposcopy and cytology were part of a follow-up regimen.

Because of the NHSCSP Sentinel Sites protocol being used in Greater Manchester, the majority of borderline/mild cytology was triaged with HPV testing and positives referred for colposcopy. Some borderline abnormalities that were repeat samples were outside this protocol and were referred if persistently abnormal. All women with high-grade abnormalities (moderate dyskaryosis or worse) were referred for colposcopy. There are outstanding colposcopy outcomes for all grades of cytology: 1.2% for low-grade and 3.2% for high-grade cytology.

Colposcopy referral rates

Table 32 shows the proportion of women referred for colposcopy broken down by arm and LBC type, as a result of high-grade cytology and HPV triage of low-grade abnormalities. Of the 72,837 samples included in the analysis, 1000 were taken at colposcopy clinics. Overall the referral rate was 4.7% (3377/71,837). When the data were restricted to routine samples from women aged 25–64 years, the proportion was 4.0% (2292/57,527), 3.9% (735/19,024) in the manual-only arm and 4.0% (1557/38,503) in the paired arm. Between the two LBC systems, 3.7% (1025/27,897) were referred following BD SurePath and 4.3% following ThinPrep cytology (1267/29,666) ($p < 0.001$) (see *Table 32*). The reason for this difference is not clear.

Histology results

The numbers and proportion of cases per 1000 of detected histological lesions are shown in *Table 33*, broken down by grade of lesion, LBC system and trial arm. These data have also been aggregated into CIN2+ and CIN3+ and shown as percentages. However the data are depicted,

TABLE 31 Management of cytology in the paired arm

Management	Colposcopy result availability	Cytology result						
		Inadequate	Negative	Low-grade				High-grade
				HPV positive	HPV negative	HPV not known	HPV not applicable (taken at colposcopy)	
Return to routine		0	38,240	0	463	0	0	0
Repeat		1415	4814	7	302	279	0	0
Refer for colposcopy	Result known	11	7	1208	20	257	0	553
	Result outstanding	1	1	6	0	8	0	20
	DNA	0	7	120	0	36	0	19
Await gynaecology/colposcopy recommendations	Result known	0	97	1	0	0	198	55
	Result outstanding	0	125	0	0	0	9	1
	DNA	0	0	0	0	0	9	1
Total		1428	43,291	1342	785	580	216	629

DNA, did not attend.

TABLE 32 Referral to colposcopy

Arm	Preparation	Number	Referred
Manual only (%)	BD SurePath	12,195	497 (4.08)
<i>Routine ages 25–64 years (%)</i>		<i>9319</i>	<i>318 (3.41)</i>
Paired (%)	BD SurePath	23,404	1030 (4.40)
<i>Routine ages 25–64 years (%)</i>		<i>18,578</i>	<i>707 (3.81)</i>
Manual only (%)	ThinPrep	12,371	626 (5.06)
<i>Routine ages 25–64 years (%)</i>		<i>9722</i>	<i>417 (4.29)</i>
Paired (%)	ThinPrep	24,867	1224 (4.92)
<i>Routine ages 25–64 years (%)</i>		<i>19,944</i>	<i>850 (4.26)</i>

TABLE 33 Histology results (all ages) – number (rate per 1000)

Histology result	Manual arm		Paired arm	
	BD SurePath	ThinPrep	BD SurePath	ThinPrep
1A+	10 (0.82)	6 (0.49)	11 (0.47)	8 (0.32)
Adenocarcinoma/CGIN	7 (0.57)	5 (0.40)	15 (0.64)	7 (0.28)
CIN3	109 (8.94)	81 (6.55)	181 (7.73)	182 (7.32)
CIN2	93 (7.63)	87 (7.03)	159 (6.79)	144 (5.79)
CIN1	117 (9.59)	94 (7.60)	195 (8.33)	159 (6.39)
% CIN2+	1.8 (<i>n</i> =219)	1.4 (<i>n</i> =179)	1.6 (<i>n</i> =366)	1.4 (<i>n</i> =341)
% CIN3+	1.0 (<i>n</i> =126)	0.74 (<i>n</i> =92)	0.9 (<i>n</i> =207)	0.8 (<i>n</i> =197)

1A+, invasive cancer stage 1A or worse (squamous cell). CGIN, cervical glandular intraepithelial neoplasia.

there was no statistical difference between the detection rate in the manual-only arm and the paired arm for both CIN2+ [398/24,566 (1.62%) vs 707/48,271 (1.46%) respectively; $p = 0.10$] and CIN3+ [218/24,566 (0.89%) vs 404/48,271 (0.84%) respectively; $p = 0.48$]. There was a higher detection rate in the BD SurePath samples than in ThinPrep for both CIN2+ [585/35,599 (1.64%) vs 520/37,238 (1.40%) respectively; $p \leq 0.01$] and CIN3+ [333/35,599 (0.94%) vs 289/37,238 (0.78%) respectively; $p = 0.02$]. When the data are restricted to routine samples at ages 25–64 years (Table 34), the rates of CIN2+ are 1.45% (337/23,219) in the manual-only arm and 1.30% (598/45,999), $p =$ not significant, in the paired arm. For CIN3+ the rates are 0.82% (191/23,219) and 0.76% (349/45,999), $p =$ not significant, respectively. The overall CIN2+ detection rates were 1.46% (493/33,784) for BD SurePath LBC and 1.25% (442/35,434) for ThinPrep LBC, $p = 0.02$. The corresponding rates for CIN3+ were 0.85% (287/33,784) and 0.71% (253/35,434) respectively, $p = 0.04$.

Primary outcome

The relative sensitivity of screening by automated or manually read cytology to detect CIN3+ and CIN2+

For the purposes of investigating sensitivity and specificity, the cytology results were translated into positive and negative outcomes for FMR and FAR.

Definition of FAR positive is a FAR result of borderline or worse, and the woman referred to colposcopy (i.e. if borderline/mild the HPV result is positive). FAR negative is any negative result or where the FAR was borderline/mild, but the HPV result was negative.

Where the cytology result was borderline or mild, but the HPV status is not known, then it is assumed to be FAR positive if the subject was sent for colposcopy. Samples where the women were referred to colposcopy, but no result has been obtained (either due to non-attendance or inadequate result) have been excluded. Samples where either the FAR or the FMR result was inadequate have also been excluded. The definition of FMR positive and negative is equivalent.

The primary outcome of the MAVARIC study is shown in *Table 35*, where paired comparisons for CIN2+ are shown for the final manual and automated readings. It is clear that there are 52 more CIN2+ lesions missed on auto than on manual reading. These data form the basis for determining relative sensitivity between manual and automated reading. Similar data are also shown for CIN3+ with 18 more lesions missed on auto than manual reading.

When the clinically less significant outcome of CIN1- is considered (*Table 36*), there were 260 more final auto readings that were negative than final manual readings and for CIN2 or less (CIN2-) the corresponding number was 294. These form the basis for determining relative specificity.

The relative sensitivity based on matched pairs was 0.92 (95% CI 0.89 to 0.95) indicating a statistically significant difference of around 8%. This means that automated reading was less sensitive for manual reading by a margin of 8% in the detection of CIN2+. The specificity for CIN2+ detection was 97.3% (44,564/45,782) for auto readings and 96.8% (44,304/45,782) for manual readings, giving a relative specificity of 1.006 (95% CI 1.005 to 1.007) in favour of auto. This means that automated reading is slightly more specific than manual, but only by a margin of 0.06%.

TABLE 34 Histology results (ages 25–64 years) – number (rate per 1000)

Histology result	Manual arm		Paired arm	
	BD SurePath	ThinPrep	BD SurePath	ThinPrep
1A+	8 (0.70)	6 (0.51)	10 (0.45)	7 (0.30)
Adenocarcinoma/CGIN	5 (0.43)	4 (0.34)	12 (0.54)	5 (0.21)
CIN 3	97 (8.43)	71 (6.06)	155 (6.96)	160 (6.75)
CIN 2	76 (6.61)	70 (5.97)	130 (5.83)	119 (5.02)
CIN 1	101 (8.78)	76 (6.49)	167 (7.49)	146 (6.16)
% CIN2+	1.52 (<i>n</i> =186)	1.2 (<i>n</i> =151)	1.3 (<i>n</i> =307)	1.2 (<i>n</i> =291)
% CIN3+	0.9 (<i>n</i> =110)	0.7 (<i>n</i> =81)	0.8 (<i>n</i> =177)	0.7 (<i>n</i> =172)

1A+, invasive cancer stage 1A or worse (squamous cell). CGIN, cervical glandular intraepithelial neoplasia.

TABLE 35 The relative sensitivity for final automated and manual readings in the paired arm

FMR	FAR			
	CIN2+ ^a		CIN3+	
	Positive	Negative	Positive	Negative
Positive	577	83	340	39
Negative	31	16	21	4
Relative sensitivity (based on matched pairs)	0.92 (95% CI 0.89 to 0.95)		0.95 (95% CI 0.91 to 0.99)	

a Total slides in the paired arm of the study with explicit colposcopy result (excluding 12 cases where the MR was inadequate) *n*=2399.

The corresponding data for CIN3+ include a relative sensitivity of 0.95 (95% CI 0.91 to 0.99), which indicates a statistically significant difference of 5%. This means that automated reading was 5% less sensitive than manual reading in the detection of CIN3+. The specificity difference for CIN3+ is exactly the same as for CIN2+, the specificity for auto reading being 96.8% (44,620/46,085) and for manual 96.2% (44,326/46,085). The relative specificity was 1.007 (95% CI 1.006 to 1.008) in favour of auto reading, meaning that the manual arm was slightly less specific for the detection of CIN3+ and by exactly the same margin as for CIN2+. Clearly, this slight gain in specificity for automated reading is outweighed by the relative loss of sensitivity.

The discordant proportions are shown in *Table 37*, which shows that FMR-positive/FAR-negative results are almost three times as common as FMR-negative/FAR-positive results in cases of CIN2+ and almost twice as common in cases of CIN3+. When all discordant pairs were counted according to our definition, they accounted for similar proportions of both CIN2+ and CIN3+ (14.9% and 16.1% respectively).

The relative sensitivity of automated versus manual reading for routine screening samples in women aged 25–64 years only is presented in *Table 38*. Automated reading was in fact 9% less sensitive for CIN2+ relative to manual for routine screening samples, and 6% less sensitive for CIN3+. Again there was a small gain in specificity of only 0.04% and 0.05% for CIN2+ and CIN3+ respectively (*Table 39*).

TABLE 36 The relative specificity for final automated and manual readings in the paired arm

	FMR	FAR			
		CIN1–		CIN2–	
		Positive	Negative	Positive	Negative
	Positive	1120	358	1357	402
	Negative	98	44,206	108	44,218
Specificity for auto		44,564/45,782 = 97.3%		44,620/46,085 = 96.8%	
Specificity for manual		44,304/45,782 = 96.8%		44,326/46,085 = 96.2%	
Relative specificity		1.006 (95% CI 1.005 to 1.007)		1.007 (95% CI 1.006 to 1.008)	

TABLE 37 Discordance rate of final automated and manual readings in the paired arm

Outcome	Discordance	<i>n</i>	%
CIN2+	FMR negative/FAR positive	31	4.4 (31/707)
	FMR positive/FAR negative	83	11.7 (83/707)
	Overall	114	16.1 (114/707)
CIN3+	FMR negative/FAR positive	21	5.2 (21/404)
	FMR positive/FAR negative	39	9.7 (39/404)
	Overall	60	14.9 (60/404)
CIN1–	FMR negative/FAR positive	98	0.2 (98/45,782)
	FMR positive/FAR negative	358	0.8 (358/45,782)
	Overall	456	1.0 (456/45,782)
CIN2–	FMR negative/FAR positive	108	0.2 (108/46,085)
	FMR positive/FAR negative	402	0.9 (402/46,085)
	Overall	510	1.1 (510/46,085)

TABLE 38 The relative sensitivity for final automated and manual readings in the paired arm: routine samples in women aged 25–64 years

FMR	FAR			
	CIN2+		CIN3+	
	Positive	Negative	Positive	Negative
Positive	362	59	225	28
Negative	22	2	13	0
Relative sensitivity (based on matched pairs)	0.91 (95% CI 0.87 to 0.95)		0.94 (95% CI 0.89 to 0.99)	

TABLE 39 The relative specificity for final automated and manual readings in the paired arm: routine samples in women aged 25–64 years

FMR	FAR			
	CIN1–		CIN2–	
	Positive	Negative	Positive	Negative
Positive	665	225	802	256
Negative	73	35,763	82	35,765
Specificity for auto	35,988/36,726 = 98.0%		36,021/36,905 = 97.6%	
Specificity for manual	35,836/36,726 = 97.6%		35,847/36,905 = 97.1%	
Relative specificity	1.004 (95% CI 1.003 to 1.005)		1.005 (95% CI 1.004 to 1.006)	

Secondary outcomes

The relative sensitivity of screening by the Becton Dickinson FocalPoint Guided Screener Imaging System or ThinPrep Imaging System and manually read cytology to detect CIN3+ and CIN2+

The MAVARIC study was not primarily designed to compare BD SurePath with ThinPrep or the BD FocalPoint GS Imaging System with the ThinPrep Imaging System, but comparing clinical outcomes using these systems was a secondary objective. The study was not formally powered for these secondary analyses and because only half of the CIN2+ lesions followed each of the systems it was likely that the study would be underpowered to detect small differences. The direct comparison of relative sensitivity between the BD FocalPoint GS Imaging System and the ThinPrep Imaging System is shown in *Table 40* and is based on routinely obtained screening samples only. For both CIN2+ and CIN3+, the ThinPrep Imaging System had a slightly higher relative sensitivity (0.92 vs 0.90 and 0.97 vs 0.91 respectively), but this did not reach statistical significance ($p=0.53$ and $p=0.52$ respectively). Relative specificity (*Table 41*) was slightly greater for the BD FocalPoint GS Imaging System, but not sufficiently to give it an advantage over the ThinPrep Imaging System.

The distribution of discordant pairs according to the LBC platform is shown in *Table 42*. There were more discordant pairs with the ThinPrep Imaging System, but most of these were associated with CIN1–. There were nine more CIN3+ lesions that were missed with the ThinPrep Imaging System than with the BD FocalPoint GS Imaging System.

TABLE 40 The relative sensitivity for the final automated and manual readings for the BD FocalPoint GS Imaging System and the ThinPrep Imaging System (restricted to age 25–64 years, routine samples only)

FMR	FAR BD FocalPoint				FAR ThinPrep Imaging System			
	CIN2+		CIN3+		CIN2+		CIN3+	
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
Positive	176	31	106	18	186	28	119	10
Negative	11	1	7	0	11	1	6	0
Relative sensitivity (based on matched pairs)	0.90 (95% CI 0.85 to 0.96)		0.91 (95% CI 0.84 to 0.99)		0.92 (95% CI 0.87 to 0.98)		0.97 (95% CI 0.91 to 1.03)	

TABLE 41 The relative specificity for the final automated and manual readings for the BD FocalPoint GS Imaging System and the ThinPrep Imaging System (restricted to age 25–64 years, routine samples only)

FMR	FAR BD FocalPoint				FAR ThinPrep Imaging System			
	CIN1–		CIN2–		CIN1–		CIN2–	
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
Positive	290	116	360	129	375	109	442	127
Negative	18	17,295	22	17,296	55	18,468	60	18,469
Specificity for automated system	17,411/17,719=98.3%		17,425/17,807=97.9%		18,577/19,007=97.7%		18,596/19,098=97.4%	
Specificity for manual	17,313/17,719=97.7%		17,318/17,807=97.3%		18,577/19,007=97.7%		18,529/19,098=97.0%	
Relative specificity (based on matched pairs)	1.006 (95% CI 1.004 to 1.007)		1.006 (95% CI 1.005 to 1.008)		1.003 (95% CI 1.002 to 1.004)		1.004 (95% CI 1.002 to 1.005)	

TABLE 42 Discordance rate of final automated and manual readings in the BD FocalPoint GS Imaging System and ThinPrep Imaging System paired comparison

Outcome	Discordance	BD FocalPoint		ThinPrep Imaging System	
		n	%	n	%
CIN2+	FMR negative/FAR positive	11	5.0 (11/219)	11	4.9 (11/226)
	FMR positive/FAR negative	31	14.2 (31/219)	28	12.4 (28/226)
	Overall	42	19.2 (42/219)	39	17.3 (39/226)
CIN3+	FMR negative/FAR positive	7	5.3 (7/131)	6	4.4 (6/135)
	FMR positive/FAR negative	18	13.7 (18/131)	10	7.4 (10/135)
	Overall	25	19.1 (25/131)	16	11.9 (16/135)
CIN1–	FMR negative/FAR positive	18	0.1 (18/17,719)	55	0.3 (55/19,007)
	FMR positive/FAR negative	116	0.7 (116/17,719)	109	0.6 (109/19,007)
	Overall	134	0.8 (134/17,719)	164	0.9 (164/19,007)
CIN2–	FMR negative/FAR positive	22	0.1 (22/17,807)	60	0.3 (60/19,098)
	FMR positive/FAR negative	129	0.7 (129/17,807)	127	0.7 (127/19,098)
	Overall	151	0.8 (151/17,807)	187	1.0 (187/19,098)

The detection rates (positive predictive value) for each category of cytology including the threshold of borderline or greater and mild dyskaryosis or greater

Table 43 shows the colposcopy outcomes in relation to the MRs in the paired arm, while Tables 44 and 45 show the colposcopy outcomes in relation to the final manual and auto results in the paired arm. There are 2411 known colposcopy results with 707 CIN2+, including 404 cases classified as CIN3+. The large majority of CIN3+ (317/404) and about two-thirds of CIN2+ (458/707) were found in women with manual cytology classified as moderate or worse. Among all borderline cytology the CIN2+ detection rate (PPV) was 5.2% (90/1747) and 12.4% (146/1176) for mild dyskaryosis. When these low-grade cytology results were triaged by HPV testing, the corresponding PPVs increased to 13.5% (80/594) and 16.4% (104/635) for borderline and mild respectively. This demonstrates that HPV status is a powerful discriminant of underlying risk irrespective of the category of low-grade abnormality. For the FMR, the PPV for all borderline cytology for CIN2+ (among those with a known colposcopy result) was 13.0% (84/648), for mild cytology 15.1% (138/915) and for moderate or worse 75.4% (450/597). For the FAR the PPVs for borderline are 11.3% (54/476), for mild 15.7% (127/808) and for moderate or worse 76.4% (429/561).

The data in Tables 44 and 45 can be used to determine the additional number of colposcopies required to achieve the added sensitivity (and slightly reduced specificity) of manual reading. If all women with borderline/HPV positive, mild/HPV positive and HPV not known, and moderate+ were referred to colposcopy, 250 additional colposcopies would have been required to detect 47 additional CIN2+, when compared with automated reading. Therefore, these extra colposcopies had a PPV of 19%. Although lower than the overall proportion of CIN2+ among the colposcopy group (707/2411, 29.32%) it is certainly consistent with the rate to be expected for women referred following HPV triage of low-grade abnormalities and would therefore be considered a worthwhile use of resource.

In Appendix 8, the results of triage are shown if they had been based on HPV genotyping as opposed to HC2. Essentially triage based on HPV 16 and/or 18 increased the PPV for CIN2+, but only from 15% to 25%, with a much lower sensitivity.

Analysis of the Becton Dickinson FocalPoint Guided Screener Imaging System no further review and quintiles

The BD FocalPoint GS Imaging System incorporates a ranking element which identifies the samples least likely to have evidence of disease, to the extent that NFR is required; this accounts for up to 25% of the slides. The remaining samples are divided into five quintiles ranging from quintile 1 (most likely to contain abnormal cytology) to quintile 5 (least likely to contain abnormal cytology). The results of the ranking are shown in Table 46. NFR was the ranking in 21.9% (4569/20,882) of slides with just four (0.02%) high-grade manual cytology readings associated with these. Table 47 shows the histological result correlated with the BD FocalPoint GS Imaging System result. Ten CIN2+ and four CIN3+ were detected, which account for only 3.1% of all CIN2+ detected using the BD FocalPoint GS Imaging System (and 2.2% of CIN3+). The proportion of the total number of CIN2+ detected in quintiles 1–5 was 63.6%, 13.7%, 8.9%, 5.5% and 5.5% respectively. It would appear that NFR could be used safely to archive cytology without further reading, which could be labour saving and cost-effective (see *Costs and cost-effectiveness of the Becton Dickinson FocalPoint Slide Profiler as a stand-alone device*), even without the use of automated reading and the Guided Screener Workstations. If NFR were restricted to only routine samples, < 1% of CIN2+ would have gone undetected.

As shown in Table 48, of the 4910 samples marked for NFR, four were associated with a histological result of CIN3+. These potential false-negatives would not have been identified by

TABLE 43 Association of MR (paired arm) with colposcopy outcomes

Colposcopy outcome	Cytology/HPV											Total	
	Negative	Inadequate	Borderline HPV positive	Borderline HPV negative	Borderline HPV not known	Mild HPV positive	Mild HPV negative	Mild HPV not known	Moderate	Severe	Q inv		Q glan
CIN3+ (%)	3 (0.7)	0 (0)	37 (9.2)	1 (0.2)	1 (0.2)	34 (8.4)	0 (0)	11 (2.7)	72 (17.8)	220 (54.5)	10 (2.5)	15 (3.7)	404 (100)
CIN2 (%)	10 (3.3)	0 (0)	43 (14.2)	0 (0)	8 (2.6)	70 (23.1)	2 (0.7)	29 (9.6)	82 (27.1)	57 (18.8)	2 (0.7)	0 (0)	303 (100)
CIN1 (%)	14 (4.0)	0 (0)	68 (19.2)	3 (0.8)	16 (4.5)	130 (36.7)	0 (0)	68 (19.2)	38 (10.7)	14 (4.0)	0 (0)	3 (0.8)	354 (100)
HPV only (%)	45 (7.9)	1 (0.2)	149 (26.2)	1 (0.2)	44 (7.7)	166 (29.2)	2 (0.4)	101 (17.8)	43 (7.6)	15 (2.6)	1 (0.2)	0 (0)	568 (100)
Colposcopy NAD (%)	34 (4.4)	11 (1.4)	277 (35.6)	11 (1.4)	83 (10.7)	235 (30.2)	1 (0.1)	93 (12.0)	14 (1.8)	13 (1.7)	0 (0)	5 (0.6)	777 (100)
Other cancer (%)	0 (0)	0 (0)	0 (0)	0 (0)	1 (20.0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (20.0)	3 (60.0)	5 (100)
DNA	7	0	58	0	14	62	0	31	8	12	0	0	192
Not referred	126	1	3	0	10	3	0	7	1	0	0	0	151
Results outstanding	43,052	1415	7	633	279	0	131	0	0	0	0	0	45,517
Total	43,291	1428	642	649	456	700	136	340	258	331	14	26	48,271

DNA, did not attend; NAD, no abnormality detected; Q Glan, query glandular neoplasia; Q Inv, query invasive.

TABLE 44 Association of FMR (paired arm) with colposcopy outcomes

Colposcopy outcome	Cytology/HPV												
	Negative	Inadequate	Borderline HPV positive	Borderline HPV negative	Borderline HPV not known	Mild HPV positive	Mild HPV negative	Mild HPV not known	Moderate	Severe	Q inv	Q gland	Total
CIN3+ (%)	24 (5.9)	0 (0)	28 (6.9)	1 (0.25)	1 (0.25)	31 (7.7)	0 (0)	10 (2.5)	70 (17.3)	214 (53.0)	10 (2.5)	15 (3.7)	404 (100)
CIN2 (%)	21 (6.9)	0 (0)	37 (2.2)	0 (0)	7 (2.3)	68 (22.4)	1 (0.3)	28 (9.2)	82 (27.1)	57 (18.8)	2 (0.7)	0 (0)	303 (100)
CIN1 (%)	29 (8.2)	0 (0)	58 (16.4)	3 (0.9)	16 (4.5)	127 (35.9)	0 (0)	68 (19.2)	36 (10.2)	14 (4.0)	0 (0)	3 (0.9)	354 (100)
HPV only (%)	70 (12.3)	1 (0.2)	132 (23.2)	1 (0.2)	43 (7.6)	162 (28.5)	2 (0.4)	99 (17.4)	42 (7.4)	15 (2.6)	1 (0.2)	0 (0)	568 (100)
Colposcopy NAD (%)	95 (12.2)	11 (1.4)	233 (30.0)	8 (1.03)	79 (10.2)	226 (29.1)	1 (0.1)	92 (11.8)	14 (1.8)	13 (1.7)	0 (0)	5 (0.6)	777 (100)
Other cancer (%)	0 (0)	0 (0)	0 (0)	0 (0)	1 (20)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (20)	3 (60)	5 (100)
DNA	26	0	44	0	12	61	0	30	7	12	0	0	192
Not referred	43,253	1,353	6	544	238	0	123	0	0	0	0	0	45,517
Results outstanding	129	1	1	0	9	3	0	7	1	0	0	0	151
Total	43,647	1,366	539	557	406	678	127	334	252	325	14	26	48,271

DNA, did not attend; NAD, no abnormality detected.

TABLE 45 Association of FAR (paired arm) with colposcopy outcomes

Colposcopy outcome	Cytology/HPV										Total		
	Negative	Inadequate	Borderline HPV positive	Borderline HPV negative	Borderline HPV not known	Mild HPV positive	Mild HPV negative	Mild HPV not known	Moderate	Severe		Q inv	Q glan
CIN3+ (%)	43 (10.6)	0 (0)	26 (6.4)	0 (0)	1 (0.2)	29 (7.2)	0 (0)	11 (2.7)	64 (15.8)	207 (51.2)	10 (2.5)	13 (3.2)	404 (100)
CIN2 (%)	54 (17.8)	0 (0)	24 (7.9)	0 (0)	3 (1.0)	58 (19.1)	2 (0.7)	27 (8.9)	78 (25.7)	55 (18.2)	2 (0.7)	0 (0)	303 (100)
CIN1 (%)	74 (21.0)	0 (0)	48 (13.6)	1 (0.3)	10 (2.8)	113 (31.9)	0 (0)	58 (16.4)	36 (10.2)	12 (3.4)	0 (0)	2 (0.6)	354 (100)
HPV only (%)	163 (28.7)	1 (0.2)	97 (17.1)	1 (0.2)	22 (3.9)	141 (24.8)	2 (0.4)	90 (15.8)	37 (6.5)	13 (2.3)	1 (0.2)	0 (0)	568 (100)
Colposcopy NAD (%)	225 (29.0)	5 (0.6)	181 (23.2)	8 (1.02)	54 (6.9)	203 (26.1)	1 (0.1)	73 (9.4)	14 (1.8)	11 (1.4)	0 (0)	2 (0.3)	777 (100)
Other cancer (%)	1 (20)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (20)	3 (60)	5 (100)
DNA	38	0	43	0	11	52	0	28	8	12	0	0	192
Not referred	44,040	872	5	321	180	0	99	0	0	0	0	0	45,517
Results outstanding	133	1	2	0	5	3	0	6	1	0	0	0	151
Total	44,771	879	426	331	286	599	104	293	238	310	14	20	48,271

DNA, did not attend; NAD, no abnormality detected.

TABLE 46 Analysis of BD FocalPoint GS Imaging System NFR and quintiles by cytology result

Final result	NFR	Process review	Rerun Q5	Review					Total
				Q4	Q3	Q2	Q1		
Inadequate	110			127	111	98	75	50	571
Negative	4360			3244	3071	2980	2879	2272	18,806
Low grade									
HPV positive	32			35	47	68	109	233	524
HPV negative	25			31	26	54	65	98	299
HPV not known	38			26	36	46	74	167	387
Moderate	1			3	6	10	15	83	118
Severe	2			4	6	9	19	110	150
Query invasive	0			2	0	0	0	5	7
Query glandular	1			2	0	5	2	10	20
Total	4569	0	0	3474	3303	3270	3238	3028	20,822 ^a

Q, quintile.

^a 23,404 samples were subjected to automated reading with the BD FocalPoint GS Imaging System, but only 20,882 could be retrieved from storage; 2522 (11.3%) were missing.

TABLE 47 Analysis of BD FocalPoint GS Imaging System NFR and quintiles by suggested management outcomes (routine samples only)

Management outcomes	NFR	Q5	Q4	Q3	Q2	Q1	Total
Not referred	4452 (3714)	3381 (2869)	3190 (2686)	3118 (2634)	3016 (2584)	2439 (2041)	19,596 (16,528)
Total referred	117 (26)	93 (41)	113 (59)	152 (83)	222 (124)	589 (370)	1286 (703)
CIN2+	10 (2)	18 (9)	18 (12)	29 (20)	41 (27)	205 (152)	321 (222)
CIN3+	4 (1)	10 (6)	10 (6)	16 (10)	28 (20)	116 (90)	184 (133)
Total	4569 (3740)	3474 (2910)	3303 (2745)	3270 (2717)	3238 (2708)	3028 (2411)	20,882 (17,231)

Q, quintile.

rapid review. Of the 26 NFR samples deemed non-negative by the rapid reviewer, the most severe histological result obtained was CIN1.

Histology outcomes of discordant results

The clinical outcomes in women in whom discordant results between manual and auto were found in the paired arm are shown in *Table 49*. Discordant results included the following matched pairs: borderline or mild/HPV positive and negative cytology; moderate dyskaryosis or worse and negative cytology; and, borderline/mild (HPV not tested) referred to colposcopy and negative. In total there were 52 additional CIN2+ associated with FMR positive/FAR negative than FMR negative/FAR positive (*Table 50*). Most of these lesions were detected in the borderline/mild HPV-positive category. An analysis of the timings of the discordant pairs revealed that discordant pairs occurred at an equal distribution throughout the duration of the study.

TABLE 48 Colposcopy outcomes for those samples marked for NFR with and without rapid review

	NFR result alone (i.e. negative)	NFR result modified by rapid review or repeated processing											
		Negative	Inadequate	Borderline HPV positive	Borderline HPV negative	Borderline HPV not known	Mild HPV positive	Mild HPV negative	Mild HPV not known	Moderate	Severe	Q Inv	Q Glan
Cancer 1b													
Cancer 1a	1	1											
Adenocarcinoma/CGIN	1	1											
CIN3	2	2											
CIN2	6	5	1										
CIN1	12	12											
HPV only	34	31	1		2								
No CIN/HPV	10	10											
Colposcopy NAD	23	20	3										
DNA	28	27	1										
Not referred	4452	4386	43		3	15		1	2		2		
Total NFR	4569	4495	45	4	3	17	0	1	2	0	2	0	0

CGIN, cervical glandular intraepithelial neoplasia; DNA, did not attend; NAD, no abnormality detected; Q Glan, query glandular neoplasia; Q Inv, query invasive.

TABLE 49 Final manual results and FARs for discordant pairs (CIN2+)

Grade of positive result	FMR positive/FAR negative ^a	FAR positive/FMR negative
Borderline	HPV positive	30
	Not tested	5
Mild	HPV positive	17
	Not tested	2
Moderate	12	2
Severe+	17	6
Total	83	31

a A negative result was negative cytology in comparison with abnormal cytology and a borderline/mild HPV negative in comparison with moderate dyskaryosis or worse.

In order to determine whether the discordant readings were due to errors in interpreting the cells as presented in the FOV, or a failure by the automated machine to locate and present the abnormal cells, a rereading of discordant pairs associated with CIN2+ was undertaken. As shown in *Table 51*, 46/61 cases involving auto negative or auto low-grade/HPV negative were considered to be due to interpretation error, i.e. the abnormality had been presented in the FOVs, but had been missed. This applied to both the ThinPrep Imaging System and the BD FocalPoint GS Imaging System. In one-quarter (15/61), no abnormality was seen on review, suggesting an automated location error. It is important to note that that these 'missed' reads on automated reading relate to instances with underlying CIN2+. We have not analysed all instances with similar discordant results where there was no underlying disease.

TABLE 50 Histology outcomes for discordant pairs

Histology Outcome	Type of mismatch between manual/auto in paired readings			
	Auto positive/manual negative		Manual positive/auto negative	
	Auto LG (HPV positive)/ manual negative	Auto HG/manual ≤ LG (HPV negative)	Manual LG (HPV positive)/auto negative	Manual HG/auto ≤ LG (HPV negative)
CIN 2	10	0	38	6
CIN 3	12	8	15	17
CGIN	1	0	1	3
Cancer	0	1 ^a	0	2 ^b
Total	23	9	54	28

CGIN, cervical glandular intraepithelial neoplasia; HG, high grade; LG, low grade.

a Micro-invasive squamous cell carcinoma.

b Contains one non-cervical adenocarcinoma and one squamous cell carcinoma stage 1.

TABLE 51 Reasons for auto/manual discordant results

Reason for discordant results		Type of mismatch between manual/auto in paired readings					
		Auto positive/manual negative			Manual positive/auto negative		
		Auto LG (HPV positive)/ manual negative	Auto HG/ manual ≤ LG (HPV negative)	Total	Manual LG (HPV positive)/ auto negative ^a	Manual HG/ auto ≤ LG (HPV negative) ^b	Total
Interpretation error	Manual	23	8	31	0	0	0
	Automated	0	0	0	29	17	46
Automated location error		N/A	N/A	N/A	12 ^c	3 ^d	15
Total		23	8	31	41	20	61

HG, high grade; LG, low grade; N/A, not applicable.

a One ThinPrep slide in this category was unable to be reviewed owing to bad optical character read – the machine was unable to recognise the slide's barcode and display the corresponding FOVs – plus eight BD SurePath slides that had been signed out on the machine in error, preventing the reviewers from accessing and reviewing the FOVs.

b Seven BD SurePath slides in this category were unable to be reviewed as they had been signed out on the machine, preventing the reviewers from accessing and reviewing the FOVs.

c Total includes three BD SurePath slides that were classified as NFR.

d Total included three BD SurePath slides that were classified as NFR.

Six slides are outstanding.

A sample of the discordant results, where the abnormal cells had been missed owing to automated interpretation error, was rescreened by the review panel in an attempt to determine why the primary screeners had missed the cells presented in the FOVs. The results are presented in *Table 52*. In the majority of cases the cells were interpreted incorrectly on the automated screening owing to biological limitations within the slide preparation – in nearly a quarter of cases the FOVs contained a scanty preparation of cells, while 16.6% of the slides reviewed were found to have FOVs containing hyperchromatic crowded groups. The remainder of the biological limitations were due to inflammatory cells, pale and small cell dyskaryosis, and blood-stained samples. Another difficulty, alongside biological limiting factors, noted by reviewers was the location of the abnormal cells in relation to the centre of the FOVs. In 17.9% of the cases reviewed, it was thought that the cytoscreeners had overlooked the abnormal cells as they

TABLE 52 Review of automated interpretation errors

Reason	Imaging system		n (%)
	ThinPrep Imaging System	BD FocalPoint	
Difficult to grade	1	1	2 (5.1)
Hyperchromatic crowded groups	3	3	6 (15.4)
Scanty	6	3	9 (23.1)
Inflammatory	3	0	3 (7.7)
Cells on edge of FOV	3	4	7 (17.9)
Pale dyskaryosis	0	1	1 (2.6)
Small cell dyskaryosis	0	1	1 (2.6)
Blood stained	1	0	1 (2.6)
No reason found	9	0	9 (23.1)
Total	26	13	39

were peripheral to the FOV presented by the automated review scopes. Peripheral cells are not thought to present a problem in manual screening owing to the practice of 'overlapping' FOVs, which is lost when primary screeners are restricted to either 10 or 22 FOVs. The practice of screening limited fields on the slide was also thought to hinder the interpretation of the biological limitations as the ability to place the cells in the overall context of the slide is lost.

Economics and organisational outcomes

Productivity

Loading and unloading time of equipment in the automated arm

Medical laboratory assistants completed 34 worksheets recording the times for loading and unloading LBC slides on the automated equipment (17 using the BD FocalPoint GS Imaging System and 17 using the ThinPrep Imaging System). The results show that MLAs on average spent 30 minutes to load and unload 160 slides using the BD FocalPoint GS Imaging System. Using the ThinPrep Imaging System on average they took 26 minutes to load and unload 251 slides. The results (*Table 53* and *Figure 9*) show that the mean (standard deviation) time for loading and unloading a sample using the BD FocalPoint GS Imaging System was 0.10 (0.06) minutes and 0.09 (0.03) minutes, respectively, amounting to a total time of 0.20 (0.07) minutes. The loading and unloading time using the ThinPrep Imaging System was 0.06 (0.04) minutes and 0.05 (0.01) minutes respectively, with a total time of 0.11 (0.04) minutes per slide. Therefore, overall it was quicker to load and unload LBC samples with the ThinPrep Imaging System and this difference was statistically significant ($p < 0.000$).

Average primary slide reading time

Two time-and-motion surveys were conducted to estimate the average time to read slides, one at 6 months and the other after staff had been reading automated slides for 3 years. The initial time and motion included 160 observations of primary slide reading time across the manual and automated arms. After cytoscreeners had been using the automated equipment for nearly 3 years a much larger study was conducted with a total of 1990 observations. The results of the surveys are reported in *Table 54*.

A very large difference was observed in the average primary review times between automated and manual technologies. These findings are expected, as with manual reading the whole slide has to be reviewed, whereas with automated technologies cytoscreeners are directed to specific

TABLE 53 Duration in minutes [mean and standard deviation (SD)] for loading and unloading LBC slides

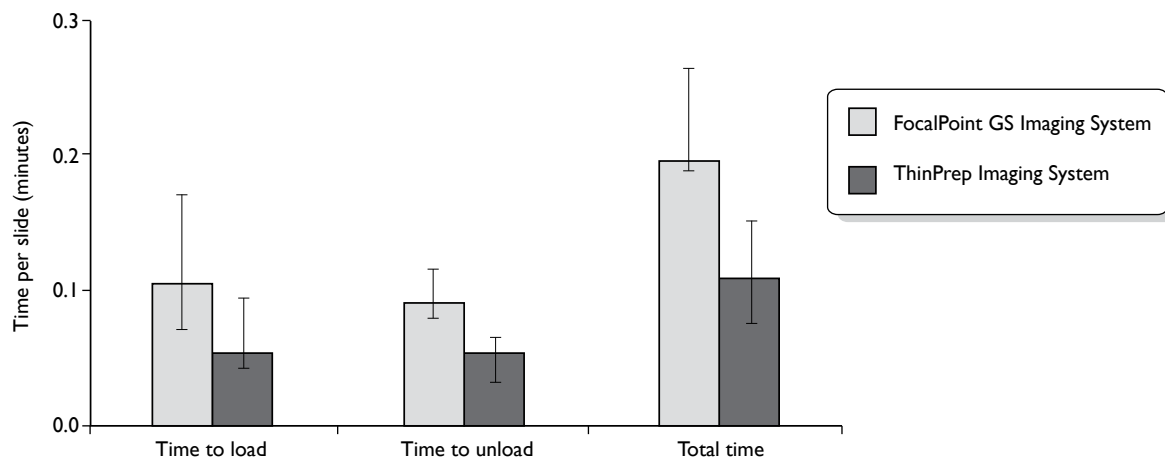
Technology	Time to load		Time to unload		Total time	
	Mean	SD	Mean	SD	Mean	SD
BD FocalPoint GS Imaging System	0.10	0.06	0.09	0.03	0.20	0.07
ThinPrep Imaging System	0.06	0.04	0.05	0.01	0.11	0.04

TABLE 54 Average slide reading times in minutes [mean and standard deviation (SD)] after 6 months and 3 years

Time period	Screening stage	Automated arm		Manual arm	
		BD FocalPoint GS Imaging System	ThinPrep Imaging System	BD SurePath LBC	ThinPrep LBC
After 6 months	Primary review ^a	1.37 (1.07)	1.48 (0.99)	5.28 (2.01)	4.11 (0.99)
	Rapid review ^b	1.14 (0.46)	1.45 (0.37)	1.08 (0.28)	1.58 (0.35)
After nearly 3 years	Primary review ^a	1.64 (1.62)	1.27 (1.58)	5.34 (1.89)	5.36 (2.48)
	Rapid review ^b	1.35 (2.11)	1.47 (0.54)	1.65 (0.54)	1.66 (0.69)

a Primary review times include average administration time.

b Rapid review in the automated arm is performed manually.

**FIGURE 9** Duration of loading and unloading per slide with automated technologies.

points on the slide that are most likely to contain abnormal cells. The results of the larger timing survey, conducted after cytoscreeners had been using the automated technologies for 3 years (see Table 54), are more likely to reflect routine practice well after any initial learning curve effect, and so only these later results are used in the main analyses. These data show a large and statistically significant difference ($p = 0.05$) in the mean time required for primary review between automated and manual reading; primary review times with automated reading were 3.26 and 4.23 times faster than manual reading for the BD FocalPoint GS Imaging System and the ThinPrep Imaging System respectively. The differences in average primary reading times between the two LBC systems were not statistically significant ($p = 0.14$), whether slides were read manually or with automated equipment. Time-and-motion estimates suggest that the hourly rate of ThinPrep Imaging System assisted screening is 37–47 slides and the corresponding rate of manual screening is about 11 primary slides.

The results from the initial time-and-motion study conducted after 6 months show very similar results. These data suggest that over time cytoscreeners took slightly longer to review slides with the BD FocalPoint GS Imaging System slides; in contrast with the ThinPrep Imaging System, cytology reading times increased slightly after 3 years.

Very little difference was observed in the average time for rapid review between the arms. This finding was not unexpected, as even with automated reading, slides are rapidly reviewed manually and the time involved follows laboratory protocols. The average time per slide for administrative activities associated with the full screen was 0.39 seconds (95% CI 0.38 to 0.40 seconds), and the average administration time involved with rapid review was 0.29 seconds (95% CI 0.28 to 0.30 seconds). The administration times are based on the manual arm only. There may be slight differences in administration time between the two automated technologies, for example the time taken to read quintile data that were not captured in the time-and-motion study.

No timing surveys were undertaken on checkers or pathologists, because these later stages of the cytology reading process are similar for both the manual and automated arms. The average times for these aspects of the slide reading pathway were therefore taken from an earlier time-and-motion study conducted in the same laboratory (*Table 55*).⁴⁹

Average number of slides screened per day

As well as the detailed time-and-motion studies, we also wanted to measure the productivity implications in terms of the overall number of slides that cytoscreeners could read per day. During a full working day cytoscreeners undertake a variety of activities, such as rapid review and filing of histological records, and also take some time off for breaks. The amount of time allowed for undertaking primary screening is restricted by national guidelines to a maximum of 5 hours per day.⁷⁵ During most of the trial, cytoscreeners would read slides in both the manual and automated reading arms on the same day, and therefore it was not possible to distinguish the total number of slides that could be read per day for each technology. Near the end of the trial, for a period of 5–6 weeks, cytoscreeners worked only on the automated or manual technologies, and during this period they completed record sheets on the average number of slides read while primary reading, as well as the number of hours spent on other activities such as rapid review, filing histological records and breaks. From these data, we again calculated the average time for primary review of slides and rapid review for purposes of comparison with the time-and-motion results.

The results are consistent with the time-and-motion study results, and show that primary review time per slide has significantly reduced owing to automated assisted screening. There are no significant differences in average time between the two automated technologies: the hourly screening rates using the BD FocalPoint GS Imaging System and ThinPrep Imaging System are 20 and 19 slides respectively (*Table 56*). The hourly screening rate using corresponding manual screening methods is nine primary slides under both technologies. Estimates of the average times for primary and rapid review in the workload survey are much higher than the time-and-motion survey results (reported in *Table 54*), suggesting that the time-and-motion survey results may fail to measure some *x*-inefficiency in the process and so underestimate the actual time involved.

Data from the workload surveys also indicated that there were no significant differences in the amount of time that cytoscreeners spent on primary review, but there were fluctuations depending on the amount of time related to rapid review and other activities that were not specific to any arm. The average time spent on other activities was on average 1.73 and 2.24 minutes per slide with automated and manual screening respectively.

TABLE 55 Duration of checking and secondary reading time in minutes [mean and standard deviation (SD)]

Examination stage	Source	Duration in minutes including administration time (SD)
Checking	ARTISTIC ⁴⁹	5:40 (1:52)
Secondary reading	ARTISTIC ⁴⁹	6:23 (2:00)

TABLE 56 Duration in minutes [mean (standard deviation)] for reading slides workload survey

	Automated arm		Manual arm	
	BD FocalPoint GS Imaging System	ThinPrep Imaging System	BD SurePath LBC	ThinPrep LBC
Primary review	3.01 (2.55)	3.18 (2.50)	6.37 (3.83)	7.04 (3.09)

Cytology reading events

Where abnormalities are identified at either primary screening or rapid review, slides are referred on for 'checking' by a senior BMS and/or for a final review by a medic or AP. With automated equipment, slides might have an ARF or be identified as requiring NFR. ARF occurs when cytology slides cannot be read by the automated equipment, because of problems either with the stain or with scanning. Where this happened, for productivity measurement and costing purposes we assumed that these slides would be read manually. In addition, with the BD FocalPoint Imaging System, up to 25% of the slides that are least likely to contain abnormal cells are identified as requiring NFR by a cytoscreener. In the trial, slides classified as NFR were sent directly for 'rapid review' without an initial primary review.

As reported in the *Clinical results* section (see *Table 16*), at the beginning of the trial very high ARF rates were observed owing to technical problems which were later rectified. Therefore, for productivity and costing purposes we have estimated average ARF rates excluding the first 12 months of the trial. The rates of cytology reading events are reported in *Table 57*.

The BD FocalPoint GS Imaging System classified on average 21.23% of the slides as requiring NFR. ARF rates were slightly higher in the ThinPrep Imaging System than in the BD FocalPoint GS Imaging System arm, 2.98% compared with 2.88%. The difference in ARF rates between technologies was not significant ($p = 0.054$).

Primary screening workloads were reduced with the BD FocalPoint GS Imaging System due to the fact that slides were identified automatically as requiring NFR. This also led to a slight increase in the rapid review workload, although increased rates of rapid review were seen across both the automated technologies. Workloads for senior staff involved in checking and secondary screening were slightly reduced owing to automated-assisted screening as a result of fewer slides being referred to checking.

Average total staff time per slide

Average total staff time per slide, including primary screening, rapid review, 'checking' and reading by the medic, was estimated by combining the results of the average time to undertake each activity [from either the time-and-motion study data or the workload survey data (see *Tables 54* and *56*)] with the probability of each cytology reading activity in the laboratory (see *Table 57*). The overall time duration for each activity is reported in *Tables 58* and *59*.

TABLE 57 Probability of cytology reading events

Probability of event	Automated arm		Manual arm	
	BD FocalPoint GS Imaging System	ThinPrep Imaging System	BD SurePath LBC	ThinPrep LBC
NFR	21.23%	0%	0%	0%
ARF	2.88% ^a	2.98% ^a	0%	0%
Primary screening	75.89%	97.02%	100%	100%
Primary rapid review	94.93%	93.21%	91.77%	90.88%
Checking	5.82%	7.65%	7.89%	9.50%
Secondary reading	5.63%	6.34%	7.57%	7.69%

a It is assumed that these slides are read manually in further analyses.

TABLE 58 Average time (minutes) per slide (95% CI) adjusted with probability of slide passing through different events – time-and-motion study

Event	Automated arm		Manual arm	
	BD FocalPoint GS Imaging System	ThinPrep Imaging System	BD SurePath LBC	ThinPrep LBC
Primary reading	1.40 (1.38 to 1.42)	1.41 (1.39 to 1.43)	5.35 (5.32 to 5.37)	5.36 (5.33 to 5.40)
Primary rapid review	1.23 (1.20 to 1.25)	1.32 (1.31 to 1.33)	1.51 (1.50 to 1.52)	1.51 (1.50 to 1.52)
Checking	0.31 (0.29 to 0.32)	0.40 (0.38 to 0.42)	0.43 (0.41 to 0.45)	0.51 (0.49 to 0.54)
Secondary reading	0.36 (0.34 to 0.38)	0.40 (0.38 to 0.42)	0.47 (0.45 to 0.50)	0.48 (0.46 to 0.50)
Average total staff time per slide	3.29 (3.24 to 3.34)	3.53 (3.49 to 3.57)	7.76 (7.71 to 7.80)	7.87 (7.82 to 7.92)

TABLE 59 Average cytology reading time (minutes) per slide (95% CI) adjusted with probability of slide passing through different events – workload survey

Event	Automated arm		Manual arm	
	BD FocalPoint GS Imaging System	ThinPrep Imaging System	BD SurePath LBC	ThinPrep LBC
Primary reading	2.47 (2.44 to 2.51)	3.29 (3.26 to 3.33)	6.38 (6.33 to 6.43)	7.05 (7.02 to 7.09)
Primary rapid review	1.25 (1.22 to 1.28)	1.33 (1.32 to 1.34)	1.51 (1.50 to 1.52)	1.50 (1.49 to 1.51)
Checking	0.31 (0.29 to 0.33)	0.39 (0.38 to 0.41)	0.43 (0.41 to 0.45)	0.51 (0.49 to 0.53)
Secondary reading	0.36 (0.34 to 0.38)	0.39 (0.38 to 0.41)	0.48 (0.45 to 0.50)	0.48 (0.45 to 0.50)
Average total staff time per slide	4.38 (4.33 to 4.44)	5.41 (5.37 to 5.46)	8.80 (8.73 to 8.86)	9.55 (9.49 to 9.60)

The results show that automation reduces the time required to read a slide. Primary reading of slides takes longer, according to the workload survey (see *Table 59*), than using results from the time-and-motion survey (see *Table 58*), a difference that applies to both automated and manual screening.

The BD FocalPoint GS Imaging system, as well as identifying slides requiring NFR, groups slides requiring a primary read into five quintiles according to the likelihood that slides contain abnormal cells. *Table 60* reports the results of the average total time per slide by quintile using data from the time-and-motion study. The average times by quintile indicate that there are no significant differences in slide reading times by quintile.

TABLE 60 Average total staff time (minutes) per slide by quintile with the BD FocalPoint GS Imaging System – mean (95% CI)

	Slide ranking by quintile ^a				
	1	2	3	4	5
Average total staff time per slide	3.39 (3.21 to 3.57)	3.19 (3.03 to 3.36)	3.22 (3.04 to 3.40)	3.32 (3.14 to 3.49)	3.18 (3.02 to 3.35)

a Slides ranked as quintile 1 have the highest, and quintile 5 the lowest, likelihood of containing abnormal cells.

Average workload per year

The annual workload of cytoscreeners was estimated using the data on the average time to perform primary reading and rapid review from the workload surveys (see *Table 59*) and combining these data on the probability of slide reading events (see *Table 58*).

The average distribution of a cytoscreener's working hours (*Table 61*) shows that with automated-assisted screening cytoscreeners spend less time on primary screening, allowing them to perform more rapid review and other activities than when manual screening. This is as expected because automated primary screening is faster than manual screening. The daily and annual workloads are reported in *Table 62*. These data indicate that the volume of slides that one primary reader could process annually was substantially increased from 8511 slides a year when slides were read manually to over 14,246 slides per year with automated-assisted screening using the ThinPrep Imaging System. A further increase in the annual number of slides processed with automated assisted screening was observed with the BD FocalPoint GS Imaging System due to the fact that workload was reduced by the NFR feature. Translating this into staffing levels, for a laboratory processing 80,000 slides per year, only six full time cytoscreeners would be required with automated reading compared with eight or nine with manual reading.

Staff satisfaction

All five cytoscreeners undertaking the automated screening completed a staff satisfaction survey after they had been reading slides for nearly 3 years. The results of the survey are presented in *Appendix 11*.

Cytoscreeners were given training on how to read automated screening slides by the commercial companies. Three cytoscreeners rated the training as 'very good', and two as 'good' and 'fairly good'. Cytoscreeners were also asked if they had any recommendations about how the training could be improved. Three cytoscreeners did not have any recommendations; one asked for more training. There was also a recommendation from one cytoscreener that training needs to be devised for the staff to find out whether they are making mistakes on an ongoing basis. It was suggested that it would be beneficial to look at the mismatches between manual reading and automated reading as they arose; this was against the trial protocol and therefore did not happen.

When staff were asked about their overall preference between screening with automated and manual reading, only one cytoscreener was indifferent between the two options, the rest preferred manual screening (see *Appendix 11, Q6*). Regarding preferences between the two technologies, most staff preferred using the BD FocalPoint GS Imaging System to the ThinPrep Imaging System for primary screening (see *Appendix 11, Q7*). Similar preferences were observed for full manual review, where all staff strongly disagreed that they prefer using the ThinPrep Imaging System to the BD FocalPoint GS Imaging System (see *Appendix 11, Q8*). As expected from the above two responses, the majority of staff (four cytoscreeners) when asked about their overall preference for using the ThinPrep Imaging System compared with the BD FocalPoint GS

TABLE 61 Distribution of hours worked daily (7.5 hours)

Task	Automated arm		Manual arm	
	BD FocalPoint GS Imaging System	ThinPrep Imaging System	SurePath LBC	ThinPrep LBC
Primary screening	2.84	3.41	4.46	4.64
Rapid screening	2.51	2.18	1.36	1.27
Other duties	2.15	1.91	1.68	1.58

TABLE 62 Annual workload of primary screeners – mean (standard deviation)

Workload	Automated arm		Manual arm	
	BD FocalPoint GS Imaging System	ThinPrep Imaging System	BD SurePath LBC	ThinPrep LBC
Average number of slides read per day (primary screening)	75 (7)	66 (7)	42 (12)	40 (11)
Average number of slides read per year (primary screening)	16,063 (1471)	14,246 (1456)	9028 (2566)	8511 (2373)

Imaging System stated that overall they prefer using the BD FocalPoint GS Imaging System (see *Appendix 11, Q9*).

Four of the five cytoscreeners stated that they found it easier to concentrate using manual screening than automated screening (see *Appendix 11, Q10*). Similarly, three cytoscreeners reported that they found their work less challenging with the automated screening system and two found it more challenging (see *Appendix 11, Q11*). All cytoscreeners strongly agreed with the statement that their work was more monotonous using the automated reading system than using manual screening (see *Appendix 11, Q12*). When asked about physical discomfort using either the manual or automated system, all the cytoscreeners reported that they had experienced physical discomfort. Each respondent mentioned that he or she had experienced discomfort with using the ThinPrep Imaging System, whereas only one respondent mentioned some discomfort with the BD FocalPoint GS Imaging System (see *Appendix 11, Q13*). The discomfort mentioned about the ThinPrep Imaging System included that it was very noisy; the microscope was heavy and could not be adjusted to each individual cytoscreeners' need; it was not ergonomic; and it caused motion sickness, eye strain, muscle strain and back pain. There have been subsequent modifications to the review scope to address these issues, but the modified Review Scope Manual Plus (Hologic) was not utilised in the trial.

Only one respondent mentioned physical discomfort with the BD FocalPoint GS Imaging System, finding it led to workstation cramp and that the machine provided less space to work with. Two respondents also mentioned that there was some level of monotony and repetition in both the automated and manual systems, which could lead to fatigue and loss of concentration.

Cost analyses

Primary care costs

These costs apply to the resources involved in screening women in general practice surgeries or community clinics where cervical samples were taken for cytological examination and/or HPV testing. The two main resource components were administration (inclusive of postal invitations to attend for screening) and staff costs for screening consultations. Administration costs were obtained from the ARTISTIC report⁴⁹ and inflated to the 2007 financial year using

the HCHS index.⁷⁶ The mean duration of screening consultations was adopted from the English pilot studies⁴⁶ (13:45 minutes, 95% CI 12:25 to 15:05 minutes) and weighted according to the likelihood that a GP or a practice nurse would be the sample taker (80% of samples were taken by a nurse and 20% were taken by a GP).¹² Staff time was costed using the relevant staff cost per minute (*Table 63*).

The weighted unit cost of staff time for taking a sample was £0.92 per minute. Average duration of consultation during sample taking in primary care was evaluated using the weighted unit cost of staff time and we found that the average cost of taking a sample was £12.37 (95% CI £11.27 to £13.85). In addition to staff costs, there are additional costs such as the administrative cost for sending invitation letters to eligible women. Combining these costs together we find that the cost of primary care per sample is £15.90 (*Table 64*).

General practices send cervical sample vials to cytology laboratories. The cost of transporting cervical samples would remain unaffected given that general practices are normally served by a hospital transport system.⁴⁹

Cytology laboratory costs

In determining the mean costs for cytology samples, costs were divided between preparation and slide reading costs.

Preparation costs include the costs of laboratory equipment, consumables, maintenance and staffing needed for processing slides prior to slide reading. To retain confidentiality over prices these costs have been blinded between the two manufacturers.

For manual reading, BD SurePath preparation costs are based on the BD SurePath LBC system, BD SurePath LBC clinic kit and BD SurePath LBC laboratory kit. With the ThinPrep technology, several different types of slide preparation systems are available. Within the MAVARIC trial, samples were processed using the T3000 machine, and therefore the cost estimates were based on 5-year rental costs of this equipment, including consumable cost and maintenance cost.

The manufacturer of the BD FocalPoint GS Imaging System estimates that working 5.5 days per week the machine has an annual capacity of 100,000 samples per annum; operating 7 days a week, a throughput of 140,000 samples per annum can be achieved. (These outputs are greater than the system used for MAVARIC because of some updated software; that system has an output of 100,000 per annum on a 7 days a week basis.) The guide costs are based on a throughput of 120,000 samples per system per annum. With this system the microscope cost is not included,

TABLE 63 Cost per minute of sample taking staff

Staff	Sources	Unit cost per minute of consultation time (£)
GP	<i>Unit Costs of Health and Social Care – 2007</i> ⁵⁸	2.60
Practice nurse time	<i>Unit Costs of Health and Social Care – 2007</i> ⁵⁸	0.50

TABLE 64 Primary care costs item per slide

Cost items	Sources of resource use and cost data	Cost (£)
Invitation letter	Pilot estimate inflated from 2002 to 2007 costs ⁵⁸	3.53
Average cost of taking a sample	Pilot (weights) and <i>Unit Costs of Health and Social Care – 2007</i> ⁵⁸	12.37
Total primary care cost for taking a sample		15.90

therefore these costs were estimated based on existing laboratory microscope costs including the lease cost of equipment, cost of consumables and cost of maintenance.

The recommended annual capacity of the ThinPrep Imaging System working 40 hours per week is 75,000 slides per annum. The manufacturers have estimated that based on running the machine overnight, which does not require operator intervention, the machine can process 110,000 slides per year.

The costs of slide preparation are presented in *Table 65*. The results indicate that preparation costs are higher with automated technologies. Costs vary over range owing to differences in indicative prices between the manufacturers.

Cytology reading costs

The costs of reading and reporting LBC slides are based on the time-and-motion survey results and unit cost of staff time (based on the new pay system). The results reported in *Tables 54* and *55* show the duration of staff time in performing different activities related to LBC screening. *Table 62* shows average time per slide adjusted for the probability of a slide going through different slide reading events. Unit costs of relevant staff in cytology laboratory are given in *Table 66*.

In the cytology laboratory, once slides are prepared they are subject to primary screening. For automated cytology, samples need to be loaded and unloaded into the corresponding machine (ThinPrep Imaging System or BD FocalPoint GS Imaging System), which is done by MLAs. Time-and-motion study results show that 86% of the primary screening and 79% of the rapid review is carried out by cytoscreeners. Abnormal slides are sent for checking in the form of a further full interpretation. 'Checking' is usually carried out by higher grade BMSs in the laboratory. We have assumed that senior BMSs perform the checking activities. 'Checking' differs from rapid review as it involves a full rescreen. Secondary reading of slides is usually performed by the pathologist. The probability of each slide going through each reviewing process (reported in *Table 57*) is multiplied by duration of screening and unit cost of staff time to get the average unit cost of each reviewing stage in the LBC laboratory. Average primary review and rapid review costs are reported in *Tables 67* and *68*. The average costs for checking and secondary reading are given in *Table 69*.

TABLE 65 Unit cost of preparation: equipment, consumables, maintenance and staff time in cytology laboratory. Average between technologies (range)

	Automated arm (£)	Manual arm (£)
Total preparation/staff costs	3.85 (3.72–3.98)	2.97 (2.66–3.29)

TABLE 66 Cost per minute for cytology laboratory staff

Laboratory staff grade	Source	Cost/minute (£)
Band 2–3	Agenda for Change ⁷⁸ and Curtis ⁵⁸	0.18
Band 4	Agenda for Change ⁷⁸ and Curtis ⁵⁸	0.22
Band 5	Agenda for Change ⁷⁸ and Curtis ⁵⁸	0.27
Band 7	Agenda for Change ⁷⁸ and Curtis ⁵⁸	0.39
Band 8a	Agenda for Change ⁷⁸ and Curtis ⁵⁸	0.47
Band 8c	Agenda for Change ⁷⁸ and Curtis ⁵⁸	0.63
Band 9	ARTISTIC ⁴⁹ and Curtis ⁵⁸	1.29

TABLE 67 Average primary reading cost per slide in the LBC laboratory – mean (95% CI) – time-and-motion study

	Automated arm		Manual arm	
	BD FocalPoint GS Imaging System (£)	ThinPrep Imaging System (£)	BD SurePath LBC (£)	ThinPrep LBC (£)
Primary reading	0.32 (0.32 to 0.33)	0.32 (0.32 to 0.33)	1.23 (1.22 to 1.23)	1.25 (1.23 to 1.25)
Primary rapid review	0.31 (0.30 to 0.31)	0.33 (0.33 to 0.33)	0.38 (0.37 to 0.38)	0.38 (0.37 to 0.38)

TABLE 68 Average primary reading cost per slide in the LBC laboratory – mean (95% CI) – workload survey

	Automated arm		Manual arm	
	BD FocalPoint GS Imaging System (£)	ThinPrep Imaging system (£)	BD SurePath LBC (£)	ThinPrep LBC (£)
Primary reading	0.57 (0.56 to 0.58)	0.76 (0.75 to 0.77)	1.47 (1.46 to 1.48)	1.62 (1.62 to 1.63)
Primary rapid review	0.31 (0.30 to 0.32)	0.33 (0.33 to 0.34)	0.38 (0.37 to 0.38)	0.38 (0.37 to 0.38)

TABLE 69 Average checking and secondary reading cost per slide in the LBC laboratory – mean (95% CI)

	Automated arm		Manual arm	
	BD FocalPoint GS Imaging System (£)	ThinPrep Imaging System (£)	BD SurePath LBC (£)	ThinPrep LBC (£)
Checking	0.12 (0.11 to 0.13)	0.16 (0.15 to 0.16)	0.16 (0.15 to 0.17)	0.20 (0.19 to 0.21)
Secondary reading	0.46 (0.43 to 0.48)	0.51 (0.49 to 0.54)	0.61 (0.58 to 0.64)	0.62 (0.59 to 0.65)

We found that the costs of primary review were lower with automated screening than with manual reading, owing to savings in the amount of staff time to read each slide. A further cost saving (£0.01–0.06 per slide) was generated with the BD FocalPoint GS Imaging System compared with the ThinPrep Imaging System owing to savings in staff time related to the need to review fewer slides overall because of the NFR option.

The workload survey results are similar to the time-and-motion study in the overall ranking of costs between the technologies. These costs are slightly higher in all arms because the workload durations are longer as they account for some *x*-inefficiency. These results indicate that the cost savings are marginally lower when the workload survey results are used.

The results in *Table 69* show that the costs of both checking and secondary reading per slide are lower with automated screening even though the average times of checking and secondary reading are similar across manual and automated pathways. This is due to the fact that automated screening leads to lower numbers of slides being forwarded to checkers and medics.

The total cost per slide includes both the preparation cost and the cost of staff time to read a slide. The total cost per slide is reported in *Table 70*. The total cost per slide with automated screening varied from £5.05 to £5.17 when staff time estimates were based on time-and-motion survey data. The corresponding average costs per slide were higher with manual screening, varying from £5.35 to £5.41. When the estimates of staff time were based on the workload survey, all the estimates of cost per slide were slightly higher. Comparative costs between technologies should be treated with caution as the preparation costs (including indicative prices) have been blinded. When assessing the range of potential costs (reflecting potential prices) with the minimum price

TABLE 70 Total cost per slide (95% CI), [range]

Item	Automated arm		Manual arm		
	BD FocalPoint GS Imaging System (£)	ThinPrep Imaging System (£)	BD SurePath LBC (£)	ThinPrep LBC (£)	
Average preparation cost ^a	3.85 [3.72 to 3.98]		2.97 [2.66 to 3.29]		
Average staff cost	Time-and-motion survey	1.20 (1.17 to 1.23)	1.32 (1.29 to 1.35)	2.37 (2.34 to 2.41)	2.44 (2.40 to 2.47)
	Workload survey	1.46 (1.43 to 1.49)	1.75 (1.72 to 1.78)	2.63 (2.59 to 2.67)	2.81 (2.78 to 2.85)
Total cost	Time-and-motion survey	5.05 [4.95–5.15]	5.17 [5.07–5.27]	5.35 [5.03–5.71]	5.41 [5.06–5.76]
	Workload survey	5.31 [5.21–5.41]	5.60 [5.50–5.70]	5.60 [5.25–5.96]	5.78 [5.44–6.14]
Minimum price difference ^b	Time-and-motion survey	4.95	5.07	5.71	5.76
	Workload survey	5.21	5.50	5.96	6.14
Maximum price difference ^c	Time-and-motion survey	5.15	5.27	5.03	5.06
	Workload survey	5.41	5.70	5.25	5.44

a Blinded to maintain price confidentiality over consumables and equipment.

b Defined as from the lowest automated price to the highest manual price.

c Defined as from the highest automated price to the lowest manual price.

between automated and manual screening, automated screening is also cost saving. In contrast, when the maximum price difference is used, automated reading is more expensive than manual screening.

Human papillomavirus testing costs

The costs of HPV testing incurred at the HPV laboratory are reported in *Table 71*. The costs of HPV testing include equipment, consumables and staff time. HPV testing on a BD SurePath LBC sample costs £16.75 and on a ThinPrep LBC sample will cost £16.94.

The small difference in HPV test costs arises from the HPV protocol required for the two LBC technologies, prior to testing with the HC2 test. ThinPrep samples were aliquoted into a labelled tube and sample conversion buffer added, as the initial stage of sample processing. With BD SurePath samples, only the original tube sample was required to be labelled and checked for adequate volume before further processing.

The total transport cost was £4206. The average number of samples sent per batch was 28 (total of 107 batches). The average cost of transport was £38.92 per batch and £1.39 per sample. There was capacity for 100 samples to be transported within the transport containers used. At full capacity, this would have brought the cost of sample transport to £0.42.

Cost of colposcopy, histology and cancer treatment

The costs of colposcopy, histology and cancer treatment were identified from published literature and are reported in *Table 72*. Slides that were borderline or mild and above in the final MR were sent for HPV testing. It was assumed that inadequate slides incurred a further sample taking and slide reading cost at the laboratory. The cost per woman also included the cost of colposcopy referral and treatment of CIN where required based on the clinical data.

The rates of different outcomes that determine the cost per woman are reported in *Table 73* based on an analysis of the clinical trial data set. Rates of events are similar between arms. These data reflect the clinical findings of the trial, in that there was a lower rate of referral to colposcopy and detection of CIN lesions with automated technology.

TABLE 71 Human papillomavirus test costs: equipment, consumables, maintenance and staff time

Process	SurePath LBC (£)	ThinPrep LBC (£)
HPV test cost ^a	16.75	16.94

a Includes LBC laboratory cost for identifying and packing samples and transport cost.

TABLE 72 Cost of histology outcome and cancer treatment

Clinical activity	Cost (£)	Reference
LBC test cost in laboratory ^a	5.46–5.72	Table 70
Consult cost – GP/nurse visit in community	15.91	Table 64
HPV test reflex costs ^a	16.75–16.94	Table 71
No CIN	282.76	Martin-Hirsch <i>et al.</i> ⁵⁰
CIN1	432.39	
CIN2	590.28	
CIN3	625.37	
Stage 1 invasive cancer	2874.02	
Stage 2 invasive cancer	4590.17	
Stage 3 invasive cancer	12,963.53	
Stage 4 invasive cancer	13,185.40	

a Sample cost for automated screening includes loading and unloading cost.

TABLE 73 Rates of inadequate samples, HPV testing, diagnosis and treatment

Item	Automated arm	Manual arm
Inadequate sample	1.91%	2.99%
Negative	93.63%	91.43%
HPV test	4.29%	5.54%
Colposcopy		
No CIN	1.76%	2.08%
CIN1	0.43%	0.49%
CIN2	0.48%	0.54%
CIN3	0.75%	0.78%

The total cost per woman was very similar for manual and automated reading where the same technology was used (Table 74). The overall costs per woman are higher with the ThinPrep technologies either where slides are read manually or with automated screening, reflecting slightly higher referral rates than with colposcopy where no CIN or CIN1 was detected.

Table 75 indicates that the average cost per case detected between automated and manual screening were very similar. These data incorporate both the total cost per woman including slide reading and downstream costs.

Figure 10 presents the results of the bootstrapping exercise on the incremental cost per case of CIN2+ detected. These data reflect the uncertainty in the comparative cost and event outcomes

TABLE 74 Total cost per woman screened including inadequate samples/HPV testing/colposcopy/± biopsy (95% CI)

Item	Automated arm (£)	Manual arm (£)
Inadequate sample and negative	20.53 (20.49 to 20.57)	20.49 (20.44 to 20.53)
HPV test	0.72 (0.69 to 0.76)	0.93 (0.90 to 0.97)
Colposcopy	4.97 (4.60 to 5.33)	5.88 (5.49 to 6.28)
No CIN		
CIN1	1.85 (1.58 to 2.13)	2.11 (1.82 to 2.41)
CIN2	2.85 (2.45 to 3.25)	3.16 (2.74 to 3.58)
CIN3	4.68 (4.15 to 5.20)	4.91 (4.37 to 5.45)
Total cost per woman, average prices	35.60 (34.79 to 36.42)	37.48 (36.63 to 38.34)

TABLE 75 Cost per case of CIN2+ and CIN3+ detected (95% CI)

Item	Automated arm	Manual arm
Average cases of CIN2+ per 1000 women	12.31 (11.23 to 13.39)	13.21 (12.09 to 14.32)
Average cases of CIN3+ per 1000 women	7.48 (6.63 to 8.32)	7.85 (6.99 to 8.71)
Cost per case of CIN2+ detected	£2892 (£2720 to £3098)	£2838 (£2676 to £3030)
Cost per case of CIN3+ detected	£4762 (£4378 to £5245)	£4775 (£4400 to £5244)

by random sampling from the trial data. The majority of results are in the south-west quadrant, indicating that automated screening is less effective in the detection of CIN2+, but is also cost saving. Approximately a quarter of the results are in the south-east quadrant, indicating that, in these random samples of the trial data, automated screening is both cost saving and more effective. A similar picture is seen in *Figure 11*, where the detection of CIN3+ is used as the outcome measure. In line with the main clinical results there is correlation between CIN detection rates and costs. This is reflective of the fact that where relatively less CIN is detected the costs are lower owing to reductions in treatment costs.

In *Figures 12* and *13* the results of the bootstrapped results have been plotted on cost-effectiveness acceptability curves. These figures indicate the probability that manual screening is cost-effective compared with automated screening for different willingness-to-pay thresholds for detecting additional cases of CIN2+ and CIN3+. In the baseline results we have used the average price of manual and automated screening. However, given the uncertainty about prices and the need to blind price differences between the two manufacturers we have also presented curves for the minimum price difference and maximum price difference.

Given a willingness to pay of £5000 for each case of CIN2+ detected, these data indicate that there is an 80% chance that manual screening is cost-effective compared with automated screening using average prices between the two manufacturers. As detailed previously, there is a high degree of uncertainty reflected in the minimum and maximum price differences between automated and manual screening. The probability of manual being cost-effective rises to around 97% at the maximum price difference and falls to around 25% at the minimum price difference.

In *Figure 13*, cost-effectiveness estimates are presented using CIN3+ as an outcome measure. This figure indicates greater uncertainty in the cost-effectiveness of manual reading compared with automated reading. Decision-makers would need to be willing to pay an additional £12,500 per additional case of CIN3+ detected to have even a 70% level of certainty that manual reading was more cost-effective than automated screening. With the minimum price difference between

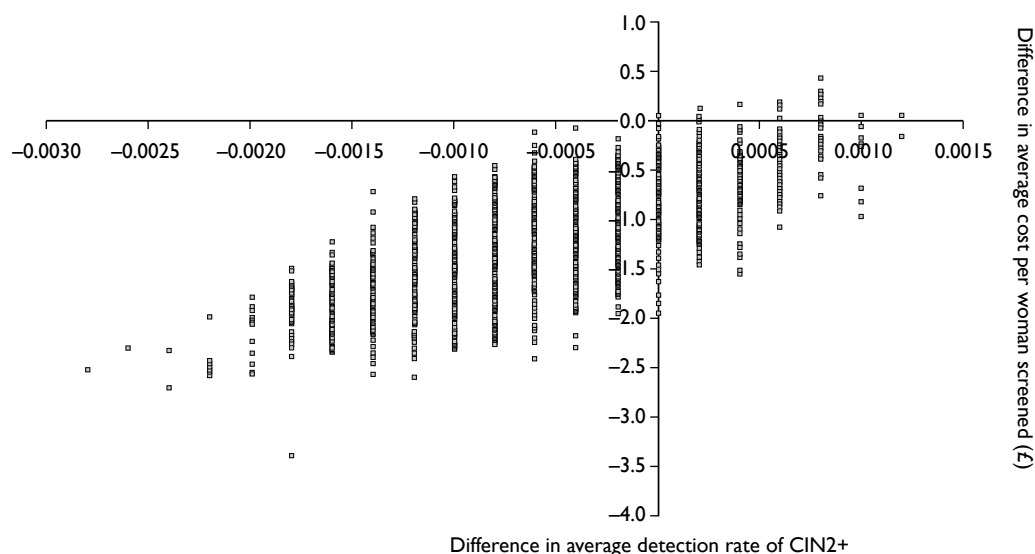


FIGURE 10 Incremental cost per case of CIN2+ detected for automated reading compared with manual reading (bootstrapped results).

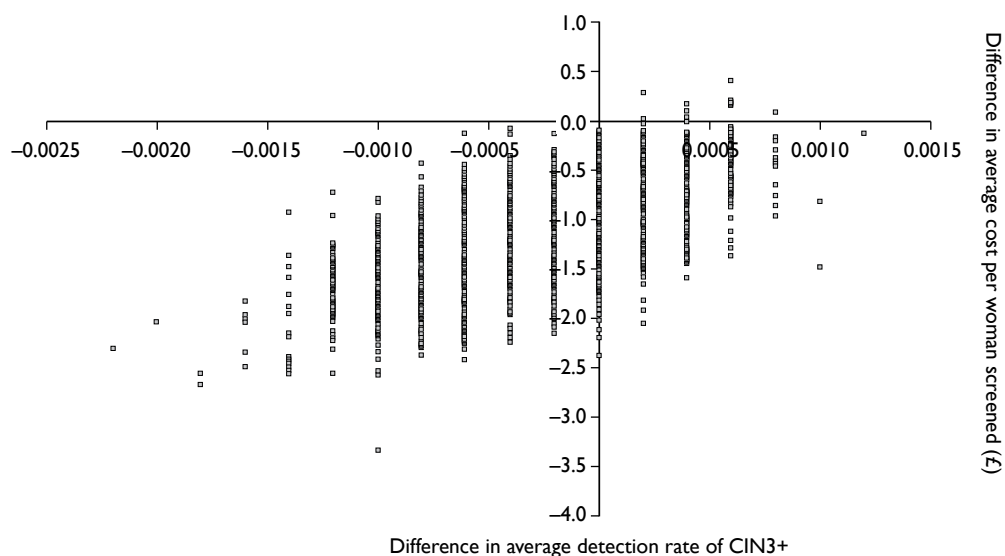


FIGURE 11 Incremental cost per case of CIN3+ detected for automated reading compared with manual reading (bootstrapped results).

automated and manual screening, the likelihood that manual screening is cost-effective falls to just under 50% at a willingness to pay of £12,500 per additional CIN3+ case.

Costs and cost-effectiveness of the Becton Dickinson FocalPoint Slide Profiler as a stand-alone device

Further analyses were conducted on the difference in the average cost of reading the slides that were identified as requiring NFR with the BD FocalPoint Slide Profiler compared with manually reading the same slides (Tables 76 and 77). The average cost per slide including staff and preparation costs was lower when the BD FocalPoint Slide Profiler was used as a stand-alone device utilising the 'NFR' option than with manual reading, regardless of whether slides were rapid reviewed. However, slightly fewer cases of CIN2+ and CIN3+ were also identified.

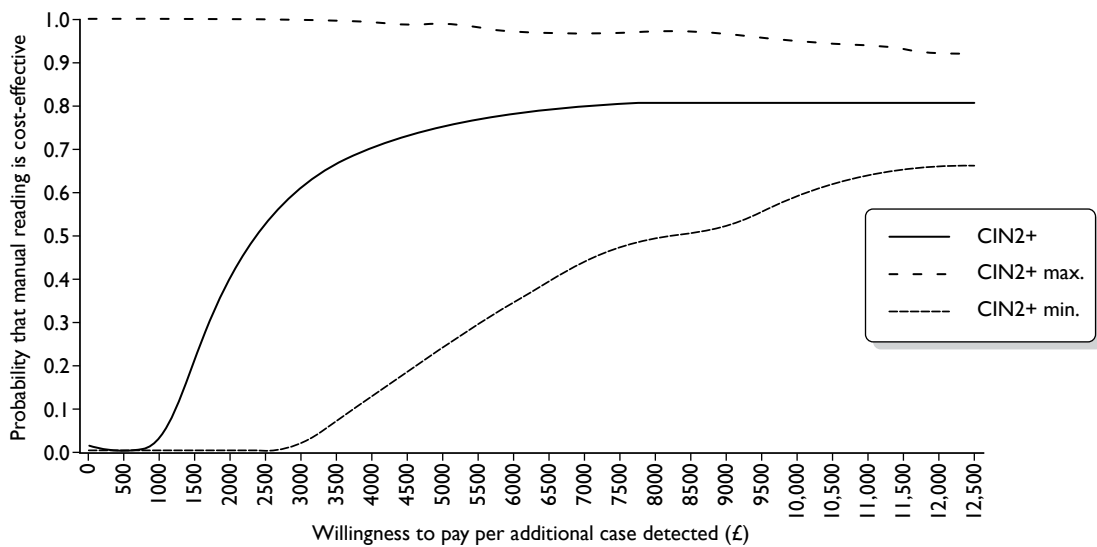


FIGURE 12 Cost-effectiveness acceptability curve for manual reading compared with automated reading for the detection of CIN2+. Min and max represent the minimum and maximum price difference between automated and manual preparation costs.

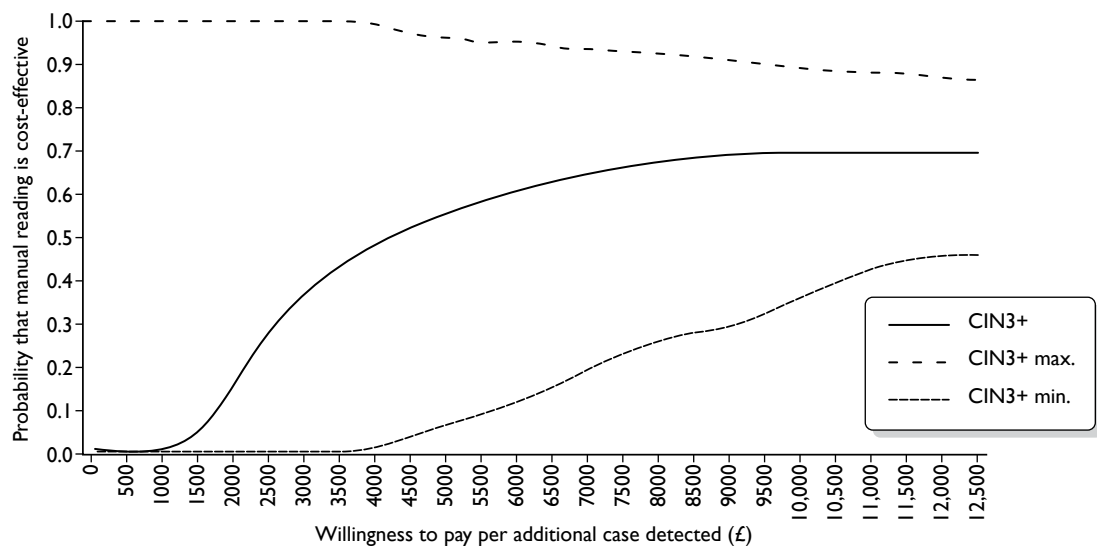


FIGURE 13 Cost-effectiveness acceptability curve for manual reading compared with automated reading for the detection of CIN3+. Min and max represent the minimum and maximum price difference between automated and manual preparation costs.

Figures 14 and 15 report the cost per case detected of CIN2+ and CIN3+ for manually reading slides compared with using the 'NFR' option on the imager. Manual reading in this case would be cost-effective compared with NFR if decision-makers were willing to pay £2500 per additional case of CIN2+ or £6000 per additional case of CIN3+ detected.

Further analyses were conducted to compare the cost of manually screening slides compared with using the 'NFR' option on the BD FocalPoint Slide Profiler and not reading slides ranked either quintile 5 or quintiles 4 and 5. These results are presented in Tables 78 and 79. Again, the overall cost of not reading slides identified with the BD FocalPoint SlideProfiler as either quintile 5 or

TABLE 76 Comparison of costs and outcomes for manual reading versus the BD FocalPoint Slide Profiler with rapid review

Item	Manual			BD FocalPoint Slide Profiler (with rapid review)		
	Mean	Lower CI	Upper CI	Mean	Lower CI	Upper CI
Staff cost	£2.63	£2.59	£2.67	£1.46	£1.43	£1.49
Preparation cost	£2.66	£2.66	£2.66	£3.72	£3.72	£3.72
Cost per slide	£5.29	£5.25	£5.33	£5.18	£5.15	£5.21
Cost per women	£36.02	£34.82	£37.22	£34.42	£33.28	£35.57
CIN2+ per 1000	13.09	11.49	14.69	12.42	10.85	13.98
CIN3+ per 1000	7.71	6.47	8.94	7.45	6.24	8.66
Cost per CIN2+	£2630	£2897	£2421	£2901	£3208	£2663
Cost per CIN3+	£4466	£5139	£3978	£4835	£5583	£4297

TABLE 77 Comparison of costs and outcomes for manual reading versus the BD FocalPoint Slide Profiler without rapid review

Item	Manual			BD FocalPoint Slide Profiler (without rapid review)		
	Mean	Lower CI	Upper CI	Mean	Lower CI	Upper CI
Staff cost	£2.63	£2.59	£2.67	£1.39	£1.35	£1.42
Preparation cost	£2.66	£2.66	£2.66	£3.62	£3.62	£3.62
Cost per slide	£5.29	£5.25	£5.33	£5.01	£4.97	£5.04
Cost per women	£36.02	£34.82	£37.22	£34.27	£33.12	£35.42
CIN2+ per 1000	13.09	11.49	14.69	12.42	10.85	13.98
CIN3+ per 1000	7.71	6.47	8.94	7.45	6.24	8.66
Cost per CIN2+	£2630	£2897	£2421	£2760	£3052	£2534
Cost per CIN3+	£4466	£5139	£3978	£4601	£5311	£4089

TABLE 78 Comparison of costs and outcomes for manual reading versus the BD FocalPoint Slide Profiler for slides identified as quintile 5

Item	Manual			BD FocalPoint Slide Profiler (without rapid review)		
	Mean	Lower CI	Upper CI	Mean	Lower CI	Upper CI
Staff cost	£2.51	£2.38	£2.63	£1.33	£1.21	£1.44
Preparation cost	£2.66	£2.66	£2.66	£3.62	£3.62	£3.62
Cost per slide	£5.17	£5.04	£5.29	£4.95	£4.83	£5.06
Cost per women	£60.19	£53.51	£66.87	£54.30	£47.97	£60.63
CIN2+ per 1000	29.99	21.15	38.82	27.89	19.36	36.43
CIN3+ per 1000	19.53	12.36	26.70	18.83	11.79	25.87
Cost per CIN2+	£1811	£2268	£1562	£2158	£2764	£1836
Cost per CIN3+	£2781	£3882	£2271	£3197	£4540	£2585

quintiles 4 and 5 is lower; however, also slightly fewer cases of CIN2+ and CIN3+ were identified. As shown in *Figures 16* and *17* these data indicate that utilising the BD FocalPoint Slide Profiler to identify slides in quintiles 4 and 5 and then not reading them is unlikely to be cost-effective if decision-makers are willing to pay > £2500 per case of CIN2+ detected.

TABLE 79 Comparison of costs and outcomes for manual reading versus the BD FocalPoint Slide Profiler for slides identified as quintiles 4 and 5

Item	Manual			BD FocalPoint Slide Profiler (without rapid review)		
	Mean	Lower CI	Upper CI	Mean	Lower CI	Upper CI
Staff cost	£2.61	£2.51	£2.70	£1.39	£1.30	£1.48
Preparation cost	£2.66	£2.66	£2.66	£3.62	£3.62	£3.62
Cost per slide	£5.27	£5.17	£5.36	£5.01	£4.92	£5.10
Cost per women	£58.17	£53.37	£62.97	£53.32	£48.75	£57.89
CIN2+ per 1000	31.92	25.35	38.48	29.74	23.40	36.09
CIN3+ per 1000	19.95	14.73	25.17	18.50	13.47	23.53
Cost per CIN2+	£1671	£1923	£1504	£1956	£2281	£1745
Cost per CIN3+	£2673	£3311	£2300	£3145	£3964	£2676

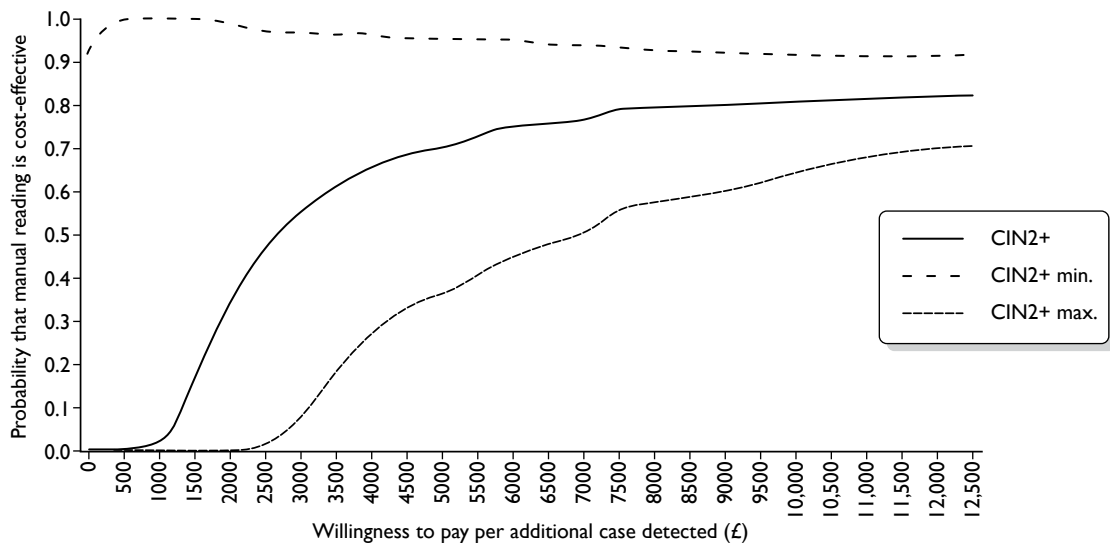


FIGURE 14 Cost-effectiveness acceptability curve for NFR compared with manual reading for the detection of CIN2+.

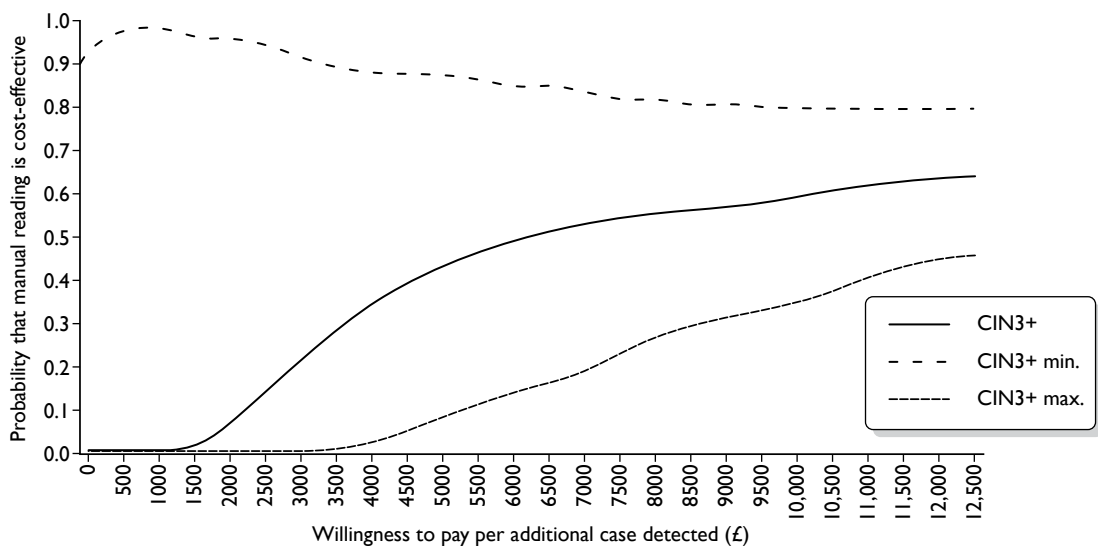


FIGURE 15 Cost-effectiveness acceptability curve for NFR compared with manual reading for the detection of CIN3+.

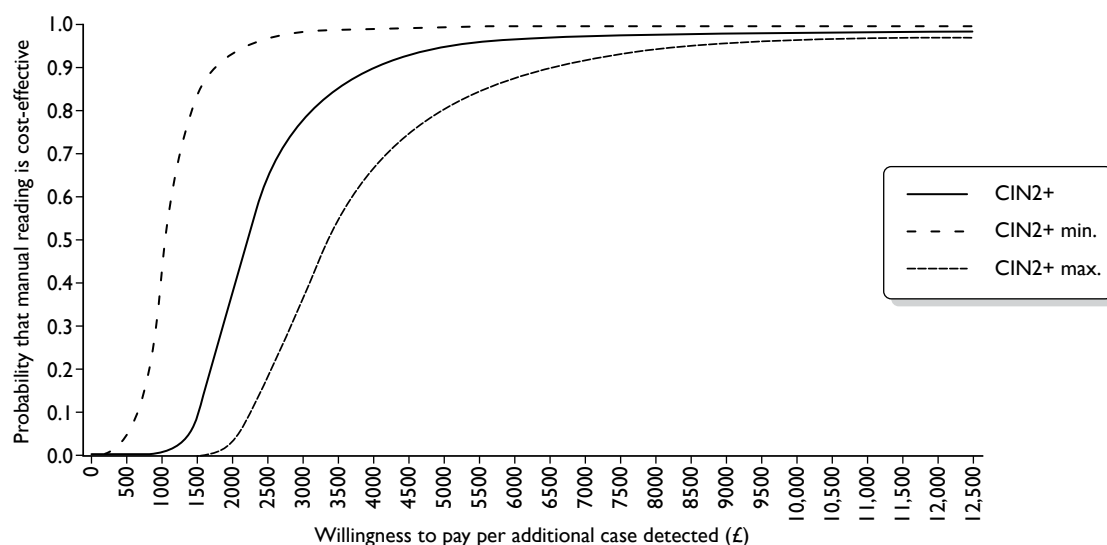


FIGURE 16 Cost effectiveness acceptability curve for manual reading compared with not reading slides identified as quintiles 4 and 5 for the detection of CIN2+.

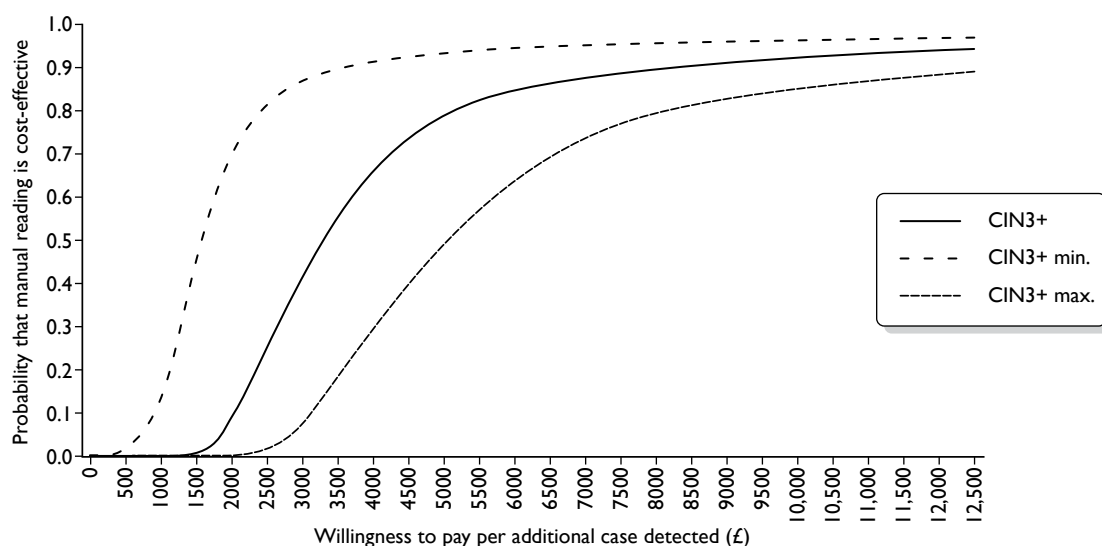


FIGURE 17 Cost-effectiveness acceptability curve for manual reading compared with not reading slides identified as quintiles 4 and 5 for the detection of CIN3+.

Lifetime modelling results

Table 80 shows the predicted lifetime costs and effects of each LBC strategy in a simulated cohort of 10,000 women. Costs and effects were discounted at 3.5% for the first 30 years and 3% thereafter. Modelling beyond trial end points predicts that automated LBC results in a small cost saving over the lifetime of a woman (approximately £12.60 per woman, discounted) and also a small loss in life-years (4.52 per 10,000 women, or approximately 4 hours per woman, discounted; Table 81). The predicted decrease in life-years associated with automated LBC, compared with manual, is primarily driven by the slight loss in sensitivity. If automated LBC was current practice, manual LBC would be associated with an incremental cost-effectiveness ratio of £27,863 per life-year saved. This is above the £20,000 per QALY figure where current NICE recommendations strongly favour adoption, but it is below the £30,000 figure where rejection is favoured.⁷⁹

TABLE 80 Baseline estimates of lifetime costs and effects associated with different LBC test technologies (per 10,000 women)

Strategy	Lifetime cost (discounted)	Life-years (discounted)	QALYs (discounted)
Manual LBC	£1,820,306	268,240	268,046
Automated LBC	£1,694,394	268,235	268,062

TABLE 81 Baseline estimates of incremental costs, incremental effects and incremental cost-effectiveness ratios associated with automated compared with manual LBC

Item	Automated LBC, compared with manual LBC
Incremental cost	–£125,912 per 10,000 women, or –£12.59 per woman (cost saving compared with manual LBC)
Incremental LYS	–4.52 per 10,000 women, or –4 hours per woman (less effective than manual in terms of life-years)
Incremental QALYs	15.83 per 10,000 women (more effective than manual LBC in terms of QALYs)
ICER (life-year saved)	Manual LBC costs £27,863 per life-year saved compared with automated LBC (automated LBC is cost saving but less effective for life-year saved)
ICER (QALY)	–£7592 per QALY gained (automated LBC is cost saving and, when quality of life is taken into account, is also more effective, ^a i.e. dominates manual LBC)

a It should be noted that quality-of-life weights for health states were obtained from the international literature and may not accurately reflect the priorities of women in the UK, therefore there is considerable uncertainty in this finding.

When QALYs are used as the outcome measure, modelling predicted that automated LBC is associated with an increase of 15.83 QALYs per 10,000 women (or approximately 13.9 quality-adjusted life-hours per woman; *Table 81*) compared with manual LBC. This finding is sensitive to choice of quality-of-life weights, but remained >0 in all cases examined during sensitivity analysis. The increase in QALYs is driven by a small increase in specificity (which decreases the disutility resulting from false-positives). When QALYs are used as the outcome measure, automated LBC dominates manual LBC as a strategy, as it is both cost-saving and more effective. It should be noted that the quality of life weights for health states were obtained from the international literature and rely on assumptions about the duration of disutility.^{67,69,80–84} It should also be noted that there is some evidence from willingness-to-pay studies that UK women express preference for improved sensitivity. These findings have not been formally integrated into a cost-utility framework, but may increase the uncertainty surrounding these results.

Modelling was also used to predict cancer outcomes over the lifetime of a cohort of 10,000 women, assuming current screening recommendations, intervals and compliance. In this cohort of 10,000 women, manual LBC was associated with 69 cancer cases and 12 cancer deaths, and automated LBC with 72 cancer cases and 13 cancer deaths (*Table 82*) over the cohort's lifetime.

Sensitivity analysis

Sensitivity analysis was conducted to explore the impact of various model assumptions. Parameters investigated during sensitivity analysis were LBC price; test characteristics of automated LBC relative to manual; HC2 price; test characteristics of HC2; utility set used to estimate QALYs; screening and follow-up compliance assumptions; proportion of CIN1 treated; and discount rate (see *Appendix 12*).

Automated LBC remained cost saving compared with manual LBC in all cases examined during sensitivity analysis, including when the unit cost of automated LBC was higher than that for manual LBC. The cost saving for a cohort of 10,000 women over their lifetimes ranged from £103,366 to £197,625 (discounted) in sensitivity analysis, or approximately £10 to £20 per

TABLE 82 Estimates of lifetime health outcomes associated with different LBC test technologies (cohort of 10,000 women)

Strategy	Cancer cases	Cancer deaths
Manual LBC	69	12
Automated LBC	72	13

woman. Predicted cost savings were larger when the unit cost saving associated with automated LBC was assumed to be highest (automated is £0.69 cheaper than manual LBC), and when attendance was assumed to be perfect. Predicted cost savings were smaller when automated LBC was more expensive (£0.21) than manual LBC, when its relative performance was worse, and when HC2 was assumed to have lower positivity rates.

Automated LBC was associated with a small loss in life-years in all cases examined during the sensitivity analysis. Manual LBC resulted in an additional 3.1–8.9 life-years saved per 10,000 women (or approximately 2.7–7.8 hours per woman) compared with the automated LBC (discounted). The value of life-years saved by manual LBC was most sensitive to assumptions regarding the accuracy of automated LBC relative to manual LBC.

Excluding discount rate (which is discussed below), manual LBC resulted in an additional 3.1–8.9 life-years saved per 10,000 women (or approximately 2.7–7.8 hours per woman) compared with automated LBC in sensitivity analysis (discounted). The value of life-years saved by manual LBC was most sensitive to assumptions regarding the relative accuracy and relative costs between automated and manual reading. With the highest relative test performance of manual compared with automated reading, the incremental cost per life-year saved was £11,881, suggesting that manual reading would be highly cost-effective. Conversely, with the worst relative test performance, the estimate of incremental life-years saved was £36,229, suggesting that manual reading would not be a cost-effective intervention compared with automated reading as it is above the NICE threshold for acceptance (based only on life-years). The estimates of the cost per life-year saved varied between £21,688 and £30,019 when the minimum and maximum relative costs were used respectively.

Automated LBC was always associated with a small increase in QALYs in all cases examined during sensitivity analysis and the cost per QALY results always suggested that automated reading dominates manual reading with higher QALYs and lower costs. This finding is due to the potential negative effects associated with follow-up and treatment. The increase in QALYs ranged from 2.1 to 21.1 QALYs per 10,000 women (or approximately 1.8 to 18.5 quality-adjusted life-hours per woman, discounted). QALYs gained were smaller when the disutility associated with various health states was assumed to be smaller, when the accuracy of automated LBC relative to manual LBC was lower, when HC2 positivity was assumed to be lower or when the initial discount rate was higher. QALYs gained were increased when compliance with screening and follow-up appointments was assumed to be perfect, when the specificity of colposcopy was assumed to be higher and when the disutility associated with various health and screening states was greater.

Changes in the discount rate had a significant impact on the incremental costs and effects, but over a broad range of discount rates the model still predicted that automated LBC would be associated with cost savings, a small loss in life-years and a small gain in QALYs. When the discount rate was increased to 6% for the first 30 years after age 10 (rather than 3.5%), the cost saving for the cohort decreased to £75,677 (or approximately £7.57 per woman), the loss in life-years decreased to 2.1 (or approximately 1.9 hours per woman), and the increase in QALYs

associated with automated LBC was reduced to 10.38 (or approximately 9.1 quality-adjusted life-hours per woman). If automated LBC was current practice, manual LBC would be associated with an additional cost of £35,345 per life-year saved. When no discounting was used, cost savings for the cohort associated with automated LBC increased to £287,431 (approximately £29 per woman). Life-years lost with automated LBC compared with manual for the cohort increased to 27.7 (or approximately 1 day per woman). The increase in QALYs associated with automated LBC was virtually unchanged compared with the base case, at 15.78. If no discounting was assumed and automated LBC was current practice, manual LBC would be associated with an additional cost of £10,375 per QALY.

Chapter 4

Discussion

This rigorous study of automation-assisted reading of cytology has indicated that, relative to manual reading, automation was found to be 8% less sensitive in the detection of CIN2+. This is considered to be inferior in performance to manual reading of cervical cytology. This trial was designed to provide as robust and unbiased a comparison as possible of automation-assisted reading with manually read cervical cytology. The study therefore comprised a standard manually read arm, to simulate 'true to life' cytology reporting, and a matched pairs arm in which the same slide was read using both automated technology and manual reading. The statistical power of the study lay in the comparison of the matched readings, with the standard arm providing a control for the manual reading in the matched pairs. This study was also designed to provide a comparison between the two types of LBC used in the NHSCSP – ThinPrep and BD SurePath. We could not randomise individually for this because the LBC used is general practice based, so in order to avoid possible bias in terms of disease prevalence, practices were stratified for deprivation index, ensuring matching in terms of this parameter between ThinPrep LBC and BD SurePath LBC. A further parameter that could introduce bias was the staining of the ThinPrep LBC slides for use with the ThinPrep Imaging System, as the system required a darker stain than that used routinely. All ThinPrep slides were therefore stained similarly in order to avoid any possible difference in sensitivity, due to the stain, between the manual-only and paired arms. This measure was designed to blind cytoscreeners as to whether the manual reading was paired with the automated read or was manual read only.

The power of the study required 50,000 matched pairs and originally we had planned a further 50,000 manual-only slides. In the event it became clear that the duration of the study would permit only 75,000 samples to be accumulated, so after 23 months, and after consulting with the Data Monitoring and Ethics Committee and the Trial Steering Committee, we changed to a 1:3 randomisation between the manual-only and paired arms which we calculated would maintain the target of 50,000 paired readings and provide 25,000 manual-only readings. This was considered to be sufficient to achieve the purpose of the control arm and maintain the statistical power of the study.

The histological primary outcome measure was considered essential because it is detection of CIN2+ which results in standard treatment designed to prevent cervical cancer. The other major feature of the design was the use of HPV triage, the purpose of which was to maximise detection of underlying CIN as quickly as possible. Triage minimised loss of power which could have resulted from default or delay in women returning for repeat cytology if this had been the criterion for referral to colposcopy. In order to determine whether automation should replace manual reading, we sought a direct comparison of the ability of both methods of reading cytology to identify those women who should be referred for colposcopic diagnosis and treatment. Our design of prospective double reading of the same slide together with immediate referral and diagnosis provided an accurate and true to life comparison. The use of the MR based on the more abnormal reading between manual and auto ensured the colposcopic assessment of women with prospectively identified discordant results, most of whom had either a negative auto or manual reading. We could not determine true sensitivity, which would have required all screened women to undergo colposcopic assessment, but our design permitted the sensitivity of one method relative to the other to be directly determined. No other studies have employed this approach. Studies that measure cytological abnormality rates could be considered too indirect to be reliable.

Split sample studies are more reliable than consecutive cohorts, but are not as analytically accurate as reading the same slide. Finally, the real life reliability of the manually read slides as a comparator with automated reading was provided by the manual-only arm in this study which provided similar results to those in the paired arm.

Cytology outcomes

The profile of cytology results in terms of grade and age was very similar to that reported nationally. Only 5.0% of the samples were from women outside the routine screening age range in England. There were proportionately fewer samples from women aged ≥ 45 years, with approximately one-third of the samples from women aged 25–34 years, one-third from women aged 35–44 years and one-third from women aged 45–64 years. One reason for this may have been a disproportionate number of women aged < 45 years attending because of the public impact by a celebrity who died from cervical cancer in 2009 aged 27 years amid a blaze of national publicity. The Manchester Cytology Centre received a marked increase in cervical cytology samples during the last 6 months of accrual to this study – a phenomenon seen in many parts of the country. Despite the inclusion of additional samples from colposcopy clinics early in the study, the distribution of cervical cytology by grade was almost identical to national reporting. This implies that the outcomes from this study will be generalisable nationally. Our study power calculations were based on prior national data, and we therefore achieved the anticipated power; we planned for 630 cases of CIN2+ based on manual reading, which was actually slightly surpassed. Because the majority of colposcopy clinic samples accrued early in the study, prior to the change in randomisation proportion, there was over-representation of cytological abnormalities in the manual-only arm. We therefore undertook additional analyses restricted to routine screening samples. In doing this, there was still adequate power for the primary outcome of relative sensitivity. Any potential ascertainment bias between automated and manual reading should have been avoided because the final MR was based on the more severe of the paired readings.

Colposcopy referral

The process of HPV triage involved samples being sent to Edinburgh for HPV testing, and this worked very efficiently, demonstrating that a similar arrangement using a hub-and-spoke model for national implementation of triage would be feasible. The overall rate of colposcopy performed (among all ages) was 2254/48,271 (4.7%) in the paired arm and 1123/24,566 (4.6%) in the manual-only arm. These are higher proportions than those currently observed in the routine NHSCSP, because triage led to $> 50\%$ of borderline/mild cytology being referred directly for colposcopy. This had the advantage of ensuring a high degree of colposcopic ascertainment of underlying disease, potentially enhancing the sensitivity of cytology. The MR with respect to borderline and mild dyskaryosis cytology showed identical HPV-positive rates between ThinPrep and BD SurePath (66%), which ensured similar triage referral to colposcopy. This is reassuring not only for comparison of these two systems, but also for the national programme, which uses both types of LBC.

Summary cytology data in the paired arm

The primary outcome was based on the comparison of the final cytology result for manual and automated reading. The assignment of readings 1 and 2 and then a final reading reflects the real-life process of slide checking in the cytology laboratory: an initial slide report then rapid review

or checking and referral to a cytopathologist when there is an abnormality. In both automated and manual reading there was a fall in overall borderline/mild dyskaryosis between readings 1 and 2 and the final reading. This reflects routine experience when reporting cervical cytology. This trend was not seen in the reporting of higher grades of cytology, which is known to have a lower interobserver variation than that seen in low-grade abnormalities.⁸⁵

The most significant difference within the paired readings was the proportion with borderline/mild dyskaryosis on final reading: 4.2% for automated and 5.5% for manual reading. These final results reflect differences in the first reading, AR1 and MR1, which were 5.3% and 7.5% respectively. This difference in final reading results meant that, compared with automated reading, manual reading resulted in more samples being HPV triaged and, of these, women who were HPV positive were referred onwards to colposcopy. This can be seen in *Table 22* in that 15% of 317 HPV-positive women who were borderline/mild on FMR and negative on FAR had CIN2+, similar to the proportion seen in the NHSCSP pilot study. This suggests that cases 'missed' on automated reading were not less significant samples classified as borderline/mild, but were representative of the borderline/mild cytology being subjected to triage in the NHSCSP Pilot Study. Discordant pairs that were FMR borderline/mild and FAR negative outweighed by a factor of more than two the 125 HPV-positive FAR borderline/mild cases reported as negative on final manual reading, of whom 16% (20/125) were found to have underlying CIN2+. There were also 12 cases of high-grade cytology on FAR which were reported negative on FMR, and 47 cases of high-grade cytology on FMR reported negative on FAR which also resulted in a net increase in CIN2+ detection for FMR compared with FAR. It is notable that there were six query invasive results in FMR, which were negative on FAR. As shown in *Chapter 3* (see *Table 49*) this amounted to 31 cases of CIN2+ detected only on automated and 83 cases of CIN2+ detected only on manual reading. This net detection of 52 lesions in favour of manual reading represents 7.6% of the total 687 CIN2+.

These discordant results between FMR and FAR are reflected in the much smaller proportion of FMR negatives than FAR negatives, in the presence of an abnormal MR: FMR 0.67% (294/43,647) versus FAR 2.08% (931/44,771).

Primary outcome

The primary outcome of the trial was the sensitivity of automated reading relative to that of manual reading within the paired arm to detect CIN2+. The outcome that automated reading was significantly less sensitive than manual reading was not expected. The 8% inferiority in terms of relative sensitivity exceeds the pre-specified limit and is too great for the rates to be considered clinically equivalent. The study was powered to demonstrate non-inferiority, which was defined as a true absolute difference of < 5% inferior in sensitivity to detect CIN2+. The observed relative sensitivity of 92% is equivalent to an absolute difference of 5% or more for values of the true sensitivity of manual screening of 65% or higher. The absolute sensitivity of manual reading with LBC for low-grade squamous intraepithelial lesion or worse (LSIL+) has been estimated as 79%;⁵⁷ under this assumption a relative sensitivity of automated reading of 92% is equivalent to an absolute difference of 6.3%.

Automated reading was also relatively less sensitive in the detection of CIN3+ by a margin of 5%. There was no pre-specified non-inferiority limit set for CIN3+ as it was not the primary outcome measure, but in terms of cancer prevention this could not be considered clinically acceptable, even in the presence of cost savings. Assessments of effectiveness in screening need to be considered with costs because cost-effectiveness in the detection of cervical lesions is important in evaluating overall performance of automated versus manual screening. It is unlikely that

automation appeared relatively less sensitive because of a bias in the detection rate in the manual reading in the paired arm, given the measures taken to conceal whether the manually read slides were in the paired or manual-only arm. In fact, the detection rates of CIN2+ and CIN3+ were higher in the manual-only arm, providing further evidence that there was little likelihood of a higher than expected detection of CIN in the manual readings in the paired arm. The colposcopy rate in the manual-only arm was 6.6% (1626/24,566) compared with 5.7% (2751/48,271) in the paired arm, which may have been due to the higher proportion of abnormal samples in the manual-only arm. Within the paired arm, colposcopy verification will have affected the automated versus manual outcomes in cases of mismatch between negative auto/non-negative manual or vice versa, resulting in colposcopy referral.

The achievement of the additional sensitivity of the manual reading was at the expense of a small drop in specificity, related to automation-assisted reading. This meant that additional colposcopies were required following manual reading compared with automated, but the 19% PPV of the additional procedures was at the high end of the range achieved following HPV triage of low-grade cytology, and was therefore considered worthwhile.

Discordant results

The review of the discordant pairs was designed to try to explain the reason for automated reading 'missing' abnormalities that were associated with CIN2+. The review also included the discordant pairs for which manual reading did not detect abnormalities picked up by automated reading. Low- and high-grade abnormalities were separated in the analysis. In one-quarter of cases no abnormality was seen, suggesting an auto location error, i.e. the abnormalities had not been shown on the FOVs. In a similar review by Halford *et al.*,¹⁶ discordant readings from a split sample study, 31/37 cases of auto misses were found to contain abnormalities on the FOVs not detected in the initial read, and in the majority of these 31 cases abnormal cells were seen in only 5 out of 22 FOVs. There are reasons why automated reading could result in false-negative reports. Peripheral location of abnormal cells in the FOVs was found in a number of cases, also noted by Halford *et al.*'s¹⁶ study. This is not a problem in manual screening given the practice of overlapping fields in routine screening practice, which is lost in location-guided screening. The nature of automated screening means that it is more monotonous, a view expressed by several staff, which could result in lower levels of vigilance. The use of new equipment presents challenges for staff used to their own workstation. It is not considered to be a learning curve issue, however, because discordant pairs occurred at an equal distribution throughout the 3 years of the study, during which staff gained considerable experience.

The most recent large-scale evaluation of automation-assisted cervical screening has just been reported from the Scottish Cytology Network for the ThinPrep Imaging System.³⁷ Around 110,000 samples were randomly allocated to either manual or automated reading using the ThinPrep Imaging System. Samples were not double read to provide direct comparison. The primary outcome measure was based on cytology grade rather than histopathology. This therefore represented a comparison of the effectiveness in detection of abnormal cytology rather than measuring comparative performance in detecting CIN. In Phase 1 of the Scottish evaluation, abnormal slides were read in the 'Autoscan' mode and in Phase 2 they were removed and read on a standard microscope. The proportion of high-grade abnormality detected was similar (1.38% manual, 1.45% auto, $p=0.512$). Sensitivity was defined as the proportion of abnormalities picked up on the first read compared with that after rapid review and checking. By this criterion, the results showed similar sensitivity in both arms (93.7% manual, 91.79% auto, p -value for high grade = 0.09). There were also similar values for low grade (7.5% manual, 7.87% auto). It is noteworthy that there was variation in these proportions of abnormality between the six participating laboratories. For manual reading the low-grade range was 4.48%–9.18% and the high-grade range was 1%–1.8%.

In the MAVARIC study paired arm, the proportion of low grades was 5.5% for manual and 4.5% for the ThinPrep Imaging System, both at the low end of the Scottish range. Similarly the proportion of high grades was 1.27% for manual and 1.14% for the ThinPrep Imaging System, again at the low end of the Scottish range. The authors of the Scottish study reported that their results indicated that automated screening would be safe and more efficient.³⁷ Cost-effectiveness was not formally evaluated and would depend on the costs of the system. It was not designed to provide the relative diagnostic performance determined in MAVARIC, and as such the two studies are not directly comparable. The abnormality rates are of interest though again not directly comparable because screening in Scotland begins at 20 years compared with 25 years in England and cytology abnormalities are particularly common in the age group 20–24 years.

One of the limitations of MAVARIC is that it was conducted in a single laboratory; however, the routine reporting from the Manchester Cytology Centre is very much in the mid-range of abnormality rates and PPVs for high-grade cytology as reported in the NHSCSP Statistical Bulletin.⁶

As stated in *Chapter 1* and shown in *Table 1*, previous studies comparing automated with manual reading have tended to indicate higher rates of cytological abnormality and some have found increased rates of CIN2+. Some of these studies^{23,27} have been performed simply comparing cytological abnormality rates in consecutive periods of time, and others have used a split sample design,^{16,24,26} whereby the same sample has been split between a conventional or LBC slide which is manually read and a slide which is subjected to automated reading. Many studies do not use histology as an outcome.^{17,18,25,29–32,34,35}

Secondary outcomes

The primary objective was to compare manual and automated-assisted reading of cervical cytology, but the study design provided an opportunity to compare BD SurePath with ThinPrep LBC and the BD FocalPoint GS Imaging System with the ThinPrep Imaging System. When both arms of the study were combined and data restricted to routinely obtained samples in women aged 25–64 years, BD SurePath had a higher detection rate than ThinPrep for both CIN2+ and CIN3+ (1.5% vs 1.25% and 0.85% vs 0.71% respectively). When the automated readings using the two systems were compared, the sensitivity of the BD FocalPoint GS Imaging System was not statistically different from that of the ThinPrep Imaging System, relative to manual reading in the paired arm, for either CIN2+ or CIN3+. There were fewer FMR positive/FAR negative results associated with CIN3+ for the ThinPrep Imaging System, but these were not sufficient to achieve a statistically significant higher sensitivity against manual reading than the BD FocalPoint GS Imaging System. It must be pointed out that the study was not formally powered to compare BD SurePath and ThinPrep.

The NFR facility of the BD FocalPoint SlideProfiler performed well in correctly identifying slides that had negative outcomes, particularly when NFR was applied to routinely obtained slides. The majority of CIN2+ which would have been missed on NFR in this study did not involve routine screening slides. For routinely obtained slides NFR looks very promising as a means of reducing the number of slides that need to be read by a cytoscreener.

Economic analysis

The study provides a detailed comparative assessment of productivity in the laboratory. It clearly demonstrates that primary reading is substantially quicker with automated equipment than a manual approach. We used two different methods to observe changes in primary reading times: time-and-motion studies and workload surveys. The time-and-motion studies indicated

that slides could be read three to four times quicker with automated screening, but there was no significant difference in reading times between the two technologies with either manual or automated reading ($p=0.14$).

By contrast, the workload survey data estimated longer slide reading times than the time-and-motion study results, suggesting some inefficiency and further administration time not captured within the time-and-motion data. It is likely that these data provide a better reflection of the productivity gains that may be achievable in a real-life setting. These data suggest that eight or nine slides can be read per hour with manual reading, compared with 19 or 20 with automated reading.

Other studies have largely been conducted on the ThinPrep Imaging System and show comparable results, although they use a range of timing methodologies and the comparator is sometimes not LBC but conventional slides. One study²⁶ reported slide reading times for the ThinPrep Imaging System of 3.4 minutes compared with 7.4 minutes for manual reading of conventional slides. An Australian study²⁴ also found that the number of slides read per hour was significantly increased with ThinPrep Imaging System-assisted reading compared with conventional slides. The mean within-reader difference was 7.2 slides per hour. Only one study³² has been identified that compared BD FocalPoint GS Imaging System-assisted screening with manual screening and it was found that interpretation time was reduced by 40%.

In addition to assessing the times for primary screening, we also estimated the overall implications for laboratory staff productivity. With the automated technologies there was a slight decrease in referral for review by checkers and medics. For primary screening the NFR option led to further increases in productivity with the BD FocalPoint GS Imaging System. Over a 7.5-hour working day we estimated an overall increase in productivity for cytoscreeners of between 60% and 80%.

One study²⁹ found slightly higher productivity increases with the ThinPrep Imaging System-assisted method, with an estimate that the rate of slides screened was typically doubled over an 8-hour day. Another study²⁶ found that the ThinPrep Imaging System-assisted screening led to a 27% productivity gain when compared with manual screening with LBC, a smaller gain than observed in our study.

With automated screening and reductions in primary screening time, the average proportion of a cytoscreener's time spent on rapid review increases, which has a significant impact on cost. Potentially, further cost savings could be made with automated screening by changing the rapid review protocols. The BD FocalPoint GS Imaging System also flags up at least 15% of all successfully processed negative or inadequate slides for QC review.

MAVARIC has produced unbiased and comparable productivity estimates across manual and automated technologies. The study has compared both BD FocalPoint GS Imaging System and ThinPrep Imaging System technologies with their manual counterparts. The slides were blinded between arms and therefore led to unbiased results, which indicate that use of automation in screening in the UK can reduce the average time taken to process a slide. The key area for savings in time is the primary screen.

The results of the staff satisfaction study indicate that staff prefer manual reading to automated technologies. Some physical and ergonomic discomfort was noted particularly with the ThinPrep Imaging System, although subsequently there has been some redesigning of the technology to address this. Another issue highlighted was the monotony of automated reading, although some cytoscreeners noted that manual and automated reading both had elements of monotony.

Comparative data on the cost per slide indicate that the additional costs associated with the automated equipment are offset by savings in the costs of staff time. For confidentiality purposes the price of equipment was blinded between the two manufacturers, but with one manufacturer the additional equipment costs were more than offset by time savings and therefore automated reading became cost saving compared with manual, whereas with the other manufacturer the additional costs were not completely offset by staff time savings. Averaging across these indicative prices, automated screening was slightly less expensive per slide than reading slides manually once staff savings were taken into account. However, these estimates are sensitive to the price of the equipment and, where the maximum price difference was used between automated and manual reading, overall it cost more to read slides using automated equipment. It should be noted that these price estimates are based on automated machines operating at maximum capacity. As the volume of slides required to operate at full capacity is higher than observed in the NHS Cancer Screening Programme, national implementation of automated cytology would require careful consideration of the need for amalgamation of existing laboratories or alternative ways of configuring or delivering services, in order to maximise efficient use of the technology. In addition, the costs of training were covered by manufacturers and it is unknown if this would be the case if automated technology were rolled out nationally. Assessment of the overall cost per woman screened including downstream costs associated with treatment of CIN, colposcopy and HPV testing indicated very similar costs between manual and automated screening from each manufacturer.

Within-trial analysis of the main trial results indicated that there is an 80% chance that manual reading is cost-effective compared with automated reading (using average prices between the two manufacturers), given a willingness to pay of £5000 for each additional case of CIN2+ detected. These results were sensitive to the price of automation, and at the minimum price difference between technologies decision-makers would need to be willing to pay £8500 per additional case of CIN2+ detected for manual reading to remain cost-effective.

Further analyses evaluated the use of the BD FocalPoint Slide Profiler as a stand-alone device with manual reading, either used to identify slides requiring NFR with or without rapid review, or not to be required to read slides in the quintiles with the lowest risk of abnormal cells. Our results indicated that when using the equipment in this way, cost savings were generated; however, slightly fewer cases of CIN2+ and CIN3+ were detected. With the NFR option only, manual reading would remain cost-effective if decision-makers were willing to pay £2500 per case of CIN2+ detected. Again, utilising the BD FocalPoint Slide Profiler to identify slides in quintiles 4 and 5 and then not reading them is unlikely to be cost-effective if decision-makers were willing to pay at least £2500 for each additional case of CIN2+ detected by manual reading. These analyses have not been applied to NFR for routine samples only.

The results of the lifetime modelling of the cost-effectiveness of automation compared with manual reading show uncertainty about the relative cost-effectiveness of automation compared with manual reading. If automated LBC was current practice, manual LBC would be associated with an incremental cost-effectiveness ratio of £27,863 per life-year saved. This is above the £20,000 per QALY figure at which current NICE recommendation strongly favour adoption, but it is below the £30,000 figure above which interventions are likely to be rejected on cost-effectiveness grounds.⁷⁹ One-way sensitivity analysis indicated that these results are highly sensitive to the relative test performance between manual and automated reading with estimates ranging from £11,881 to £36,229 per life-year saved.

Quality-adjusted life-year estimates were also derived from modelling, and these indicated that automated reading might produce a small QALY gain due to the difference in specificity and potential disutility associated with 'overtreatment' of lesions that might regress. This finding

remained the same in all options explored in the sensitivity analysis, including when minimum values were used for disutility and relative test performance. These QALY results should, however, be treated extremely cautiously, as the empirical evidence on utilities came not from the trial but from the international literature.^{67,69,80-84} In particular, the true duration of disutility for women associated with overtreatment of pre-invasive cervical cancer lesions is difficult to determine. In particular, more data are required on the true value and duration of the disutilities (reported in *Table 96*) associated with treatment for CIN and colposcopy referral: in the model these apply for the 6-month cycle in which the event occurs. It may be that disutility from a false-positive result is shorter. Further studies from the UK, evaluating women's overall preferences between cervical cancer screening strategies with comparatively higher sensitivity, at the cost of lower specificity, indicated an overall preference for comparative gains in sensitivity when traded with lower specificity.^{86,87} While we have performed extensive one-way sensitivity analysis on the modelling results, we have not performed a probabilistic sensitivity analysis. The modelling exercise suggests, however, that the key area of uncertainty for drawing more affirmative conclusions on the true cost-effectiveness of automated compared with manual reading rests with the need for improved understanding and empirical research on the quality-of-life implications and women's preferences for trading for improvements in sensitivity.

A published systematic review¹⁴ provided an analysis of automation in cervical screening programmes in the UK. This review of cost-effectiveness studies indicated strong limitations in the evidence used to populate previous models and therefore lack of certainty about any conclusions. Our results have significantly reduced the uncertainty relating to the costs of automated compared with manual reading, but substantial uncertainty remains concerning lifetime quality-adjusted survival.

Implications for the NHS Cervical Screening Programme

Despite the potential for increased throughput in slides, by shortening the reading times, there is no evidence that automation produced any clinical benefit. Indeed, automation-assisted reading achieved 8% less sensitivity relative to manual reading in the detection of CIN2+, which is deemed to warrant treatment in cervical screening programmes. It was also less sensitive than manual in the detection of CIN3+. There is strong evidence that automation significantly increases productivity in the laboratory, generating savings in the cost of staff time, but incurs additional equipment costs. There is variation in the indicative prices of automated equipment. Given the minimum price difference between the technologies, automation would be less expensive than manual reading. Modelling results indicate that the relative cost-effectiveness of manual and automated reading is in the threshold area of uncertainty where NICE would have difficulty in reaching firm conclusions based on economic evidence alone. Without clear-cut benefit in terms of specificity and cost-effectiveness, the increased productivity of automation in reducing pressure on the cytology screening service cannot be achieved at the expense of a significant reduction in sensitivity. The analysis of the discordant pairs revealed missed abnormalities and these were more frequent in auto reading.

However, the observation that the NFR has a high and clinically acceptable NPV provides a basis for considering the use of the NFR option in primary screening as a potential means of not having to read up to 25% of samples. Indeed, had NFR been restricted to routinely obtained slides, as recommended by the FDA, < 1% of CIN2+ detected would have been missed. Rapid review did not appear to add significantly to the detection of lesions among the group classified as NFR. If NFR alone were used there would be no need for the clinical workstations' component of the automated equipment. Although the cost per slide was cheaper than manual reading,

because fewer cases of CIN2+ and CIN3+ were detected, manual screening would be cost-effective given a willingness to pay of £2500 per additional CIN2+ detected.

Within 5 years in England, women vaccinated in the national catch-up programme will be invited for screening; in countries with a younger screening threshold this has already begun or is imminent. Vaccinated women can be expected to have a 60%–70% reduction in CIN2+, which will affect the rate of abnormal cytology and raises concerns that vigilance may be lessened and the predictive value of cytology reduced. Automation might be considered helpful in this regard by drawing cytoscreeners' attention to abnormal areas, and using the 'NFR' facility in the BD SurePath automated system to reduce the number of negative cytology slides; currently the ranking facility selects around 20% for NFR. In a screened population with lower rates of CIN it might be possible to envisage NFR being applied to a larger proportion of slides. The impact of vaccination on low-grade cytological abnormalities will, however, be less than high-grade owing to the broad range of high-risk HPV types associated with mild abnormalities. An alternative scenario is that in the postvaccination era, HPV testing could provide the means to filter out the large majority of HPV-negative women who would be at negligible risk over the next screening interval and, by restricting cytology to HPV-positive women, the proportion of abnormal slides would be somewhat similar to present levels, or perhaps even greater owing to the bias presented to the cytoscreener by the knowledge of a positive HPV test.

Research recommendations

Further research could be carried out to develop strategies for avoiding the non-detection of low-grade as well as high-grade abnormalities. Only a small proportion of CIN2+ cases missed with automated reading were due to location error. This may be related to peripheral distribution of abnormal cells in the FOVs, but further investigation is warranted.

The following studies could be recommended for NFR:

1. Follow-up studies on those samples that were reported as NFR and negative on manual reading would be useful in terms of 3-year follow-up and rate of HPV detection.
2. A vaccinated population can be expected to have an increased rate of negative cytology and the cost-effectiveness of NFR might increase if a larger proportion of screened women could have their slides archived without further reading.
3. The effect of NFR in an HPV-positive screened population would be worthy of further investigation.

A cost-effectiveness analysis of NFR for routine screening samples would also be recommended.

It would be relevant to have additional insight into the quality-of-life implications for women subjected to cervical screening strategies with varying levels of sensitivity and specificity. In particular, more data are required on the true duration of disutility associated with treatment for CIN and colposcopy referral.

Acknowledgements

The MAVARIC Trial Study Group

Chief investigators

- HC Kitchener, Clinical Principal Investigator, School of Cancer and Enabling Sciences, University of Manchester.
- S Moss, Statistics and Epidemiology, CSEU, Institute of Cancer Research, Sutton.

Trial co-ordinators

- R Albrow, School of Cancer and Enabling Sciences, University of Manchester.
- J Mather, Manchester Cytology Centre, Central Manchester University Hospitals NHS Foundation Trust.

Epidemiology/Statistics

- R Blanks, CSEU, Institute of Cancer Research, Sutton.
- G Dunn, Health Science Research Group, University of Manchester.

Data management

- L Gunn, CSEU, Institute of Cancer Research, Sutton.
- E O'Brien, CSEU, Institute of Cancer Research, Sutton.

Cytopathology

- M Desai, Manchester Cytology Centre, Central Manchester University Hospitals NHS Foundation Trust.
- DN Rana, Manchester Cytology Centre, Central Manchester University Hospitals NHS Foundation Trust.

Virology

- H Cubie, Specialist Virology Centre, Royal Infirmary of Edinburgh.
- C Moore, Specialist Virology Centre, Royal Infirmary of Edinburgh.

Health economics

- R Legood, Health Services Research Unit, London School of Hygiene and Tropical Medicine; Health Economics Research Centre, University of Oxford.
- A Gray, Health Economics Research Centre, University of Oxford.
- Z Sadique, Health Services Research Unit, London School of Hygiene and Tropical Medicine.

This study was conducted under the guidance of a Steering Committee. The independent members are as follows:

- Professor David Torgerson, Independent Chair, Alcuin Research Resource Centre, Department of Health Sciences, University of York.
- Dr Maggie Cruickshank, School of Medicine, University of Aberdeen/Department of Obstetrics and Gynaecology, Aberdeen Maternity Hospital.
- Dr Karin Denton, Department of Cellular Pathology, Southmead Hospital, North Bristol NHS Trust.

The study Data Monitoring and Ethics Committee comprised the following members:

- Professor Paula Williamson, Chairperson, Centre for Medical Statistics and Health Evaluation, University of Liverpool.
- Dr John Smith, Department of Histopathology, Sheffield Teaching Hospitals NHS Foundation Trust.
- Mr Patrick Walker, Department of Obstetrics and Gynaecology, Royal Free Hampstead NHS Trust.

Cost-effectiveness modelling beyond the study end points was performed using a model previously developed by a group at Cancer Council New South Wales (NSW), Australia and adapted for this project in collaboration with Dr Rosa Legood. We are grateful to Dr Karen Canfell, Jie Bin Lew, Megan Smith and Robert Walker at Cancer Council NSW for their assistance in accessing the model, conducting the lifetime modelling and drafting the document sections relating to this.

We are grateful to Yvonne Hughes, Laboratory Manager, for her co-operation and effort in accommodating the MAVARIC study in the Manchester Cytology Centre.

We acknowledge the use of a BD FocalPoint GS Imaging System provided free of charge by Source Bioscience (formerly Medical Solutions) for part of the study.

We thank Qiagen for providing substantially discounted HC2 kits.

Contributions of authors

Professor Henry Kitchener (Professor of Gynaecological Oncology) was the Chief Investigator for the study. He contributed to the conception and design of the study, the interpretation of the data and drafting the final report and gave final approval to publish.

Dr Roger Blanks (Epidemiologist) contributed to the conception and the design of the study, the analysis and interpretation of the data, revising the final report and gave final approval to publish.

Dr Heather Cubie (Director, Scottish HPV Reference Laboratory) contributed to the conception and design of the study, the interpretation of the data, revising the final report and gave final approval to publish.

Dr Mina Desai (Consultant Cytopathologist/Clinical Head) contributed to the conception and design of the study, revising the final report and gave final approval to publish.

Professor Graham Dunn (Professor of Biomedical Statistics) contributed to the conception and design of the study, the analysis and interpretation of data, revising the final report and gave final approval to publish.

Dr Rosa Legood (Lecturer in Decision Modelling, Health Economics) contributed to the conception and design of the study, the analysis and interpretation of data, drafting the final report and gave final approval to publish.

Professor Alastair Gray (Professor of Health Economics/Director of Health Economics Research Centre) contributed to the conception and design of the study, the interpretation of data, revising the final report and gave final approval to publish.

Dr Zia Sadique (Research Fellow, Health Economics) contributed to the design of the economic study, the analysis and interpretation of data, revising the final report and gave final approval to publish.

Dr Sue Moss (Reader in Epidemiology/Associate Director of CSEU) contributed to the conception and design of the study, the analysis and interpretation of data, drafting the final report and gave final approval to publish.

References

1. Peto J, Gilham C, Fletcher O, Matthews F. The cervical cancer screening epidemic that screening has prevented in the UK. *Lancet* 2004;**364**:8.
2. National Health Service Cancer Screening Programmes. *NHS Cervical Screening Programme Annual Review 2008*. Sheffield: NHS Cancer Screening Programmes; 2008.
3. National Health Service Cancer Screening Programmes. *HPV Sentinel Sites implementation project*. 2008. URL: www.cancerscreening.nhs.uk/cervical/hpv-sentinel-sites.html (accessed on 22 December 2009).
4. NHS Cancer Screening Programmes. Achievable standards, benchmarks for reporting and criteria for evaluating cervical cytopathology. Sheffield: NHS Cancer Screening Programmes; 2000.
5. NHS Cervical Screening Programme. *Laboratory Organisation: A Guide For Laboratories Participating in the NHS Cervical Screening Programme*. Sheffield: NHS Cancer Screening Programmes; 2003.
6. The Health and Social Care Information Centre. *Cervical Screening Programme, England 2008–09*. London: The Health and Social Care Information Centre; 2009.
7. The Health and Social Care Information Centre. *Cervical Screening Programme, England 2007–08*. London: The Health and Social Care Information Centre; 2008.
8. Department of Health. *Cancer Reform Strategy*. London: Department of Health; 2007.
9. Lord Carter of Coles. *Report of the Review of NHS Pathology Services in England*. London: Department of Health; 2006.
10. Advisory Committee on Cervical Screening. Extraordinary meeting to re-examine current policy on cervical screening for women aged 20–24 years taking account of any new evidence and make recommendations to the National Cancer Director and Ministers held on 19 May 2009. Sheffield: NHS Cancer Screening Programmes; 2009.
11. Department of Health. *New Cervical Cancer campaign*. 2009. URL: www.wired-gov.net/wg/wg-news-1.nsf/0/9B54517FC34EBBE8802575DF00506FF2 (accessed on 25 June 2009).
12. Moss SM, Gray A, Marteau T, Legood R, Henstock E, Maissi E. *Evaluation of HPV/LBC Cervical Screening Pilot Studies - Report to the Department of Health (revised October 2004)*. Sutton: Institute of Cancer Research; 2004.
13. National Institute for Clinical Excellence. *Guidance on the use of liquid-based cytology for cervical screening (Technology Appraisal 69)*. London: National Institute for Clinical Excellence; 2003.
14. Willis BH, Barton P, Pearmain P, Bryan S, Hyde C. Cervical screening programmes: can automation help? Evidence from systematic reviews, an economic analysis and a simulation modelling exercise applied to the UK. *Health Technol Assess* 2005;**9**(13).
15. Broadstock M. Effectiveness and cost effectiveness of automated and semi-automated cervical screening devices: a systematic review of the literature. *N Z Health Technol Assess Rep* 2000;**3**(1).
16. Halford JA, Batty T, Boost T, Duhig J, Hall J, Lee C, *et al*. Comparison of the sensitivity of conventional cytology and the ThinPrep Imaging System for 1,083 biopsy confirmed high-grade squamous lesions. *Diagn Cytopathol* 2010;**38**:318–26.

17. Wilbur DC, Black-Schaffer WS, Luff RD, Abraham KP, Kemper C, Molina JT, *et al.* The Becton Dickinson FocalPoint GS Imaging System: clinical trials demonstrate significantly improved sensitivity for the detection of important cervical lesions. *Am J Clin Pathol* 2009;**132**:767–75.
18. Pacheco MC, Conley RC, Pennington DW, Bishop JW. Concordance between original screening and final diagnosis using imager vs. manual screen of cervical liquid-based cytology slides. *Acta Cytol* 2008;**52**:575–8.
19. Papillo JL, St John TL, Leiman G. Effectiveness of the ThinPrep Imaging System: clinical experience in a low risk screening population. *Diagn Cytopathol* 2008;**36**:155–60.
20. Passamonti B, Bulletti S, Camilli M, D'Amico MR, Di Dato E, Gustinucci D, *et al.* Evaluation of the FocalPoint GS system performance in an Italian population-based screening of cervical abnormalities. *Acta Cytol* 2007;**51**:865–71.
21. Lozano R. Comparison of computer-assisted and manual screening of cervical cytology. *Gynecol Oncol* 2007;**104**:134–8.
22. Troni MG, Cariaggi MP, Bulgaresi P, Houssami N, Ciatto S. Reliability of sparing Papanicolaou test conventional reading in cases reported as No Further Review at AutoPap assisted cytological screening - Survey of 30,658 cases with follow up cytological screening. *Cancer Cytopathol* 2007;**111**:93–8.
23. Miller FS, Nagel LE, Kenny-Moynihan MB. Implementation of the ThinPrep® Imaging System in a high-volume metropolitan laboratory. *Diagn Cytopathol* 2007;**35**:213–17.
24. Davey E, d'Assuncao J, Irwig L, Macaskill P, Chan SF, Richards A, *et al.* Accuracy of reading liquid based cytology slides using the ThinPrep Imager compared with conventional cytology: prospective study. *BMJ* 2007;**335**:31.
25. Schledermann D, Hyldebrandt T, Ejersbo D, Hoelund B. Automated screening versus manual screening: A comparison of the ThinPrep® Imaging System and Manual Screening in a Time Study. *Diagn Cytopathol* 2007;**35**:348–52.
26. Roberts JM, Thurloe JK, Bowditch RC, Hyne SG, Greenberg M, Clarke JM, *et al.* A Three-Armed Trial of the ThinPrep Imaging System. *Diagn Cytopathol* 2007;**35**:96–102.
27. Dziura B, Quinn S, Richards K. Performance of an Imaging System vs. manual screening in the detection of squamous intraepithelial lesions of the uterine cervix. *Acta Cytol* 2006;**50**:309–11.
28. Bulgaresi P, Cariaggi MP, Troni MG, Ciatto S. Quality control of the AutoPap screening system employed as a primary screening device: Rapid review of the smears coded as no further review. *Tumori* 2006;**92**:276–8.
29. Biscotti CV, Dawson AE, Dziura B, Galup L, Darragh T, Rahentulla A, *et al.* Assisted primary screening using the automated ThinPrep Imaging System. *Am J Clin Pathol* 2005;**123**:281–7.
30. Parker EM, Foti JA, Wilbur DC. FocalPoint slide classification algorithms show robust performance in classification of high-grade lesions on SurePath liquid-based cytology slides. *Diagn Cytopathol* 2004;**30**:107–10.
31. Stevens MW, Milne AJ, Parkinson IH, Nespolon WW, Fazzalari NL, Arora N, *et al.* Effectiveness of AutoPap system location-guided screening in the evaluation of cervical cytology smears. *Diagn Cytopathol* 2004;**31**:94–9.
32. Ronco G, Vineis C, Montanari G, Orlassino R, Parisio F, Arnaud S, *et al.* Impact of the AutoPap (currently FocalPoint) primary screening system location guide use on interpretation time and diagnosis. *Cancer Cytopathol* 2003;**99**:83–8.

33. Confortini M, Bonardi L, Bulgaresi P, Cariaggi MP, Cecchini S, Ciatto S, *et al.* A feasibility study of the use of the AutoPap Screening System as a primary screening and location guided rescreeing device. *Cancer Cytopathol* 2003;**99**:129–34.
34. Wilbur DC, Parker EM, Foti JA. Location-guided screening of liquid-based cervical cytology specimens - a potential improvement in accuracy and productivity is demonstrated in a preclinical feasibility trial. *Am J Clin Pathol* 2002;**118**:399–407.
35. Vassilakos P, Carrel S, Petignat P, Boulvain M, Campana A. Use of automated primary screening on liquid-based, thin-layer preparations. *Acta Cytol* 2002;**46**:291–5.
36. Hologic. *Scottish ThinPrep Imager Feasibility Study*. Insight: Hologic Diagnostic News, Crawley, UK, 2009 (Spring 2009):12–13.
37. Scottish Cervical Cytology Review Group Feasibility Sub Group. *Cervical Cytology ThinPrep Imager (TIS) Feasibility Study - Report from the Feasibility Sub Group to Cervical Cytology Review Group*. Edinburgh: Scottish Pathology Network; 2009.
38. Walboomers J, Jacobs M, Manos M, Bosch F, Kummer J, Shah K, *et al.* Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol* 1999;**189**:12–19.
39. Schiffman M, Clifford G, Buonaguro FM. Classification of weakly carcinogenic human papillomavirus types: addressing the limits of epidemiology at the borderline. *Infect Agent Cancer* 2009;**4**:8.
40. Clifford G, Smith J, Plummer M, Munoz N, Franceschi S. Human papillomavirus types in invasive cervical cancer worldwide: a meta-analysis. *Br J Cancer* 2003;**88**:63–7.
41. The Future II Study Group. Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. *N Engl J Med* 2007;**356**:1915–27.
42. Paavonen J, Naud P, Salmeron J, Wheeler C, Chow S.-N., Apter D, *et al.* Efficacy of human papillomavirus (HPV) - 16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): final analysis of a double-blind, randomised study in young women. *Lancet* 2009;**374**:301–14.
43. Schiffman M, Solomon D. Findings to date from the ASCUS-LSIL Triage Study (ALTS). *Arc Pathol Lab Med* 2003;**127**:946–9.
44. Cuzick J, Szarewski A, Cubie H, Hulman G, Kitchener H, Luesley D, *et al.* Management of women who test positive for high-risk types of human papillomavirus: the HART study. *Lancet* 2003;**362**:1871–6.
45. Wright J, Rader J, Davila R, Powell M, Mutch D, Gao F, *et al.* Human papillomavirus triage for young women with atypical squamous cells of undetermined significance. *Obstet Gynecol* 2006;**107**:822–9.
46. Moss S, Gray A, Legood R, Vessey M, Patnick J, Kitchener H. Effect of testing for human papillomavirus as a triage during screening for cervical cancer: observational before and after study. *BMJ* 2006;**332**:83–5.
47. Kitchener HC, Walker PG, Nelson L, Hadwin R, Patnick J, Anthony GB, *et al.* HPV testing as an adjunct to cytology in the follow up of women treated for cervical intraepithelial neoplasia. *BJOG* 2008;**115**:1001–7.
48. Bulkman N, Berkhof J, Rozendaal L, van Kemenade F, Boeke A, Bulk S, *et al.* Human papillomavirus DNA testing for the detection of cervical intraepithelial neoplasia grade 3 and cancer: 5 year follow-up of a randomised controlled implementation trial. *Lancet* 2007;**370**:1764–72.

49. Kitchener HC, Almonte M, Dowie R, Stoykova B, Sargent A, Roberts C, *et al.* ARTISTIC: A randomised trial of HPV testing in primary cervical screening. *Health Technol Assess* 2009;**13**(51).
50. Naucler P, Ryd W, Tornberg S, Strand A, Wadell G, Elfgren K, *et al.* Human papillomavirus and papanicolaou tests to screen for cervical cancer. *N Engl J Med* 2007;**357**:1589–97.
51. Sargent A, Bailey A, Wheeler P, Kitchener H, Corbitt G, Peto J. A comparison of the Digene Hybrid Capture 2 assay and the Roche Amplicor Human Papillomavirus (HPV) test for the detection of 'high-risk' HPV genotypes in DNA extracts from liquid-based cytology samples collected from women whose cytology was graded 'borderline'. Proceedings of the 21st International Papillomavirus Conference, 2004, Mexico City, Mexico, abstract no. 143.
52. Cubie HA, Cuschieri KS, Zmijewski FM, Moore C. Evaluation of the sensitivity and specificity of the Roche Amplicor HPV test, the Roche prototype line blot assay and Digene Hybrid Capture 2 test for the detection of HPV in archived cervical samples with borderline cytology. Proceedings of the 21st International Papillomavirus Conference, 2004, Mexico City, Mexico, abstract no. 120.
53. Cuschieri KS, Seagar AL, Moore C, Gilkison G, Kornegay J, Cubie HA. Development of an automated extraction procedure for detection of human Papillomavirus DNA in liquid based cytology samples. *J Virol Methods* 2003;**107**:7.
54. Sargent A, Bailey A, Turner A, Almonte M, Gilham C, Baysson H, *et al.* Optimal Cut-off for Positive Hybrid Capture 2 Test for the detection of Human Papillomavirus: data from the ARTISTIC trial. *J Clin Microbiol* 2010;**48**:554–8.
55. Evans DM, Hudson EA, Brown CL, Boddington MM, Hughes HE, Mackenzie EF, *et al.* Terminology in gynaecological cytopathology: report of the working party of the British Society for Clinical Cytology. *J Clin Pathol* 1986;**39**:933–44.
56. Solomon D, Davey D, Kurman R, Moriarty A, O'Connor D, Prey M, *et al.* The 2001 Bethesda System: terminology for reporting results of cervical cytology. *JAMA* 2002;**287**:2114–19.
57. Arbyn M, Bergeron C, Klinkhamer P, Martin-Hirsch P, Siebers AG, Bulten J. Liquid compared with conventional cervical cytology- a systematic review and meta-analysis. *Obstet Gynecol* 2008;**111**:167–77.
58. Curtis L. *Unit Costs of Health and Social Care*. 2007 URL: www.pssru.ac.uk/pdf/uc/uc2007/uc2007.pdf (accessed on 26 June 2009).
59. Department of Health. *Payment by results 2008/09*. 2008. URL: www.dh.gov.uk/en/Managingyourorganisation/Financeandplanning/NHSFinancialReforms/index.htm (accessed on 26 June 2009).
60. Martin-Hirsch P, Rash B, Martin A, Standaert B. Management of women with abnormal cervical cytology: treatment patterns and associated costs in England and Wales. *BJOG* 2007;**114**:408–15.
61. Legood R, Gray AJW, Moss S. Lifetime effects, costs and cost-effectiveness of testing for human papillomavirus to manage low grade cytological abnormalities: results of the NHS pilot studies. *BMJ* 2006;**332**:79–85.
62. Canfell K, Barnabas R, Patnick J, Beral V. The predicted effect of changes in cervical screening practice in the UK: results from a modelling study. *Br J Cancer* 2004;**91**:530–6.
63. Medical Services Advisory Committee. Human Papillomavirus triage test for women with possible or definite low-grade squamous intraepithelial lesions. Canberra: Medical Services Advisory Committee; 2009.

64. Medical Services Advisory Committee. *Automation assisted and liquid based cytology for cervical cancer screening*. Canberra: Medical Services Advisory Committee; 2009.
65. Office For National Statistics. *Interim life tables for England 2005–2007 (Females)*. 2009. URL: www.statistics.gov.uk/downloads/theme_population/Interim_Life/ILTEng0608Reg.xls#’2005–07!A1 (accessed on 14 December 2009).
66. Paraskevaïdis E, Arbyn M, Sotiriadis A, Diakomanolis E, Martin-Hirsch P, Koliopoulos G, *et al*. The role of HPV DNA testing in the follow-up period after treatment for CIN: a systematic review of the literature. *Cancer Treat Rev* 2004;**30**:205–11.
67. Myers ER, Green S, Lipkus I. Patient preferences for health states related to HPV infection: Visual analogue scales vs. time trade-off elicitation. Proceedings of the 21st International Papillomavirus Conference, 2004, Mexico City, Mexico, abstract no. 390.2.
68. Gold MR, Franks P, McCoy KI, Fryback DG. Toward consistency in cost-utility analyses: using national measures to create condition-specific values. *Med Care* 1998;**36**:778–92.
69. Stratton KR, Durch JS. Appendix 11: human papillomavirus. In Stratton KR, Durch JS, Lawrence RS, editors. *Vaccines for the 21st century: A tool for decision making*. Washington, DC: National Academies Press; 2000. pp. 213–22.
70. International Agency for Research on Cancer. *Cancer incidence in five continents vol. VIII*. 2002. URL: www.iarc.fr/en/publications/pdfs-online/epi/sp155/ci5v8-cover.pdf (accessed on 14 December 2009).
71. Office For National Statistics. *Cancer registrations in England, 2006*. 2008. URL: www.statistics.gov.uk/downloads/theme_health/2006cancerfirstrelease.xls (accessed on 14 December 2009).
72. Office For National Statistics. *Mortality statistics: Cause - review of the Registrar General on deaths by cause, sex and age in England and Wales 2005*. London: Office for National Statistics; 2006.
73. Kitchener HC, Almonte M, Wheeler P, Desai M, Gilham C, Bailey A, *et al*. HPV testing in routine cervical screening: cross sectional data from the ARTISTIC trial. *Br J Cancer* 2006;**95**:56–61.
74. Briggs AH, Goeree R, Blackhouse G, O’Brien BJ. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. *Med Decis Making* 2002;**22**:290–308.
75. British Society for Clinical Cytology. *Recommended code of practice for laboratories participating in the UK cervical cancer screening programmes*. London: British Society for Clinical Cytology; 2009.
76. HM Treasury. *GDP Deflator Figures*. 2007. URL: www.hm-treasury.gov.uk/economic_data_and_tools/gdp_deflators/data_gdp_index.cfm (accessed on 14 December 2009).
77. Curtis L, Netten A. *Unit Costs of Health and Social Care 2006*. Canterbury: Personal Social Services Research Unit, University of Kent at Canterbury; 2006.
78. NHS Employers. Pay Circular (A for C) 1/2006. *Pay and Conditions for NHS staff covered by the Agenda for Change agreement*. London: NHS Employers; 2006.
79. National Institute for Health and Clinical Excellence. *The guidelines manual 2009*. London: National Institute for Health and Clinical Excellence; 2009.
80. Insinga RP, Glass AG, Myers ER, Rush BB. Abnormal outcomes following cervical cancer screening: event duration and health utility loss. *Med Decis Making* 2007;**27**:414–22.

81. Mandelblatt JS, Lawrence WF, Womack SM, Jacobson D, Yi B, Hwang YT, *et al.* Benefits and costs of using HPV testing to screen for cervical cancer. *JAMA* 2002;**287**:2372–81.
82. Sanders GD, Taira AV. Cost effectiveness of a potential vaccine for human papillomavirus. *Emerg Infect Dis* 2003;**9**:37–48.
83. Goldie SJ, Kohli M, Grima D, Weinstein MC, Wright TC, Bosch FX, *et al.* Projected clinical benefits and cost-effectiveness of a human papillomavirus 16/18 vaccine. *J Natl Cancer Inst* 2004;**96**:604–15.
84. Kim JJ, Wright TC, Goldie SJ. Cost-effectiveness of alternate triage strategies for atypical squamous cells of undetermined significance. *JAMA* 2002;**287**:2382–90.
85. Robertson JH, Woodend BE, Crozier EH, Hutchinson J. Risk of cervical smear associated with mild dyskaryosis. *BMJ* 1988;**297**:18–21.
86. Whynes DK, Woolley C, Philips Z. Management of low-grade cervical abnormalities detected at screening: which method do women prefer? *Cytopathology* 2008;**19**:355–62.
87. Philips Z, Avis M, Whynes DK. Introducing HPV triage into the English cervical cancer screening program: consequences for participation. *Women Health* 2006;**43**:17–34.
88. Poljak M, Fujs K, Seme K, Kocjan BJ, Vrtacnik-Bokal E. Retrospective and prospective evaluation of the Amplicor HPV test for detection of 13 high-risk human papillomavirus genotypes on 862 clinical samples. *Acta Dermatovenerol Alp Panonica Adriat* 2005;**14**:147–52.
89. Hardie A, Moore C, Patnick J, Cuschieri K, Graham C, Beadling C, *et al.* High-risk HPV detection in specimens collected in SurePath preservative fluid: comparison of ambient and refrigerated storage. *Cytopathology* 2009;**20**:235–41.
90. Bouvard V, Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, *et al.* A review of human carcinogens – Part B: biological agents. *Lancet Oncol* 2009;**10**:321–2.
91. Poljak M, Marin IJ, Seme K, Vince A. Hybrid Capture II HPV Test detects at least 15 human papillomavirus genotypes not included in its current high-risk probe cocktail. *J Clin Virol* 2002;**25**(Suppl. 3):S89–97.
92. Halfon P, Trepo E, Antoniotti G, Bernot C, Cart-Lamy P, Khiri H, *et al.* Prospective evaluation of the Hybrid Capture 2 and AMPLICOR human papillomavirus (HPV) tests for detection of 13 high-risk HPV genotypes in atypical squamous cells of uncertain significance. *J Clin Microbiol* 2007;**45**:313–16.
93. Castle PE, Solomon D, Wheeler CM, Gravitt PE, Wacholder S, Schiffman M. Human papillomavirus genotype specificity of hybrid capture 2. *J Clin Microbiol* 2008;**46**:2595–604.
94. Poljak M, Kocjan BJ, Kovanda A, Lunar MM, Lepej SZ, Planinic A, *et al.* Human papillomavirus genotype specificity of hybrid capture 2 low-risk probe cocktail. *J Clin Microbiol* 2009;**47**:2611–15.

Appendix 1

Time-and-motion survey questionnaire for loading and unloading of automated sample

Appendix 2

Time-and-motion survey questionnaires

Manual samples

MAVARIC TIME AND MOTION STUDY Technology (SP) (TP) please tick

How long have you been reading automated slides (months)

Instructions: This slide is to be included in the MAVARIC time and motion study. Please record your staff grade for the relevant activity and the time spent reading the slide.

1/ Times should be recorded using a stop watch in minutes and seconds

e.g. Ten minutes and 5 seconds should be written 10:05

2/ Times should NOT include any administration time including completing this form, looking at history, putting slide on stage or updating records on the computer. The timing starts the instant you start to screen the slide and ends the instant you finish screening the slide

Staff grade C = checker, CS = cytoscreener, or BMS

Slide no	1 st manual screen time	Staff grade	Rapid review manual time	Staff grade	Admin time 1 full screen (looking into history)	Admin time 2 full screen (updating records)	Admin 1 time for rapid screen (looking into history)	Admin 2 time for rapid screen (updating records)

Appendix 3

Primary screener worksheet

Primary Screeners Weekly Worksheet

Day	Technology	Routine full screen		Routine rapid screen		MAVARIC full screen		MAVARIC rapid screen		Other duties		Time for breaks
		Time	No	Time	No	Time	No	Time	No	Time	Describe	
Mon	TP											
	SP											
Tues.	TP											
	SP											
Wed	TP											
	SP											
Thurs.	TP											
	SP											
Friday	TP											
	SP											

Screener's name..... Week beginning.....

Please complete daily and return to XXXX at the end of the work week. These data will also be used as part of the MAVARIC trial to compare the overall workload between conventional and automated. * Please be specific e.g. obtaining 30 histopathology records

Appendix 4

Staff satisfaction questionnaire

Staff satisfaction questionnaire – MAVARIC Trial

We would be extremely grateful if you could complete the questionnaire.

1. What is your staff grade?

2. How many **months** have you been using automated reading?

3. How many **years** have you been working as a cyto-screener?

TRAINING

4. Were you satisfied with the training for using automated reading?

Excellent	Very good	Good	Fair	Poor
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

(Please tick appropriate box)

5. Do you have any specific comments about how the training could be improved?

.....

.....

.....

MANUAL AND AUTOMATED SYSTEMS

6. Overall I prefer using the automated reading systems compared with only using manual reading:

Strongly agree	Agree	Neutral	Disagree	Strongly disagree

(Please tick appropriate box)

7. I prefer using the IMAGER™ compared with FOCAL POINT™ for **primary screening**:

Strongly agree	Agree	Neutral	Disagree	Strongly disagree

(Please tick appropriate box)

8. I prefer using the IMAGER™ compared with FOCAL POINT™ for **full manual review**:

Strongly agree	Agree	Neutral	Disagree	Strongly disagree

(Please tick appropriate box)

9. Overall I prefer using the IMAGER™ compared with FOCAL POINT™:

Strongly agree	Agree	Neutral	Disagree	Strongly disagree

(Please tick appropriate box)

10. I find is easier to concentrate using the automated system compared with manual reading:

Strongly agree	Agree	Neutral	Disagree	Strongly disagree

(Please tick appropriate box)

11. My work is more challenging using the automated reading system compared to manual reading:

Strongly agree	Agree	Neutral	Disagree	Strongly disagree

(Please tick appropriate box)

12. My work is more monotonous using the automated reading system compared to manual reading:

Strongly agree	Agree	Neutral	Disagree	Strongly disagree

13. Do you experience any physical discomfort using either the manual or automated system:

Yes	No

(Please tick appropriate box)

14. Please describe any physical discomfort (e.g. noise, strain, motion sickness) you experience and specify whether this is associated with a particular machine?

.....

.....

.....

Many thanks for completing this questionnaire - please return to Jean Mather

Appendix 5

Roche Amplicor human papillomavirus testing

Initially, the Amplicor HPV MWP test was used because local testing^{51,52} and published data⁸⁸ had suggested an apparently greater analytical sensitivity for HPV DNA screening. Before recruitment to MAVARIC commenced, a small study was undertaken between the Manchester cytology laboratory and the Specialist Virology Centre in Edinburgh to ensure a robust transport protocol and also to test the Amplicor HPV assay on BD SurePath LBC samples sent at room temperature. The manufacturer recommends that LBC samples collected in BD SurePath medium should be topped up with BD SurePath medium if required, stored at 2–8°C and tested with Amplicor HPV MWP within 2 weeks. These conditions could not be met and samples were transported and stored at room temperature. Three batches of samples were sent with 16 paired samples (original collection ‘pot’ and ‘processed tube’).

Concordance within pairs was 85%. There were three pairs where the ‘pot’ was HPV negative (β -globin positive) and the ‘tube’ was HPV positive. However, all three had low RLU indices (0.432; 1.2; 1.8) which would have been considered negative for clinical management. Two ‘pot’ samples were positive for HPV despite negative β -globin results. With ‘tube’ samples, the only β -globin-negative result was also negative for HPV despite positive HPV and β -globin results in the ‘pot’ sample. Although more HPV-positive results were obtained with ‘tube’ than ‘pot’ samples, the results were generally low, raising the question of potential carry-over of HPV during the initial cytological processing stage. The ‘pot’ results seemed more robust, including picking up HPV-positive results in the absence of detectable β -globin, suggesting HPV presence in non-cellular fluid and thus supporting the potential for carry-over without carry-over of cellular material. The HPV testing laboratory recommended using ‘pot’ samples only, provided adequate closure of the hole in the pot lid (created as part of the cytology processing) could be achieved by covering it with a water-resistant adhesive disc prior to transport to prevent both evaporation and spillage. Subsequently, a supply of lids was made available from the manufacturer and used to secure the sample pot for transport.

Processing of samples for human papillomavirus testing

Relevant LBC samples corresponding to a low-grade abnormality were collected, screened and collated for dispatch to Edinburgh in the Manchester Cytology Centre. Patient names were removed prior to sending. The identifier used for subsequent interaction between Manchester and Edinburgh was the sample number assigned by the Manchester laboratory.

Samples received at the Specialist Virology Centre, Edinburgh, were accorded an internal sample number for HPV testing. A MAVARIC trial sample identification worksheet and laboratory checklist were completed in the laboratory throughout the testing process.

Nucleic acids were extracted from a 1-ml aliquot using the QIAamp 96 DNA Swab Kit on a BioRobot 9604 platform (Qiagen) with a protocol validated in Edinburgh for use with ThinPrep LBC medium.⁵³ Where weekly sample numbers were small (< 22), nucleic acids were extracted manually using the Roche Diagnostics AmpliLute Liquid Media Extraction Kit.

Samples were amplified by PCR using primers from the L1 region of HPV according to manufacturer's instructions and including full kit controls. Amplicons (165 base pairs in length) were detected colorimetrically in MWP's following hybridisation to oligonucleotide probes for 13 high-risk HPV types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 68) and to cellular (β -globin) targets to control for adequate cellularity of sample. Amplicor HPV MWP is a qualitative in vitro test for the detection of HPV in clinical specimens and is CE marked for use on ThinPrep and BD SurePath LBC samples under defined conditions.

Test data were entered into the Microsoft ACCESS database and results returned to the Manchester Trial Centre electronically as a Microsoft EXCEL password-protected file after each batch run.

Results of Roche Amplicor human papillomavirus testing

In total, 676 LBC samples were tested with the Amplicor HPV MWP test between 11 May 2006 and 10 July 2007. These were sent from Manchester in 34 batches. The turnaround time for receipt/testing/reporting averaged 3.6 days, with a range of 2–8 days.

Using the Roche Amplicor HPV test, 310 ThinPrep LBC samples and 366 BD SurePath LBC samples were tested. The results of the testing are summarised in *Tables 83 and 84*.

TABLE 83 Results of Roche Amplicor HPV test for ThinPrep LBC samples

Cytology	HPV negative	HPV positive	HPV invalid	Not tested	Total
N/A	0	0	0	1	1
Inadequate	1	0	0	0	1
Negative	15	3	0	0	18
Mild	21 (18.9%)	90 (81.1%)	0	0	111
Borderline	69 (40.1%)	99 (57.6%)	4	0	172
Borderline query high grade	0	1	0	0	1
Moderate	0	5	0	0	5
Severe	0	1	0	0	1
Total	106	199 (64.4%)	4 (1.3%)	1	310

N/A, not applicable.

TABLE 84 Results of Roche Amplicor HPV test for BD SurePath™ LBC samples

Cytology	HPV negative	HPV positive	HPV Invalid	Not tested	Total
N/A	0	1	0	0	1
Inadequate	0	1	0	0	1
Negative	41	6	6	0	53
Mild	17 (16.3%)	75 (72.1%)	11	1	104
Borderline	75 (40.9%)	76 (41.5%)	29	3	183
Borderline query high grade	0	1	0	0	1
Moderate	3	14	0	0	17
Severe	0	5	0	0	5
Delete from file	1	0	0	0	1
Total	137	179 (49.4%)	46 (12.6%)	4	366

Discussion

The Roche Amplicor HPV test was initially selected for testing samples showing a low-grade cytological abnormality as a means of triaging women for colposcopy. The test was used throughout the first year of the trial, when recruitment was far lower than expected, and consequently only 676 samples were processed with the Amplicor test. Despite the low number of samples tested there was a marked variation in positivity rates between ThinPrep LBC samples and BD SurePath LBC samples. ThinPrep LBC samples were reported as HPV positive in 64.4% of cases compared with 49.4% of BD SurePath LBC samples. The figures also show a high invalid rate with BD SurePath samples (12.6%), which gave rise to concern about the compatibility of the Roche Amplicor test and the BD SurePath LBC medium. It was initially thought that this may be due to the BD SurePath LBC samples being stored and transported at ambient temperature, rather than being kept in a cold chain as recommended by the manufacturers. The logistics involved in keeping the BD SurePath LBC samples in a cold chain within a routine screening programme were impractical and a decision was made to switch to HC2 for triage as LBC samples could be stored at ambient temperature and tested within 4 weeks of being taken.

Recent data from the QuASAR (Quality Assurance SurePath Ambient v. Refrigeration) study showed a high concordance rate between HC2 and Amplicor with both ambient and refrigerated BD SurePath LBC samples (87.7% and 89.2% respectively).⁸⁹ The QuASAR study also showed that BD SurePath LBC could be tested with Roche Amplicor within 3 weeks of collection, after being stored at ambient temperature. In comparison, data from the ARTISTIC study (using ThinPrep LBC) showed that Amplicor has a higher sensitivity than HC2, yet does not provide any additional clinical benefit and may result in a significantly higher number of women being triaged to colposcopy.⁴⁹ The NHSCSP HPV Special Interest Group are currently assessing the clinical utility of various new HPV tests and results will provide further insight into the utility of newer HPV tests for triage within the national screening programme [e.g. Abbott RealTime High Risk HPV (*rtHPV*), see *Appendix 8*].

Appendix 6

Automated cytology training

The training was provided by representatives of both companies.

Hologic (ThinPrep Imaging System)

Stain validation

Prior to training the ThinPrep Imaging System stain had to be validated by two cytopathologists and the laboratory trial co-ordinator (see *Chapter 2, ThinPrep Imaging System stain validation process*).

Training

The training took place over 3.5 days and comprised:

1. Presentations and an introduction to the review scope.
2. Review scope training over 1.5 days which comprised three modules – two training modules (10 slides in each) followed by a test module (25 slides). The results of the review scope training are provided in *Table 85*. This session was attended by the laboratory trial co-ordinator, two cytopathologists, one chief BMS, one senior cytoscreener and seven cytoscreeners.
3. Training in the use of the ThinPrep Imaging System; this 1-day session covered guidance on loading and unloading the machine, maintenance and troubleshooting. This session was attended by the laboratory trial co-ordinator and seven MLAs.

TABLE 85 Results of the ThinPrep Imaging System review scope training

Screener	Module I (10 slides)	Module II (10 Slides)	Test set (25 slides)	Overcall	Undercall
A	10	10	23	1 × HG, 1 × BL	0
B	9	10	24	1 × BL	0
C	10	9	23	1 × BL, 1 × LG	0
D	10	9	24	0	1 × BL
E	9	9	20	2X BL, 1 × LG, 1 × HG	1 × BL
F	10	8	21	3 × BL	1 × BL
G	10	9	21	2 × BL, 1 × MOD	1 × BL
H	9	9	23	1 × BL, 1 × MOD	0
I	9	9	25	0	0
J	10	10	24	0	1 × BL
K	10	10	13	9 × BL, 2 × LG, 1 × HG	0
L	10	10	23	2 × LG	0

BL, borderline nuclear changes; HG, high-grade dyskaryosis; LG, mild dyskaryosis/koilocytes.

Hologic were satisfied with the training results, and positive feedback was given by those who had taken part in the training.

Becton Dickinson Diagnostics (Becton Dickinson FocalPoint Guided Screener Imaging System)

The training took place over five days and comprised six modules:

- Module 1 – Presentation.
- Module 2 – Practical training with the BD FocalPoint GS Review Station to familiarise staff with its functions. This session included 10 technical training slides.
- Module 3 – Open discussion and question session.
- Module 4 – Location verification session including techniques for screening with the 10 FOVs. This session included 10 training slides.
- Module 5 – Open discussion and question session.
- Module 6 – Diagnostic performance session (comprising 100 test slides).

The results of all the slides screened during the BD FocalPoint GS Imaging System training are given in *Table 86*.

Training on screening slides using the BD FocalPoint GS Imaging System was delivered to the laboratory trial co-ordinator, two cytopathologists, one chief BMS, one senior cytoscreener and seven cytoscreeners. The laboratory trial co-ordinator and seven MLAs were also given training on loading and unloading the machine plus maintenance and troubleshooting guidance.

BD Diagnostics decided that all participants should go through a further test set of 100 slides owing to the number of undercalls in the first training set. The second test set showed excellent correlation with no undercalls. BD Diagnostics was satisfied with these results and those who took part in the training gave positive feedback.

TABLE 86 Results of the BD FocalPoint GS Imaging System training

Screener	Test 1 (10 slides)	Test 2 (10 slides)	Test 3 (102 slides)	Undercalls ^a	Overcalls ^b
1	10	10	92	1	9
2	10	8	95	4	3
3	10	10	93	1	8
4	10	10	96	2	4
5	10	10	91	2	9
6	10	10	94	1	7
7	10	10	94	2	6
8	10	8	89	4	9
9	10	10	94	5	3
10	10	7	95	0	7

a An undercall was classed as any abnormality of borderline changes or greater called negative.

b An overcall was classed as any negative called borderline changes or greater.

Appendix 7

Proforma for the review of discordant pairs

Appendix 8

Human papillomavirus genotyping

A new CE-marked HPV test, the Abbott *rt*HPV test, which could be carried out on an automated platform (M2000) had been trialled in Manchester and Edinburgh with favourable results. It involves nucleic acid extraction using magnetic beads, followed by real-time PCR amplification of target. In collaboration with Abbott it was agreed to test all HPV HC2-positive ThinPrep LBC samples using the manufacturer's cut-off of RLU/CO of 1.0 (887 samples) with genotyping of discrepant samples using Roche HPV LINEAR ARRAY. In addition, 469 samples from LBC specimens that gave HC2-negative results (RLU/CO < 1.0) were identified. An inadequate number of BD SurePath LBC samples were available for similar testing, but as the Abbott *rt*HPV test is not validated for BD SurePath medium, testing would have been inappropriate.

Methods

ThinPrep LBC cervical samples testing positive with HC2 were tested using the *rt*HPV test. This is a qualitative in vitro test for the detection of DNA for 14 high-risk HPV genotypes: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 56, 66 and 68. A cellular control amplicon is generated using primers and probe targeting a human β -globin sequence. The Abbott m2000sp robot and a residual 600- μ l volume of sample were used in the DNA extraction protocol. The Abbott m2000rt instrument was used in sample amplification and detection of the extracted DNA.

Discrepancies between the Abbott and Digene assays' results were resolved using the Roche LINEAR ARRAY genotyping test, which detects 37 anogenital HPV types, 6, 11, 16, 18, 26, 31, 33, 35, 39, 40, 42, 45, 51, 52, 53, 54, 55, 56, 58, 59, 61, 62, 64, 66, 67, 68, 69, 70, 71, 72, 73, 81, 82, 83, 84, IS39 and CP6108, and includes a human β -globin gene control. The Roche LINEAR ARRAY test is CE marked and validated for use with ThinPrep LBC samples.

Results

Results were available from 1356 samples and showed good overall concordance (89%) between HC2 and *rt*HPV (*Table 87*). One hundred and fifty-five discordant samples were genotyped, including five which contained specific HPV types at a copy number below Abbott's cut-off. Analysis of discordant samples is shown in *Table 88*. Of those which had no high-risk HPV detected using the Roche LINEAR ARRAY test, there were 47 false-positives by HC2 compared with three by Abbott *rt*HPV. Of the HC2 false-positives, 45% (21/47) had a RLU/CO < 2 and would have been reported as negative using the MAVARIC reporting protocol. In addition, 47 HC2 positives were associated with HPV types not present in the HC2 probe cocktail (30 with HPV 53, 12 with HPV 67, 4 with HPV82 and 3 with HPV 29). There were more false-negatives detected with Abbott *rt*HPV (45 vs 32 according to manufacturer's cut-off for both tests). Five samples gave a low positive RLU index with HC2 (RLU/CO > 1, < 2) and a different five were detected at low level with *rt*HPV [i.e. gave cycle threshold (Ct) values beyond cut-off, but showed graphic evidence of specific amplification]. Of these five, two contained HPV 16, two contained HPV 18 (one with HPV 59 and 66) and one contained HPV 52 and 58.

TABLE 87 Assessment of 1356 LBC samples with low-grade abnormalities using both HC2 and Abbott *rtHPV* (using manufacturer's cut-off for both tests)

Abbott <i>rtHPV</i>	HC2 high-risk HPV		Total
	HPV positive	HPV negative	
HPV positive	767	31	798
HPV negative	120	438	558
Total	887	469	1356

TABLE 88 Resolution of Digene HC2 and Abbott *rtHPV* discrepant samples using the Roche LINEAR ARRAY test

HPV LINEAR ARRAY	HC2			<i>rtHPV</i>		Total samples
	RLU/CO > 2	RLU/CO > 1, < 2	RLU/CO < 1	Positive	Negative	
High-risk HPV only	31	1	12	21	23	44
High- and low-risk HPV detected	41	4	15	38	22	60
Low-risk HPV only	22	16	1	1	38	39
No HPV detected	4	5	3	2	10	12
Subtotal	98	26	31	62	93	
Total	155			155		155

Comparing the *rtHPV* results with the genotypes detected by the Roche LINEAR ARRAY test, HPV 16 was detected in 110 and HPV 18 in 30 mono infections. HPV 16 was also present in a further 95 samples as a dual or more than dual infection with HPV 18 in 44 samples as a dual or more than dual infection. Other HPV types were found in 500 as mono infections and in association with HPV 16 and/or 18 in a further 110 samples (*Table 89*).

Discussion

- The *rtHPV* assay was more specific than HC2 with only three false-positive results compared with 47 HC2 false-positives.
- HC2 was more sensitive than *rtHPV* and gave fewer false-negative results (32 vs 45). The *rtHPV* test produced five samples with negative readings using the manufacturer's cut-off, but showing evidence of HPV-specific late amplification suggesting a low copy number. Cut-offs in every biological assay are open to scrutiny and may affect clinical algorithms.
- More false-positive results with HC2 were associated with detection of HPV types not present in the probe cocktail. This included 47 samples containing Group 2B HPV types which are considered 'high risk or probably high risk'.^{39,90} Cross-reaction was also detected in HC2 with low-risk HPV types (38 samples). Cross-hybridisation with both high- and low-risk HPV types has been reported with HC2,⁹¹⁻⁹⁴ but poses a problem for clinical management especially where HC2 gives a false-positive result.
- Thirty-five per cent of high-risk HPV infections were associated with HPV 16 or HPV 18, either alone or in association with other types.
- Only 18% of infections were associated with HPV 16 or HPV 18 mono infections. Comparison with clinical data will be required to assess the utility of the new Abbott *rtHPV* assay.

TABLE 89 Distribution of HPV genotypes in 767 LBC samples with low-grade abnormalities using the Abbott *rt*HPV test

HPV types found	HPV 16 mono infection	HPV 18 mono infection	Other HPV types mono infection	Dual HPV 16 and 18	Dual HPV 16 + other(s)	Dual HPV 18 + other(s)	Triple HPV 16 and 18 + other(s)
Number of samples	110	30	500	7	83	32	5

TABLE 90 CIN lesions by HPV type^a for all women with low-grade ThinPrep cytology triaged with HC2

	CIN2		CIN3+		CIN1 ^b	
	HPV 16 and/or 18 ^c	Non-16/18	HPV 16 and/or 18 ^c	Non-16/18	HPV 16 and/or 18 ^c	Non-16/18
Paired arm	18	22	20	12	119	256
Manual arm	13	12	8	4	59	121
Total	31	34	28	16	178	377
PPV	59/[59+178] × 100 = 24.89%					
NPV	377/[377+50] × 100 = 88.29%					

a All typed lesions were HC2 positive at a cut-off of 2 RLU/CO.

b Includes biopsy proven CIN1– and women with negative colposcopy and no biopsy.

c These lesions may contain multiple types.

In order to determine the value of HPV typing as a means of achieving greater specificity for HPV triage, all ThinPrep borderline/mild samples that were HC2 positive were typed using the Abbot *rt*HPV typing assay. These data are shown in *Table 90*, which indicates the clinical outcome by typing results for both arms of the study. The data have been classified as HPV 16 and/or 18, and non-16/18. HPV 16/18 are together the most prevalent in high-grade CIN. Out of the 109 CIN2+ lesions, 50 were associated with non-16/18 types and 59 with 16/18. If detection of HPV 16/18 were used to triage, colposcopy referral would therefore have been one-third of that using HC2 to triage, but 46% of CIN2+ would have been undetected and would have to be sought subsequently by repeat cytology, which risks non-attendance and failure to detect. The numbers for all known CIN2+ outcomes are shown and the PPV for type 16/18 is 25% compared with 15% for HC2.

Appendix 9

National Screening Committee's criteria for appraising the viability, effectiveness and appropriateness of a screening programme

The criteria, which are set out below, are based on the classic criteria first promulgated in a WHO report in 1966, but take into account both the more rigorous standards of evidence required to improve effectiveness and the greater concern about the adverse effects of health care; regrettably some people who undergo screening will suffer adverse effects without receiving benefit from the programme.

These criteria have been prepared taking into account international work on the appraisal of screening programmes, particularly that in Canada and the USA. It is recognised that not all of the criteria and questions raised in the format will be applicable to every proposed programme, but the more that are answered will obviously assist the National Screening Committee to make better evidence-based decisions.

All of the following criteria should be met before screening for a condition is initiated:

The condition

1. The condition should be an important health problem.
2. The epidemiology and natural history of the condition, including development from latent to declared disease, should be adequately understood and there should be a detectable risk factor or disease marker and a latent period or early symptomatic stage.
3. All the cost-effective primary prevention interventions should have been implemented as far as practicable.

The test

4. There should be a simple, safe, precise and validated screening test.
5. The distribution of test values in the target population should be known, and a suitable cut-off level defined and agreed.
6. The test should be acceptable to the population.
7. There should be an agreed policy on the further diagnostic investigation of individuals with a positive test result and on the choices available to those individuals.

The treatment

8. There should be an effective treatment or intervention for patients identified through early detection, with evidence of early treatment leading to better outcomes than late treatment.
9. There should be agreed evidence-based policies covering which individuals should be offered treatment and the appropriate treatment to be offered.
10. Clinical management of the condition and patient outcomes should be optimised by all health-care providers prior to participation in a screening programme.

The screening programme

11. There must be evidence from high-quality randomised controlled trials that the screening programme is effective in reducing mortality or morbidity.
12. Where screening is aimed solely at providing information to allow the person being screened to make an 'informed choice' (e.g. Down's syndrome, cystic fibrosis carrier screening), there must be evidence from high-quality trials that the test accurately measures risk. The information that is provided about the test and its outcome must be of value and readily understood by the individual being screened.
13. There should be evidence that the complete screening programme (test, diagnostic procedures, treatment/intervention) is clinically, socially and ethically acceptable to health professionals and the public.
14. The benefit from the screening programme should outweigh the physical and psychological harm (caused by the test, diagnostic procedures and treatment).
15. The opportunity cost of the screening programme (including testing, diagnosis, treatment, administration, training and quality assurance) should be economically balanced in relation to expenditure on medical care as a whole (i.e. value for money).
16. There must be a plan for managing and monitoring the screening programme and an agreed set of quality assurance standards.
17. Adequate staffing and facilities for testing, diagnosis, treatment and programme management should be made available prior to the commencement of the screening programme.
18. All other options for managing the condition should have been considered (e.g. improving treatment, providing other services) to ensure that no more cost-effective intervention could be introduced or that current interventions increased within the resources available.
19. Evidence-based information, explaining the consequences of testing, investigation and treatment, should be made available to potential participants to assist them in making an informed choice.
20. Public pressure for widening the eligibility criteria for reducing the screening interval, and for increasing the sensitivity of the testing process, should be anticipated. Decisions about these parameters should be scientifically justifiable to the public.

References

- Department of Health. *Screening of pregnant women for hepatitis B and immunisation of babies at risk*. Department of Health, 1998. (Health Service Circular: HSC 1998/127).
- Wilson JMG, Jungner G. *Principles and practice of screening for disease*. Public Health Paper Number 34. Geneva: WHO; 1968.
- Cochrane AL, Holland WW. Validation of screening procedures. *Br Med Bull* 1971;**27**:3.
- Sackett DL, Holland WW. Controversy in the detection of disease. *Lancet* 1975;**2**:357–9.
- Wald NJ, editor. *Antenatal and neonatal screening*. Oxford University Press, 1984.
- Holland WW, Stewart S. *Screening in healthcare*. The Nuffield Provincial Hospitals Trust; 1990.
- Gray JAM. *Dimensions and definitions of screening*. Milton Keynes: NHS Executive Anglia and Oxford, Research and Development Directorate; 1996.

Appendix 10

Tables restricted to routine samples from women aged 25–64 years

TABLE 91 Comparison of MRs

Arm	Inadequate	Negative	Borderline	Mild	Moderate	Severe	Q Inv	Q Glan	Total
Manual	534	17,486	465	356	87	103	3	7	19,041
	2.80%	91.83%	2.44%	1.87%	0.46%	0.54%	0.02%	0.04%	100%
Paired	1205	35,067	1136	703	144	235	10	22	38,522
	3.13%	91.03%	2.95%	1.82%	0.37%	0.61%	0.03%	0.06%	100%

Q Glan, query glandular neoplasia; Q Inv, query invasive.

TABLE 92 Comparison of MR1 results

Arm	Inadequate	Negative	Borderline	Mild	Moderate	Severe	Q Inv	Q Glan	Total
Manual	488	17,204	871	299	86	88	2	3	19,041
	2.56%	90.35%	4.57%	1.57%	0.45%	0.46%	0.01%	0.02%	100%
Paired	1045	34,779	1735	578	164	197	9	15	38,522
	2.71%	90.28%	4.50%	1.50%	0.43%	0.51%	0.02%	0.04%	100%

Q Glan, query glandular neoplasia; Q Inv, query invasive.

TABLE 93 Comparison of MR2 results

Arm	Inadequate	Negative	Borderline	Mild	Moderate	Severe	Q Inv	Q Glan	Total
Manual	538	17,076	932	311	87	92	2	3	19,041
	2.83%	89.68%	4.89%	1.63%	0.46%	0.48%	0.01%	0.02%	100%
Paired	1157	34,505	1869	597	203	166	9	16	38,522
	3.00%	89.57%	4.85%	1.55%	0.53%	0.43%	0.02%	0.04%	100%

Q Glan, query glandular neoplasia; Q Inv, query invasive.

TABLE 94 Comparison of FMR results

Arm	Inadequate	Negative	Borderline	Mild	Moderate	Severe	Q Inv	Q Glan	Total
Manual	534	17,486	465	356	87	103	3	7	19,041
	2.80%	91.83%	2.44%	1.87%	0.46%	0.54%	0.02%	0.04%	100%
Paired	1153	35,323	964	679	140	231	10	22	38,522
	2.99%	91.70%	2.50%	1.76%	0.36%	0.60%	0.03%	0.06%	100%

Q Glan, query glandular neoplasia; Q Inv, query invasive.

Appendix 11

Staff satisfaction survey results

Q4. Were you satisfied with the training for using automated screening?

	Responses
Excellent	0
Very good	3
Good	1
Fair	1
Poor	0

Q6. Overall I prefer using the automated reading systems compared with only using manual reading

	Responses
Strongly agree	0
Agree	0
Neutral	1
Disagree	4
Strongly disagree	0

Q7. I prefer using Imager compared with Focal point for primary screening

	Responses
Strongly agree	0
Agree	0
Neutral	1
Disagree	3
Strongly disagree	1

Q8. I prefer using Imager compared with Focal point for full manual review

	Responses
Strongly agree	0
Agree	0
Neutral	0
Disagree	0
Strongly disagree	5

Q9. Overall I prefer using Imager compared with focal point

	Responses
Strongly agree	0
Agree	0
Neutral	1
Disagree	2
Strongly disagree	2

Q10. I find it easier to concentrate using the automated system compared to manual reading

	Responses
Strongly agree	0
Agree	0
Neutral	1
Disagree	2
Strongly disagree	2

Q11. My work is more challenging using the automated reading system compared to manual reading

	Responses
Strongly agree	0
Agree	2
Neutral	0
Disagree	1
Strongly disagree	2

Q12. My work is more monotonous using the automated reading system compared to manual reading

	Responses
Strongly agree	5
Agree	0
Neutral	0
Disagree	0
Strongly disagree	0

Q13. Do you experience any physical discomfort using either the manual or automated system

	Responses
Yes	5
No	0

Q14. Please describe any physical discomfort (e.g. noise, strain, motion sickness) you experience and specify whether this is associated with a particular machine

Respondent	IMAGER	FOCAL POINT
1	<p>Back pain (between the shoulder blades) probably due to less ergonomic microscope. Difficult to adjust due to several operators using same microscope. Feel it is too noisy also.</p>	<p>Work station cramped, could do with more room, especially when using the microscope all day.</p>
2	<p>The imager is noisy when changing objectives to a higher magnification. The imager microscope is heavy and in one fixed position and cannot be adjusted to each individual screeners needs. Doing a full screen on the imager is very uncomfortable due to the continuous movement of the slide, resulting in a felling of motion sickness. This is probably the worst aspect of the imager.</p> <p>There is an element of monotony and repetition in both systems, which could lead to fatigue and loss of concentration</p>	
3	<p>Imager system - microscope not ergonomic, very heavy to adjust motion sickness when doing a full manual screen. Very noisy and eyes feel very tired when on this machine for a long period of time.</p> <p>Both machines have their good and bad points - I believe that there is an updated imager machine on trial at present therefore the above option would change</p>	<p>No problems to date.</p>
4	<p>The imager is very noisy when moving objectives etc -it is not ergonomic and doing a manual screen causes motion sickness.</p>	

	Both system cause the work to be more monotonous which lead to tiredness and lack of concentration	
5	Motion sickness on imager when doing scan, muscle strain and discomfort on the imager	

Appendix 12

Results of model fitting and additional parameters used in sensitivity analyses

The model of natural history and screening in England predicts an age-standardised incidence of 7.74 per 100,000 women (all ages, standard European population), and an age-standardised mortality of 2.20 per 100,000 women (all ages, standard European population). The results of the model fitting are shown in *Figures 18 and 19* and *Table 95*. It also predicts an age-specific prevalence of high-risk HPV which is consistent with that seen in ARTISTIC (*Figure 20*).¹

Parameters examined during modelling sensitivity analysis

The test characteristics of automated LBC were varied during sensitivity analysis to simulate (i) the worst performance consistent with MAVARIC data relative to manual LBC (lowest relative sensitivity and specificity) and (ii) the best performance consistent with MAVARIC data relative to manual LBC (highest relative sensitivity and specificity). Targets for relative performance were based on *Tables 35 and 36*. When it was not possible to meet the targets owing to competing constraints, we made assumptions that were favourable to automated LBC.

Reference

1. Kitchener HC, Almonte M, Wheeler P, Desai M, Gilham C, Bailey A, *et al.*, HPV testing in routine cervical screening: cross sectional data from the ARTISTIC trial. *Br J Cancer* 2006;**95**:56–61.
2. West Midlands Cancer Intelligence Unit. Invasive cervical cancer relative survival by stage in the West Midlands: tumours diagnosed 1995–7 followed up to the end of 2002. Birmingham: WMCIU; 2006.

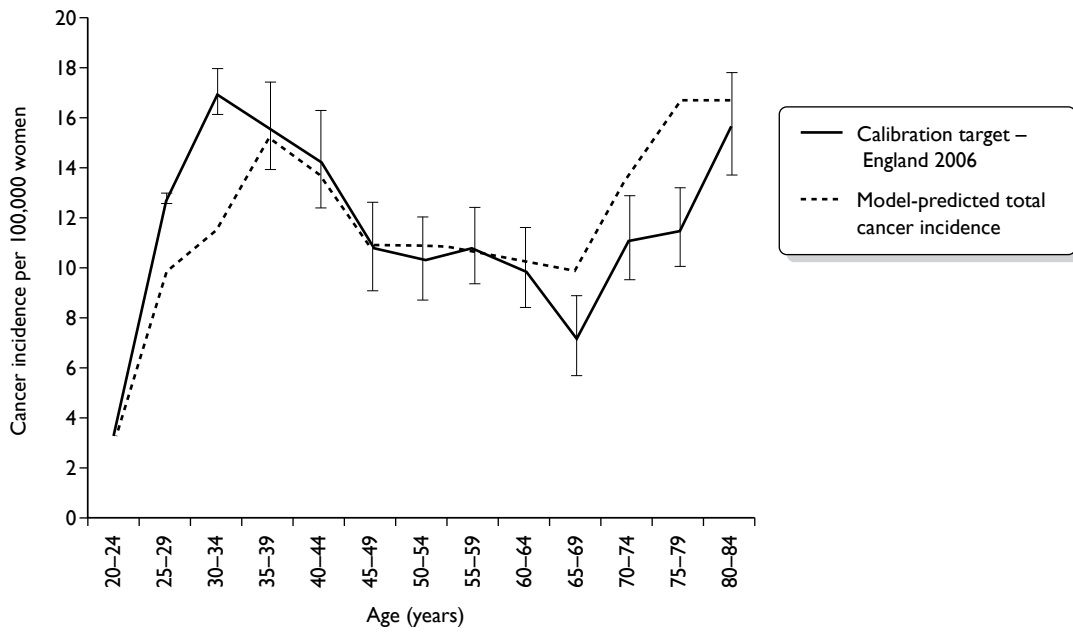


FIGURE 18 Predicted age-specific cervical cancer incidence compared with registry data (England, 2006).

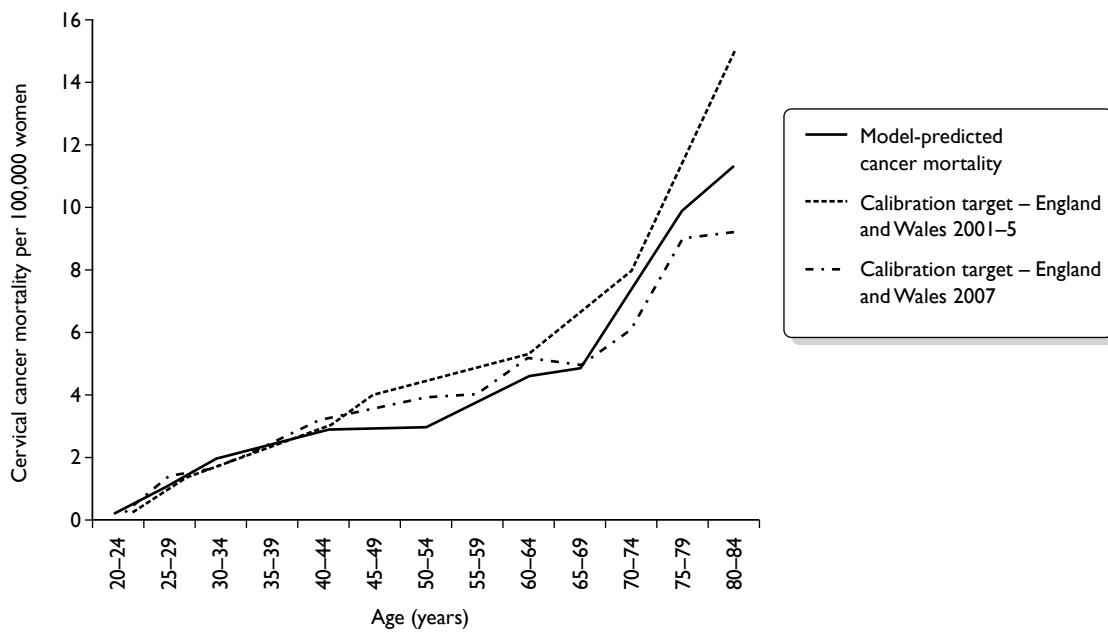


FIGURE 19 Predicted age-specific cervical cancer mortality compared with data for England and Wales (2001-5, 2007).

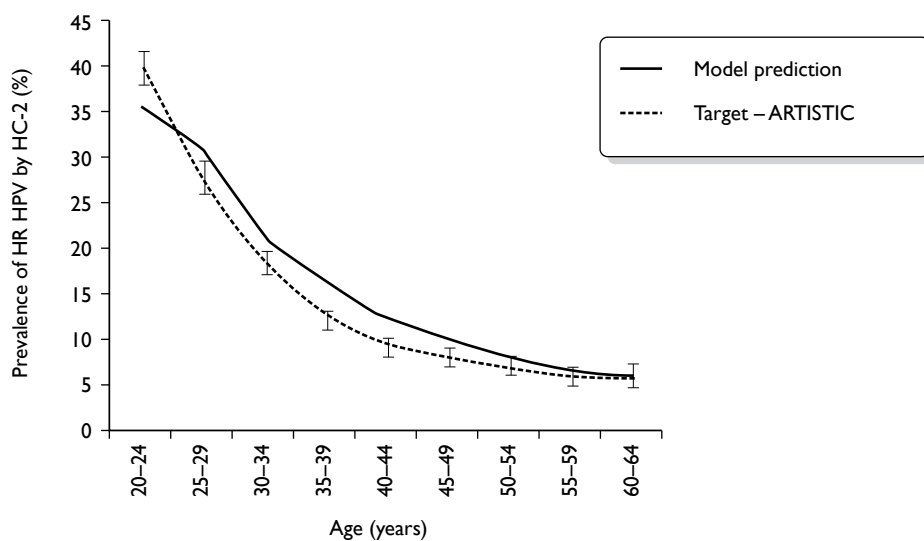


FIGURE 20 Predicted prevalence of high-risk HPV (detected by HC2) in a screening population compared with ARTISTIC data.¹

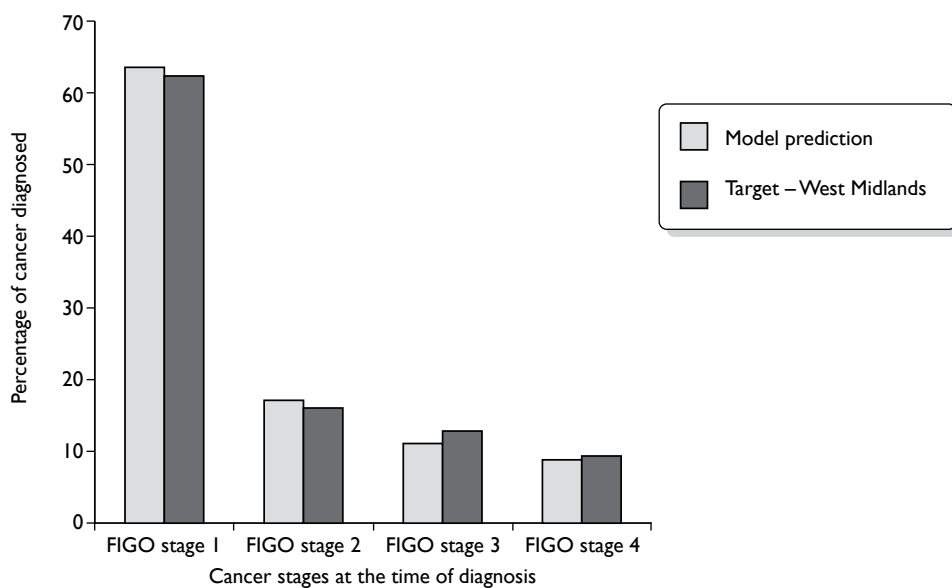


FIGURE 21 Cancer stage at time of diagnosis – model prediction vs data from West Midlands (2006).² FIGO, International Federation of Gynecology and Obstetrics.

TABLE 95 Model prediction versus actual data for cervical cancer incidence and mortality.

Output ^a	Model prediction	Target
Cancer incidence per 100,000 women – England (all ages)	7.74	8.1 (7.0–9.3) (average England 2004–6)
Cancer incidence per 100,000 women – (ages 25–64 years)	11.73	12.8 (11.2–14.6) (average England 2004–6)
Cancer cases – England (age ≤ 84 years)	2,199	2221 (actual cases England, 2006)
Cancer cases – England (ages 25–64 years)	1,612	1745 (actual cases England, 2006)
Cancer mortality per 100,000 women – England and Wales (all ages)	2.20	2.75–3.15 [England and Wales 2001–5, (average) 2007]
Cancer mortality per 100,000 women – England and Wales (ages 25–64 years)	1.75	1.96–2.08 (England and Wales 2001–5, (average) 2007)
Cancer deaths (all ages)	734	798 (actual death England and Wales, 2005)
Cancer deaths (ages 25–64 years)	404	427 (actual death England and Wales, 2005)

a Rates are standardised to the European population, unless otherwise specified.

TABLE 96 Summary of parameters used in the cost-effectiveness model

Item	Baseline	Minimum	Maximum
Management variables			
Yearly discount rate costs ^a	3.5%	0%	6%
Yearly discount rate effects ^a	3.5%	0%	6%
Attendance			
Routine smear (within 5 years) ^b			
Age <20 years	0.1%	Perfect compliance (0%)	
Age 20–24 years	18.3%		
Age 25–49 years ^b	80.3%–99.2%	Perfect compliance (100% every 3 years)	
Age 50–64 years ^b	84.9%–89.8%	Perfect compliance (100% every 5 years)	
Age 65–84 years ^b	8.4%–56.1%	Perfect compliance (0%)	
Repeat smear in 6 months	85%		100%
Repeat smear in 12 months	83%		100%
Colposcopy	84%		100%
Proportion of histological CIN1 referred for immediate treatment	7%	0%	
Proportion of women never screened ^c	2.2 %	0%	
Test characteristics			
Cytology	See below		
Cytology inadequate rate			
Manual LBC	2.99% ^d	2.6%	2.98%
Automated LBC	1.91% ^d	1.70%	1.94%
HC2		All values lowest positivity rates	All values highest positivity rates

Item	Baseline	Minimum	Maximum	
Costs – 2009 prices				
Cytology (laboratory cost)				
Manual LBC	£5.69	£5.35	£6.05	
Automated LBC	£5.455	£5.36	£5.56	
HPV reflex test	£16.85			
Histology outcome ^a				
No CIN ^b	£282.76	–	–	
CIN1	£432.29	–	–	
CIN2	£590.28	–	–	
CIN3	£625.37	–	–	
Cancer	<i>Stage I</i>	£2874.02	–	
	<i>Stage II</i>	£4590.17	–	
	<i>Stage III</i>	£12,963.53	–	
	<i>Stage IV</i>	£13,185.40	–	
Utilities				
False-positive ^d	0.96	0.95	0.97	
CIN1	0.89	0.85	1	
CIN2	0.88	0.87	1	
CIN3	0.89	0.83	1	
Cancer	<i>Stage I</i>	0.76	0.49	0.81
	<i>Stage II</i>	0.67	0.42	0.7
	<i>Stage III</i>	0.56	0.42	0.7
	<i>Stage IV</i>	0.48	0.36	0.6

a Three per cent after 30 years.

b Age specific.

c Prior to age 85 years.

d Based on *Table 73*.

e Applies to negative histology and cases where histology was not available (negative colposcopy, inadequate colposcopy without diagnostic excisional biopsy).

f Defined as cytology moderate or worse, or mild/borderline cytology with positive HPV triage test, where subsequent colposcopy/histology does not confirm CIN.

g Assumed to include all relevant colposcopy, biopsy and treatment costs, and is applied on the basis of histologically confirmed outcome – applies to negative histology and cases where histology is not available (negative colposcopy, inadequate colposcopy without diagnostic excisional biopsy).

TABLE 97 Accuracy of automated LBC relative to manual used for sensitivity analysis

Scenario being modelled	Relative sensitivity		Relative specificity	
	CIN2+	CIN3+	CIN2+	CIN3+
Baseline	0.924 (0.92)	0.956 (0.95)	1.007 (1.006)	1.008 (1.007)
Best performance assumption for automated LBC (target)	0.947 (0.95)	0.971 (0.99)	1.007 (1.007)	1.008 (1.008)
Worst performance assumption for automated LBC (target)	0.887 (0.89)	0.908 (0.91)	1.005 (1.005)	1.006 (1.006)

TABLE 98 Assumed test characteristics of HC2 test, based on international data

Model health state	Gold standard used	HC2 positivity rate			
		HPV triage (from borderline)		HPV triage (from mild)	
		Baseline (%)	Range ^a (%)	Baseline (%)	Range ^a (%)
Normal	PCR negative, normal cytology	1.4	1.4–4.2	1.4	1.4–4.2
HPV (no CIN)	PCR positive, normal cytology	92.5	49.7–92.5	92.5	49.7–92.5
CIN1	Histology (or cytology if no histology)	92.5	69.4–98.9	92.5	69.4–98.9
CIN2	Histology	92.5	90.1–94.9	97.2	95.6–98.9
CIN3	Histology	95.6	92.8–98.4	97.0	93.9–100.0

a Ranges given are derived from the literature⁶⁴ and used in sensitivity analysis.

Appendix 13

Additional tables relating to the comparison of results between manual and automated readings in the paired arm

TABLE 99 First manual result versus FMR

FMR	MR1		HPV positive	HPV negative	HPV not known	Borderline/mild	
	Inadequate	Negative				Moderate+	Total
Inadequate	1179	130			57		1366
Negative	46	42,520	18	33	1021	9	43,647
Borderline/mild							
HPV positive	4	59	1088			66	1217
HPV negative	3	61		606		14	684
HPV not known	1	41			657	41	740
Moderate+	4	15	13		143	442	617
Total	1237	42,826	1119	639	1878	572	48,271

Concordant results 46,492 (96.3%); discordant results 1779 (3.7%).

TABLE 100 First automated result versus FAR

FAR	AR1		HPV positive	HPV negative	HPV not known	Borderline/mild	
	Inadequate	Negative				Moderate+	Total
Inadequate	713	133			33		879
Negative	215	43,786	3	19	719	29	44,771
Borderline/mild							
HPV positive	8	65	849			103	1025
HPV negative	4	79		334		18	435
HPV not known	4	67			464	44	579
Moderate+	3	21	14		108	436	582
Total	947	44,151	866	353	1324	630	48,271

Concordant results 46,582 (96.5%); discordant results 1689 (3.5%).

Appendix 14

Comparison of results between manual-only and paired arm

TABLE 101 Comparison of MR1 and MR2 results in the manual and paired arm by reading and cytology results

Read	Cytology result (<i>n</i> and %)					Total
	Inadequate	Negative	Borderline/mild	Moderate	Severe+	
MR1 (manual)	584 (2.38%)	21680 (88.25%)	2400 (8.15%)	162 (0.66%)	140 (0.58%)	24566 (100%)
MR1 (paired)	1238 (2.56%)	42825 (88.72%)	3636 (7.53%)	276 (0.57%)	296 (0.62%)	48271 (100%)
MR2 (manual)	641 (2.61%)	21486 (87.46%)	2131 (8.68%)	164 (0.67%)	144 (0.59%)	24566 (100%)
MR2 (paired)	1378 (2.85%)	42415 (87.87%)	3892 (8.07%)	279 (0.58%)	307 (0.63%)	48271 (100%)

Appendix 15

Final trial protocol

A comparison of automated technology and manual cervical screening

Version 6 December 2007

Kitchener HC, Moss S, Cubie H, Desai M, Rana DN, Blanks R, Gray A, Legood R and Dunn G

Planned investigation

Background

Cervical screening is widely accepted as an effective and cost-effective means of reducing deaths from cervical cancer. Its inherent problems include limited sensitivity, maximised only by including the lowest grade of abnormality (borderline) for further investigation. This lowers the specificity of cervical screening and causes unnecessary anxiety to women and colposcopic workload. In addition, a reliance on manual reading is very time consuming and requires a very committed and large laboratory workforce. Previous laboratory failures have attracted widespread adverse publicity which has undermined the public image of cervical screening, and has also resulted in cytoscreeners feeling under fire. The importance of the cervical screening programme to the public and in particular to individual women cannot be underestimated. However, women need and expect the most accurate and reliable screening service, and the public the most cost-effective service. This project comparing automated and manual reading will determine the most efficient system.

The NHSCSP has become recognised as one of the world's leading cervical cancer prevention programmes. The basis for this is a quality-assured process with a population uptake in excess of 80%. This has seen a reduction in cancer incidence of 50% since 1988 and a corresponding fall in deaths. The remaining challenges include achieving higher sensitivity and even wider coverage in order to increase detection and at the same time achieving a sustainable service recognising the pressure on cytoscreeners. Harnessing technology to achieve these aims is a key strategy to improve the service. New technologies are not limited to developments in cytology but include complementary developments as seen in the field of HPV detection. The recent announcement that LBC is to be implemented highlights the commitment by the NHS to evidence-based strategies for improving screening. The availability of automated technology to facilitate cytological testing offers another opportunity to increase the efficiency and cost-efficiency of cervical screening. For automated technology to be perceived as a viable strategy it would need to be demonstrably superior to current manual reading, in terms of detection rates of abnormality, and/or practicability and cost-effectiveness. It is important to note that equivalence of automated and manual reading in terms of detection of high-grade CIN could still enable major advantages in terms of cost-effectiveness by greater efficiency.

Automated technologies that could be compared with manual screening

Desirable advances in cytology would alleviate some of these problems by improving sensitivity, specificity and reducing human workload. During the last 25 years gradual progress resulted in the emergence of two systems that automate the presentation of abnormalities on a cervical cytology slide. Both use location guiding which offers a means of standardising and thereby

quality assuring the scanning of slides, if not the actual interpretation of abnormal FOVs presented for review.

FocalPoint (TriPath). This is a location-guided system which can work on either conventional or LBC. In this study we would use the SurePath equipment designed for LBC. The location guiding works by identifying the 15 most abnormal locations on the slide designated from the most abnormal to least abnormal location. A computerised platform guides the slide so that the screener can visualise the FOVs. In addition to location guiding this technology can assign slides below a primary threshold which do not require human viewing, i.e. can be designated negative without the need for viewing and would only be backed up by rapid review. This has been approved by the FDA in the USA for a threshold representing the low 25%. In addition the slides requiring review can be ranked into quintiles (1–5) for likelihood of abnormalities. One machine can process up to 60,000 slides per year.

Imager (Cytyc). This system which has a similar capacity scans the ThinPrep slide and from a total of 120 FOVs selects the most abnormal 22 FOVs. These are then presented to the cytoscreener who can mark and interpret the abnormal cells. This system does not sort the slides as not requiring further review – it is purely location guided.

Previous research

Automated cytology

Much of the published research relates to the PapNet system, but this system was withdrawn before it became available. Other published studies evaluate the AutoPap 300 QC (supplied by TriPath and the precursor to the FocalPoint).¹

In 2001 an Italian study² was published which evaluated the AutoPap Primary Screening System, backed up by manual reading. Out of 14,779 consecutive conventional cervical smears, 10% were not processed because of technical defects. Of the remaining 13,261, 10,349 (78%) were selected for 'review' and 2912 (22%) as 'No Further Review'. Of the slides selected for review, 90% of abnormal smears were categorised by the device as in the first and second quintile rank while of those selected as 'No Further Review', 2905 were manually read as within normal limits, and the remaining seven as abnormal squamous cells of undetermined significance (ASCUS) or LSIL.

Recent data on FocalPoint have been presented by Cleary *et al.*³ from University College Hospital, Galway. They reported the impact of FocalPoint on lab reporting rates, based on 8632 slides pre-FocalPoint and 11,580 post-FocalPoint; all were conventional and not LBC. Unsatisfactory smears were reported at an increased rate by FocalPoint (7.8% vs 4.9%) but this may be resolved by LBC. There was a small but insignificant increase in the rates of all grades of abnormality.

There are as yet no peer-reviewed published data on the Imager system (Cytyc). FDA approval for the system was based on a four-centre trial sponsored by Cytyc. The outcomes are sensitivity and specificity of manual versus Imager read slides using LBC. These data were provided by Cytyc, and the key points were on Cytyc's website. The gold standard was not colposcopy/biopsy determined, but a 'truth adjudication' by two or three cytopathologists agreeing to a consensus cytological diagnosis. Specificity was defined as the percentage of 'true' classified slides by either system. 9550 slides were included, reflecting population screening. Seven per cent were rejected (Imager 'review') because of air bubbles, etc. There was a significantly increased 'sensitivity' to identify ASCUS, but not higher grades of abnormality. Specificity was broadly equivalent between Imager and manual for all grades of abnormality except for a very small increase in specificity for high-grade abnormalities. The conclusion was that the Imager system was safe and cost-effective.

The New Zealand HTA published a systematic review in 2000⁴ on effectiveness and cost-effectiveness of automated and semi-automated cervical screening devices. The section concerned with automated screening (part of the assessment was concerned with LBC) identified just one primary research study relevant to AutoPap. Verification was limited and did not permit direct estimates of test sensitivity and specificity. The assessment concluded that there was increased detection of low-grade lesions but not high-grade lesions. It also concluded that higher quality research was required to generate valid estimates of test sensitivity and specificity including methodology to address appropriate reference standards for verification of cytological diagnosis including test negatives. More robust health economic analysis is also required.

A recent systematic review commissioned by the HTA concluded that reliable conclusions about automated screening could not be drawn owing to the lack of sufficiently rigorous evaluations and trials. Further high quality primary research is required.

Our assessment of the published literature is that there is a need for a large publicly funded study which enables unbiased comparison of manual and automated cytology as well as head to head comparison of the two technologies which have emerged; location guiding with and without slide ranking.

HPV triage

In a recently published meta analysis of four pooled HPV triage studies,⁵ HC2 demonstrated a 16% increase in sensitivity compared with repeat cytology at a positive threshold of ASCUS (similar to borderline) cytology to detect CIN+.

In a follow-up paper on the original report of the ASCUS-LSIL Triage Study trial in the USA, also using HC2 for HPV testing, triage of ASCUS cytology by means of HPV testing to select those for colposcopy was at least as sensitive as colposcopying all subjects, and required only half the number of colposcopies. Although numbers are relatively small, both the HART (HPV in Addition to Routine Testing) study⁶ and the Kaiser Permanente Study⁷ demonstrated a very high NPV for HPV triage.

Key considerations in assessing diagnostic accuracy

1. New technologies should be compared against the existing method in terms of sensitivity, specificity and PPV. Both sensitivity (the proportion of subjects truly with the disease called positive by the screening test) and specificity (the proportion of subjects truly disease free called negative by the screening test) require a reference (gold) standard to determine the true-positives; for cervical screening the appropriate gold standard is colposcopy and biopsy, leading to histological diagnosis.
2. In practice, it is neither ethical or practical to colposcope women found negative on all screening tests performed, and we thus lack information on the reference standard in these women. Relative sensitivity (and specificity) of two methods can be compared for both paired and unpaired data.
3. We will maximise our estimate of sensitivity by including HPV triage of women with borderline or mild dyskaryosis. A 'positive' screening test will be one that leads to immediate referral to colposcopy (i.e. moderate dyskaryosis or worse OR borderline/mild dyskaryosis and HPV positive).
4. As a reference standard, CIN 2 or greater represents the threshold for treatment and will be used to determine true-positives. However, in terms of protection against invasive cancer and death from the disease, detection of CIN3 is a more valid outcome, and will also be used as a clinical outcome in the analysis.
5. Invasive cancer is too rare an outcome, even in a study of this size, to be informative. Flagging of subjects would therefore provide little benefit for this study. We will obtain

information on cytological and histological diagnosis at 3 years in those women attending for routine repeat smear.

In order to reflect real life, the project should be embedded within routine practice in the NHSCSP.

Key considerations in assessment of economic analysis and organisation impact

Automated equipment is expensive, but there may be productivity/workload savings if smear readers could: read slides faster, have fewer slides to review (NFR with FocalPoint) or refer fewer slides for 'checking'. These factors affect the overall cost, sensitivity and specificity of reading a smear.

Assessment of full cost-effectiveness requires assessment of life-years/QALYs. As it is not feasible to obtain data on incidence of cervical cancer, it is necessary to model how alternative screening technologies would affect the underlying incidence and disease progression and regression. Estimates from the clinical study of the true sensitivity and specificity, and information on costs and productivity, are required to inform the cost-effectiveness model.

Research objectives

- To determine the comparative diagnostic performance of automated and manual reading in terms of relative sensitivity, specificity and PPV.
- To determine how automated reading compares with manual reading when used in conjunction with HPV triage of low-grade abnormalities.
- To evaluate the ranking module of FocalPoint in terms of the NFR, i.e. whether a proportion can be reported negative without being read.
- To compare the two technologies, i.e. location-guided (Imager) versus location-guided and slide ranking (FocalPoint).
- To assess inadequate rates with both technologies.
- To evaluate productivity gains of automation in relation to laboratory throughput and reporting times.
- To determine by economic analysis the costs and long-term cost-effectiveness of the two systems in comparison with manual.
- To investigate cytoscreeners' experience and satisfaction with automated systems.
- To investigate the organisation changes that automation would require and achieve, whether beneficial or detrimental.

Study design

Overall considerations in the study design

This is a randomised trial of automated versus manually read liquid-based cervical cytology, involving assessment of both FocalPoint and Imager systems.

Randomisation to technology will be performed at general practice level as it is not feasible for both technologies to be used within one practice. Therefore there is cluster randomisation between automated technologies.

Samples received using each supplier's collection devices will be individually randomised between an arm with double reading by both manual and automated systems (paired comparison) or to a manual reading only arm. The primary statistical analysis will include the paired comparisons within the double reading arm. Such paired comparison has the advantage of providing greater

statistical power (by avoiding between subject variability). Because it is necessary to demonstrate equivalence of the new technique to manual reading before automated cytology can be used as the sole screening test, our statistical plan is designed to demonstrate equivalence in sensitivity and specificity between manual and automated and between the automated technologies.

We do not feel that the knowledge that a separate manual reading is being done will significantly affect the interpretation of the automated reading. This issue will be avoided for manual reading by having a separate manual reading only arm, so that the screener performing the manual reading is, as far as possible, 'blind' as to whether or not automated screening is also taking place.

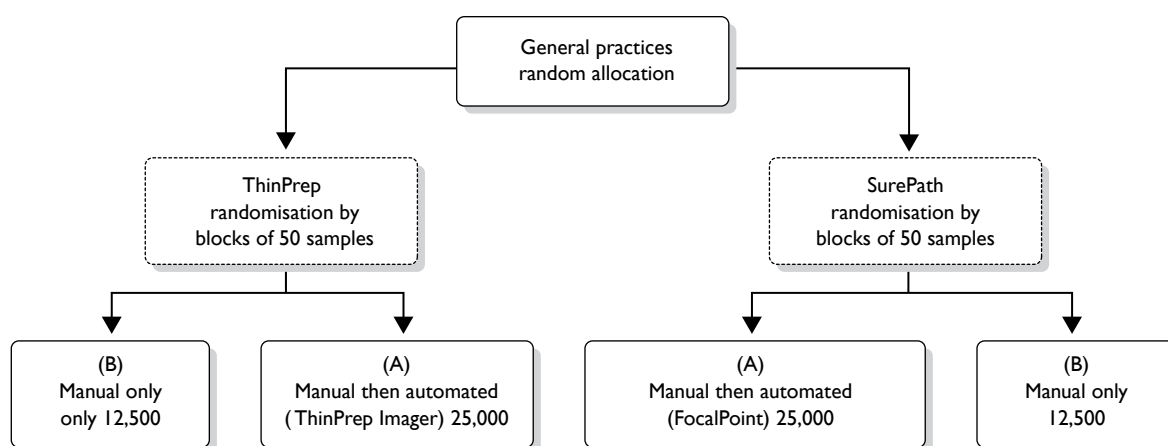
Because the two automated technologies use different fixative, each automated system will need to have a separate 'manual reading only' arm.

The primary comparison will be of each automated technology with manual reading, and equivalence would need to be demonstrated for each technology before either could be used alone. We will also undertake comparisons of each automated technology with manual reading in terms of cost and cost-effectiveness. HPV triage will be used for women in both arms with borderline and mild dyskaryosis. By using this for both grades of cytology we will minimise any verification bias that could result from differing rates of reporting either grade, between the manual and automated arms.

Study design

Screened population in Greater Manchester

General practices will be randomly allocated to use either SurePath or ThinPrep LBC kits. For each technology, on receipt at the laboratory, samples would be randomised to the double reading (automated and manual) or the manual-only arm.



In the double reading arms (A) management will be based on the manually read result, with the exception of a normal manual result and an abnormal automated result after checking. In this case borderline or mild are sent for HPV triage and moderate and severe dyskaryosis are sent to colposcopy.

Colposcopy will be performed for a single report showing moderate or severe dyskaryosis. If colposcopy is abnormal an appropriate biopsy and treatment will be performed.

For first borderline cytology or mild dyskaryosis only a reflex HPV test would be used to select women for colposcopy (as in recent NHS pilots). For subsequent borderline or mild dyskaryosis, repeat in 6 months. If HPV test is negative, return to routine recall.

Cytology taken as part of follow-up protocol following initial screen will be manually read.

The reason for including the two automated systems is that:

- (a) Both of the LBC systems (ThinPrep and SurePath) will be in place in the NHSCSP.
- (b) The slide ranking module of the FocalPoint™ is of potential importance because if indeed the least abnormal 25% slides can be filed without reading, there would be major efficiency saving.
- (c) Head to head comparison in the manual arm alone will be informative (as requested by NICE).

Primary outcomes

The primary outcome would be the relative sensitivity of screening by automated or manually read cytology to detect CIN3/invasive cancer (CIN3+) and CIN2, 3 and invasive cancer (CIN2+).

Other outcomes – clinical

1. The detection rates of CIN2+ and CIN3+ in each arm.
2. The detection rates (PPVs) for each category of cytology including the threshold of borderline or greater and mild dyskaryosis or greater.
3. Relative specificity rates of screening by automated and manual reading.
4. All of the above comparing FocalPoint and Imager.
5. The reliability of NFR in FocalPoint in terms of NPV using negative manual reading in the paired reading and the reference standard.
6. To assess inadequate rates with both technologies.

Other outcomes – economics and organisational

- Comparative throughput and reporting times (for each stage of screening).
- Detailed cost estimates of the total cost of processing smear at the laboratory and total cost per smear including consideration of inadequate rates and using NFR at different cut-off levels.
- Estimate of the comparative cost-effectiveness of automated versus manually read cytology using trial data and modelled lifetime costs and effects.
- Assessment of cytoscreeners' experience and satisfaction with automated systems and the organisational changes that automation would require in implementation.

Planned interventions

Cytology

On receipt of the LBC specimen at the Manchester Cytology Centre, for each technology random blocks of 50 will be allocated to either automated plus manual or manual reading (later to automated only or manual only). Details of exactly how slides will be handled are described in *Appendix 1*. The need for separate manual arms for SurePath and ThinPrep is based on their distinct liquid preservative medium and for the Imager system a distinct staining system. To compare the Imager automated reading with SurePath stained slides for manual reading would not be valid. The full conversion of the Manchester Cytology Centre will mean that these separate manual arms will be available anyway in terms of capacity; the only costs will be data inputting, and possibly the additional cost of ThinPrep if the PCTs purchase only SurePath LBC.

The rate of reading of slides allocated to the double reading arm(s) is constrained by the additional workload involved in double reading, which will be done largely in overtime. Manual reading will be done prior to being processed for FocalPoint in order to blind the cytoscreeners to whether or not the slide is also being read automatically. For the Imager system, as compared with routine ThinPrep manual read slide, a different stain is used which has undergone reformulation and is satisfactory for manual reading. The ThinPrep specimens randomised to manual reading will therefore, be distinguishable from manual reading in the double reading arm.

In the event of slides being rejected by the automated systems as either 'process review' or simply not read by the machine (up to 10%), a second slide will be prepared and the end result will be based on that result.

We did consider developing an additional slide from the liquid residue for back-up manual reading but this would be expensive and probably not as valid as paired readings on exactly the same slides.

Human papillomavirus testing

Primary research has indicated conclusively that HPV testing is capable of selecting women at increased risk of having underlying high-grade CIN from those who have a very low likelihood of having high-grade CIN.^{7,8} This triage by HPV testing can be used to increase the sensitivity of cytology by investigating women with low-grade abnormality while at the same time maintaining colposcopy investigation at a manageable level. The use of HPV triage in this study will achieve three objectives:

- It will enable a more sensitive determination of underlying disease than would routine NHSCSP guidelines. It will therefore enable a more accurate determination of the relative sensitivity of each cytology system.
- It will achieve a more rapid diagnosis of underlying disease than if the outcome of reported low grade were required based on repeat cytology for up to 12 months. This will allow the project to be completed in a shorter time scale and with less default.
- It will allow manual and automated cytology to be compared in conjunction with HPV triage, which may be incorporated into future NHSCSP protocol if the NHS pilot studies confirm its clinical utility.

Women who test cytology-negative manually, but mild on automated will be triaged, if indicated, after the discrepancy has been resolved by a medic.

Referral for colposcopy

Women with moderate and severe dyskaryosis will be referred for colposcopy as dictated by NHS guidance. In addition, women who have borderline or mild dyskaryosis who test HPV positive will also be referred for colposcopy. Those testing HPV negative will undergo surveillance according to current NHS guidance, and be referred for colposcopy if the abnormality persists.

Currently around 3% of screened women are referred for colposcopy on the basis of low-grade cytological abnormalities. Data from the ARTISTIC cohort in Greater Manchester indicate that 9.6% are either borderline or mild dyskaryosis in women aged between 25 and 64 years. Of these 36% are HPV positive (25% of borderline and 61% of mild dyskaryosis). This represents 3.45% of the screened population and could therefore be accommodated in local colposcopy clinics.

Avoidance of bias

Bias in the comparison of automated and manual reading will be avoided by randomisation in blocks of 50.

General practices will be allocated to use one or other of the LBC kits: ThinPrep or SurePath. This cluster randomisation will only affect comparisons of the two technologies. To avoid bias in terms of underlying risk of cytological abnormality, the practices will be randomised to either of the systems stratified by Townsend Deprivation Scores which is a measure at the PCT level. Areas will therefore be evenly balanced in the use of both technologies.

Inclusion/exclusion criteria

All women in the cervical screening age group will be eligible if they are attending for a routine cervical screening test or repeat test for mild abnormalities. Following the recent announcement from the Department of Health this age group will be 25–64 years. We will also include cytology samples from colposcopy clinics because these will have a higher proportion of abnormalities which will help to achieve a greater power. We will attempt to achieve a balance of ThinPrep and SurePath by allocation between colposcopy clinics.

Ethical considerations

The study has full ethical approval from the Central Manchester LREC. Women will receive an information leaflet with their call/recall letters from the PCT. In some PCTs where there are only a small number of GP practices participating, or if staff at the PCT find it difficult to disperse leaflets to practices, we will distribute the information leaflet to practices ourselves, so that women can collect it when they make appointments. Should a woman decline HPV testing, the smear takers have been asked to note this on the cervical cytology request form to inform the lab of the decision. A telephone hotline will be set up for women with concerns or queries.

Statistical analysis and sample size determination

Referring to *Table 1*, the letters D+/D–, M+/M– and A+/A– indicate the results of the colposcopy (CIN2+ or not CIN2+, for example), manual smear test procedure and automated smear test procedure, respectively. The outcome of colposcopy is taken to be the gold standard, but it is only available for those women who are smear positive (that is, a positive smear test using either method for the paired data, or smear positive using the manual method for the unpaired data; a so-called ‘screen positives design’).⁹ Smear test characteristics are estimated as illustrated in *Table 1*. Note that numbers enclosed brackets are those, which from the nature of the design, cannot be directly observed.

The paired data in each arm of the study will provide estimates of the ratio of the sensitivities (relative TPR) of the manual (M) and automated smear tests [ThinPrep – A1 or SurePath – A2 in the two arms, respectively – see *Table 1a*], but not their separate values. Similarly, the paired data will provide estimates of the relative false-positive rate (rFPR) for the two tests, where the FPR = 1 – specificity. The unpaired data can similarly be used for the comparison of M used alone with M used on the same sample as ThinPrep (A1) or SurePath (A2), in terms of both the relative TPR (rTPR) and rFPR. Detection rates and PPVs can also be estimated from both the paired and the unpaired data. The statistical methods for the construction of valid CIs for these characteristics (used to evaluate equivalence or non-inferiority of the two tests) are described in Pepe⁹ and in Alonzo *et al.*¹⁰ For the comparison of two test procedures (using the paired data) we wish to demonstrate equivalence for *both* TPR *and* FPR with a global significance test (with significance level $\alpha = 0.05$, say) and therefore use $\alpha^* = 1 - (1 - 0.05)^{0.5} \approx 0.025$ as the significance level for each characteristic separately. The clustering of participants introduced by the cluster

randomisation to ThinPrep or SurePath should have no effects on the paired comparisons. It is possible, however, that the clustering might increase the sampling variability of the estimates from the unpaired data and robust standard errors and associated CIs will be estimated to check for this.

In a final series of analyses, which will enable the investigators to make full use of the potential of all of the information from the complex design, data from both paired and unpaired smear tests will be jointly analysed for the comparison of the two automated tests (ThinPrep and SurePath), the comparison of each automated tests with the Manual procedure, with and without the assumption that the performance of the Manual smear procedure is the same in both arms and for both paired and unpaired smears. These analyses will involve the fitting of a series of latent class models, allowing for the complex pattern of missing data determined by the design (i.e. avoiding work-up biases) as described in Chapter 5 of Dunn.¹¹

We have based our sample size calculations on a proposed test of non-inferiority of the automated smear test in terms of its sensitivity (relative to that of the Manual method) based only on data from the paired observations. Inclusion of the unpaired data will increase statistical power, but we have chosen a conservative approach based solely on the paired comparisons. Sample sizes for the paired comparison are determined by the numbers of D+ participants needed to evaluate relative TPRs. When the number of D+s is about 630, a paired test with a 0.025 one-sided significance level will have 80% power to reject the null hypothesis that the sensitivities are not equivalent [the difference in sensitivities (TPRs) is 0.050 or farther from zero in the same direction] when the expected difference in proportions is 0, assuming that the proportion of discordant pairs is 0.200 (NQUERY ADVISOR, Version 3). The sample size estimation is sensitive to the assumed value for the proportion of discordant pairs. We think that 0.2 is likely to be the upper limit. The power would increase to about 95% if the proportion of discordant pairs were actually 0.1. In the latter case the study would have about 70% power to exclude a difference in the TPRs of 0.03 or farther from zero in the same direction. If the proportion of women who are D+ in the population is about 3% we need to obtain a total of about 23,000 participants in each of the two arms to have a probability of 0.975 that it contains at least 630 D+s. We have chosen a conservative estimate of 25,000 smears in each arm for the paired comparison, and an equal number of unpaired smears (hence a total of $4 \times 25,000 = 100,000$ smears in the trial overall).

Numbers within square brackets [] are missing. The TPR (TPR = sensitivity) of test M is $(a + c)/(a + b + c + [d])$; that for A1 is $(a + b)/(a + b + c + [d])$. These cannot be determined, but their ratio (rTPR) is estimated by $(a + c)/(a + b)$. Similarly, the FPR (FPR = $1 - \text{specificity}$) for M is $(e + f)/(e + f + g + [h])$, for A1 is $(e + g)/(e + f + g + [h])$ and for their ratio, the rFPR, is $(e + f)/(e + g)$. The PPV of M is $(a + b)/(a + b + e + f)$ and for A1 is $(a + c)/(a + c + e + g)$. The detection rate for A1 is estimated by $(a + c)/N$, where $N = [n_{D+}] + [n_{D-}]$ is the total number of paired smears randomised to this arm (i.e. N will be about 25,000).

TABLE 1a Paired data from the ThinPrep (A1) arm

D+	D-		D-	
	A1+	A1-	A1+	A1-
M+	a	b	M+	e
M-	c	[d]	M-	g
$[n_{D+}] = a + b + c + [d]$			$[n_{D-}] = e + f + g + [h]$	

TABLE 1b Unpaired data from the ThinPrep (A1) arm

	D+	D-
A1+	A	C
A1-	[B] [N _{D+}]=A+[B]	[D] [N _{D-}]=C+[D]

Numbers within square brackets [] are missing. Note that randomisation implies approximate equality of $A + [B]$ and the corresponding count in the SurePath arm, and also of $C + [D]$ and the corresponding count in the SurePath arm. The TPR for ThinPrep is $A/[B]$, and the corresponding FPR is $C/[D]$. The corresponding parameters for SurePath are defined similarly. None of these can be estimated directly but their ratio (rTPR and rFPR respectively) can (because randomisation ensures that, on average, the missing denominators are equal in the ThinPrep and SurePath arms). The PPV of A1 is estimated by $A/(A + C)$ and its detection rate by A/N , where $N = [N_{D+}] + [N_{D-}]$ which is again about 25,000.

Health economic assessment

Economic analysis and organisational impact assessment

An economic analysis will be conducted alongside this trial, with the objectives of:

1. Assessing the productivity implications and organisational impact of automated screening.
2. Estimating the incremental costs, effects and cost-effectiveness of the two automated screening technologies being evaluated, in comparison with manually read cytology.

In conducting this analysis we will be able to draw on the methods, questionnaire designs and modelling procedures used when we undertook the evaluation for the Department of Health¹² of the national screening programme's pilot sites using LBC and HPV triage.

Productivity and organisational impact

A detailed assessment will be made of the productivity implications and broader organisational impact of automated screening throughout the trial. Prospective survey instruments, observations and questionnaires will be employed. The design of these instruments will be piloted, but we will adapt the methods and questionnaire designs used in our LBC/HPV pilot sites evaluation.¹² Cytoscreeners will be interviewed 1 year into the study and at the conclusion of the study.

Productivity of laboratory staff, including both smear readers and laboratory assistants operating the automated equipment, will be measured in the implementation period and throughout the trial. This will permit study of whether productivity improvements can be realised in practice through changes in actual numbers of staff required.

The broader organisational impacts of automated screening will also be assessed. The training requirements and logistical implications will be fully documented. Data on staff acceptability of the automated screening will be collected through questionnaires. Quality assurance will be closely monitored at the laboratories and guidance developed to assist other laboratories in the event of a national roll-out of the technology.

Costs and cost-effectiveness, future outcomes modelling

Costs per smear

The economic evaluation will pay particular attention to estimating the incremental costs of the technologies: including the capital equipment and consumable costs, staff costs, and the effects of any changes in laboratory productivity and throughput. Transition costs such as the costs of staff training, logistical and organisational change will be recorded.

Method: A bottom-up costing method will be used as this has been found to give more reliable estimates than a top-down approach.¹³ We will use the same combination of questionnaires, surveys, observations and interviews to estimate these as we employed in our evaluation of the LBC/HPV pilot sites.¹² The collection of costing data will be fully integrated with the assessment of productivity and organisational impact. This will allow the development of detailed costings, encompassing assessment of factors such as the impact of different cut-off values for NFR and whether changes in staffing costs can be realised financially.

Analysis: The total laboratory cost to screen one woman's sample will be estimated by combining data in a cost model. These data include the average time for preparation, primary screen, rapid review and checking slides, consumables, equipment and overhead costs. As well as estimating the average time and resource use for each stage of the laboratory process, the range and distribution of uncertainty in each component of cost will be assessed. Total average cost estimates will combine data on both the average costs and the uncertainty around total costs. The cost estimates will be used in the cost-effectiveness model.

Cost-effectiveness assessment

Methods: Assessment of long-term outcomes and cost-effectiveness requires assessment of life-years gained/QALYs. Modelling is required as the trial data do not collect data on cancer incidence and mortality. We believe the most appropriate and validated way of modelling long-term cost-effectiveness in this study will be to use an adapted UK version of the Myers (US Agency for Health Care Policy and Research) Markov model.¹⁴ This model was developed for the US Agency for Health Care Policy and Research (US Department of Health and Human Resources) to help evaluate national screening programmes, and is well validated. It has clear advantages over other existing models in that it permits modelling of long-term health outcomes of cytological abnormality and HPV detection, and has previously been applied successfully by us to information from the Department of Health LBC/HPV pilot evaluation (final report). The model incorporates simulation of the natural history of disease including HPV status, CIN and invasive cancer states (I–IV) and will incorporate UK data such as invasive cancer 5-year survival data.

The main model parameters that will be obtained from this study will be:

- accurate estimates of the cost of processing smears
- relative sensitivity (by smear grade) and specificity
- cost-consequences of smear results, in particular colposcopy referral rates.

The study will provide not only baseline estimates, but also information on the range and distribution of uncertainty in these estimates. Trial estimates of relative sensitivity and specificity cannot be used directly in the model because the model requires estimates of true sensitivity and specificity given underlying disease. It will be necessary to adjust for verification bias when estimating true sensitivity and specificity estimates. The statistician and health economists will draw on further data from the literature, where women have been followed up with negative manual cytology results to adjust the relative estimates obtained in the trial to predict the true sensitivity and specificity.

For other parameters (including the effectiveness of colposcopy, natural history, invasive cancer treatment costs and primary care costs and utilities) the literature will be searched to ensure that we are using the most up to date and valid estimates.

Analysis: This model already reflects current UK screening policy including comprehensive modelling of the management of different types of cytology results as well as the management of women with negative smears and the current age range of 25–64 years). The model will be adapted to permit comparison between automated screening systems and other screening options including using cytology alone (LBC or conventional) and HPV testing. The cost and cost-effectiveness analysis will also simulate optimal cut-off values for abnormalities in the automated procedures. Results will be presented within a probabilistic framework, using cost-effectiveness acceptability curves and net-benefit statistics.

Consumer input

We have consulted Dr Pat Wilke, an experienced lay advisor to NHS bodies, currently serving with the Royal College of Pathologists. She approves of the project and has contributed some comments. She has accepted our invitation for her to join the Trial Management Group.

Milestones

Months 1–6	Set up study Detail which practices will be involved Train practice nurses where required Develop database and data collection system Organise HPV collection Get cytology lab staff trained and equipment installed
Months 7–36	Trial
Months 37–48	Complete follow-up and analyse data Prepare final report, publications, etc.

This time scale fits well with the time required to implement LBC across England, in the sense that it will be 4–5 years before LBC is completely rolled out, and the system ready for further change.

Justification of costs

This project needs to be on a large scale in order to demonstrate in a convincing manner whether or not automated cytology should be introduced to the NHSCSP. The potential productivity gains of both slide sorting and location guiding are such that a major investment in the primary research will be justifiable. In order to offset the costs to a degree, our academic institutions have agreed to a reduced overhead of 30%.

The project is not inherently complex but its scale and practical issues will require adequate manpower.

Research costs

Staffing

Project manager: This individual will provide direct day-to-day supervision of the project including contact with primary care, the cytology laboratory, consumable suppliers and

equipment manufacturers. He or she will oversee the data collection and ensure adequate backing up of data.

Project secretary: The project manager will require a secretary to deal with data inputting, obtaining results for colposcopy, follow-up cytology and histopathology. He or she will be required to take telephone calls and provide hour-to-hour commitment to the project.

Statistician: The database for the project will be held at the CSEU from the outset, to permit necessary management/data validation to take place. A junior statistician/data manager will be required for the duration of the project to design and manage the database, liaise with the trial centre in Manchester and with economic researchers to ensure appropriate data collection, and perform all analyses. Supervision by Dr Moss and Dr Blanks, together with statistical advice from Professor G Dunn, will be provided at no additional cost.

Health economist: A health economist is requested on scale RS2 (D32.05) at 0.6 whole time equivalent (WTE) over the duration of the study to prepare a detailed economic analysis plan, prepare and check data collection instruments for resource use and outcomes, collect unit cost information, measure productivity and organisational impact by field work and other methods, attend meetings and liaise with investigators, sponsors and collaborators, prepare progress reports and any interim analyses of health economics data, conduct data modelling and simulation of long-term results, prepare manuscript(s), prepare presentations, attend relevant conferences, and deal with all queries concerning economic analyses and results.

Biomedical scientist – virology: A BMS2/MT04 is required to analyse up to 10,000 samples in each phase for HPV testing during months 7–36. This includes receipt/logging of specimens, DNA extraction and amplification, running the tests and sending data to the trial centre.

Biomedical scientist – cytology: A BMS3/MT05 is required to supervise the automated machinery to manually check all of the doubly read cytology results in Phase 1, both manual and automated, in order to authorise and sign off the final cytological reports. He or she would also provide additional manpower for reading the cytology given the necessity for double reading 25,000 slides in Phase 1.

Medical laboratory assistant – cytology: Daily duties to be performed by the MLA for the HTA trial will include extra sorting and filing of slides into trays, cleaning and removal of slides prior to loading and unloading the automated machines. There will be extra remounting of slides and restaining of rejected slides. The machines will need to be maintained. An extra staining machine will be provided for the automated ThinPrep samples, which will also need to be operated and maintained. Vials will need to be retrieved from the archive and packed for transport to Edinburgh.

References

1. Lee JSJ, Kuan L, Seho Oh, Patten FW, Wilbur DC. A feasibility study of the AutoPap system location-guided screening. *Acta Cytol* 1998;**42**:221–5.
2. Alasio LM, Alphandery C, Grassi P, Ruggeri M, De Palo G, Pilotti S. Performance of the AutoPap primary screening system in the detection of high-risk cases in cervicovaginal smears. *Acta Cytol* 2001;**45**:704–8.
3. Cleary J, Rabbitt L, Kenny B, Bennani F, Fitzpatrick B. Impact of Focal Point™ slide profiler on reporting of results. IACC meeting, 16–17 May 2003, Sligo, Ireland.

4. Broadstock M. Effectiveness and cost effectiveness of automated and semi-automated cervical screening devices: A systematic review. *N Z Health Technol Assess* 2000;**3**(1).
5. Arbyn M, Buntinx F, Van Ranst M, Paraskevardis E, Martin-Hirsch P, Dillner J. Virologic vs cytologic triage of women with equivocal pap smears: a meta analysis of the accuracy to detect high grade intraepithelial neoplasia. *J Natl Cancer Inst* 2004;**96**:280–93.
6. Cuzick J, Szarewski A, Cubie H, Hulman G, Kitchener H, Luesley D, *et al.* Management of women who test positive for high-risk types of human papillomavirus: the HART Study. *Lancet* 2003;**362**:1871–6.
7. Manos MM, Kinney WK, Hurley LB, Sherman ME, Shieh-Ngai J, Kurman RJ, *et al.* Identifying women with cervical neoplasia: using HPV DNA testing for equivocal Papanicolaou results. *JAMA* 1999;**281**:1605–10.
8. Solomon D, Schiffman M, Tarone R, for the ALTS Group. Comparison of three management strategies for patients with atypical squamous cells of undetermined significance: baseline results from a randomised trial. *J Natl Cancer Inst* 2001;**93**:293–9.
9. Pepe MS. *The statistical evaluation of medical tests for classification and prediction*. Oxford: Oxford University Press; 2002.
10. Alonzo TA, Pepe MS, Moskowitz CS. Sample size calculations for comparative studies of medical tests for detecting presence of disease. *Stat Med* 2002;**21**:835–52.
11. Dunn G. *Statistical evaluation of measurement errors*. 2nd edn. London: Arnold; 2004.
12. Moss S, Gray A, Legood R, Henstock E. *Evaluation of HPV/LBC cervical screening pilot studies. First report to the Department of Health on evaluation of LBC*. Sutton: Cancer Screening Evaluation Unit, Institute of Cancer Research; 2003.
13. Helms LJ, Melnikow J. Determining costs of health care services for cost effectiveness analysis. The case of cervical cancer prevention and treatment. *Med Care* 1999;**7**:652–61.
14. Myers ER, McCrory DC, Nanda K, Bastian L, Matchar DB. Mathematical model for the natural history of human papillomavirus infection and cervical carcinogenesis. *Am J Epidemiol* 2000;**151**:1158–69.

Appendix 1: Protocol for the management of cytology samples

All ThinPrep (TP) and SurePath (SP) samples will have their specimen type entered at request entry as per current office protocols.

A query will be set up by the laboratory manager from which an electronic list will be produced of all the TP and SP cervical samples from women between the ages of 25 and 64 years.

The statistician will provide the laboratory with a randomisation list for both TP and SP with numbers from 1 to 25,000, this will include whether the sample will be read automatically and manually or manually only.

The electronic list will be added to the randomisation list and in sets of no more than 20 will be added to a reclassification list. The reclassification lists will contain the sample number, which arm of the trial the sample is in and, in the case of those randomised to automated reading the screener, the rapid screener and the results.

Private patients will be excluded from the trial.

All the departments' SP cervical samples will go through the FocalPoint location-guided screening machine to facilitate a print run being performed more easily (120 slides have to be run on the FocalPoint for a print run to be produced).

All the trial TP samples will go through the Cytoc Imager location-guided machine.

Once the slides have been imaged by both systems they will be passed to the laboratory co-ordinator/BMS 3 and they will organise the slides in the automated arm into slide trays with the request form and slide sheet and pass them to the screeners trained to read them.

After the primary screen on the location-guided screening microscopes (Slide Wizard for SP and Review Scope for TP) the slides, forms and sheets will be passed by the laboratory co-ordinator to another screener for rapid review. From the SP system up to 25% can be classified as NFR, these will just have a rapid screen on the automated arm of the trial.

No ink marks will be made on the slides while reading them on the location-guided microscopes, but electronic marks can be added.

After the automated read and rapid rescreen, the slides and request forms will be placed back in their original slide trays and placed on the shelf in the screening room in numerical order to be manually read, the manual reader will be screening the slides without knowing the outcome of the automated read.

After the trial slides have been manually read and rapid rescreened they will be passed to the laboratory co-ordinator who will add the manual or manual and automated result to the request notes on the laboratory computer system. If the result is negative or inadequate the laboratory co-ordinator will authorise these following the laboratory reporting protocols to generate a printed report.

Any results that are abnormal will be passed to a medic/AP to report. Any sample showing a borderline/mild dyskaryosis result will be sent to Edinburgh for HPV testing and the result of these samples will not be sent until the HPV result is known.

The MLA will pick out the HPV samples and pack them ready for transporting to Edinburgh.

The HPV samples will be sent via Citysprint on a Monday provided there are at least 15 samples.

Any discrepancies between the manual and automated readings will be passed to a medic/AP to report.

The samples showing borderline/mild dyskaryosis will be reported as per the MAVARIC trial protocol.

The laboratory co-ordinator will be responsible for the automated machines and the flow of the trial samples through the laboratory.

Appendix 2: Human papillomavirus testing protocols

Logistics

LBC samples will be collected in Manchester.

LBC samples will be transported to Edinburgh, weekly in batches using appropriate approved packaging by designated courier. The LBC samples will be the specimen volume remaining in the 'tubes' following cytological slide processing of specimens collected in SurePath Preservative Fluid using the TriPath Imaging Prestain Slide Processor and the 'original vial' containing the remaining specimen volume of PreservCyt® Solution after ThinPrep Pap Test Slides are prepared according to Cytoc protocol.

Samples will be logged in a secure database with unique identifiers for each sample.

HPV screening will be carried out using Digene Hybrid Capture High-Risk HPV DNA test™. This involves an in vitro nucleic acid hybridisation assay with signal amplification using microplate chemiluminescence for the qualitative detection of HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 68 in cervical specimens.

Results will be returned to Manchester after each batch run.

Human papillomavirus testing rationale

Hybrid Capture has been selected for several reasons:

- First commercially available HPV test, which is both CE marked and FDA approved.
- No nucleic acid extraction procedures are required.
- Although LBC samples need to be prepared prior to the hybridisation stage of the assay, there are validated Digene protocols to follow for both specimens in PreservCyt Solution and SurePath Preservative Fluid.
- PreservCyt Solution specimens may be held for up to 3 months at temperatures between 2 and 30 °C following collection and prior to processing for the HC2 high-risk HPV DNA test. After cytological analysis, SurePath specimens may be stored for up to 4 weeks at 2–30 °C prior to processing for the HC2 high-risk HPV DNA test.
- The HC2 high-risk HPV DNA test can be performed manually or using the Rapid Capture System Instrument for high-volume, sample throughput testing. Although not available in Edinburgh, the Rapid Capture System is a general-use automated pipetting and dilution system, handling up to 352 specimens in 8 hours including a 3.5-hour period during which user intervention is not required.
- 'Invalid' HPV interpretation, possible with Roche Molecular Systems Amplicor MWP test does not occur with the Digene HC2 high-risk HPV DNA test, as there is no internal housekeeping gene control to determine if the cellular content is adequate.
- Using only trained and validated laboratory personnel and following validated protocols, the risk of either false-positive or false-negative results should be minimised.

Appendix 16

The standards for the reporting of diagnostic accuracy studies checklist

Section and topic	Item number		On page number
TITLE/ABSTRACT/KEYWORDS	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity')	iv
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups	13, 14
METHODS			
<i>Participants</i>	3	The study population: The inclusion and exclusion criteria, setting and locations where data were collected	15, 16
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	15
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in items 3 and 4? If not, specify how participants were further selected	14, 15
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	22–25
<i>Test methods</i>	7	The reference standard and its rationale	21–23
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard	17–21
	9	Definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard	18, 21–23
	10	The number, training and expertise of the persons executing and reading the index tests and the reference standard	18–21
	11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers	20, 21
<i>Statistical methods</i>	12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals)	27
	13	Methods for calculating test reproducibility, if done	N/A
RESULTS			
<i>Participants</i>	14	When study was performed, including beginning and end dates of recruitment	39
	15	Clinical and demographic characteristics of the study population (at least information on age, gender, spectrum of presenting symptoms)	40–44
	16	The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended)	42
<i>Test results</i>	17	Time interval between the index tests and the reference standard, and any treatment administered in between	25
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition	N/A
	19	A cross-tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard	60–62
	20	Any adverse events from performing the index tests or the reference standard	N/A

Section and topic	Item number		On page number
<i>Estimates</i>	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals)	55–58
	22	How indeterminate results, missing data and outliers of the index tests were handled	25,43–45
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centres, if done	N/A
	24	Estimates of test reproducibility, if done	N/A
DISCUSSION	25	Discuss the clinical applicability of the study findings	89–91, 94, 95

N/A, not applicable.

Appendix 17

The consolidated standards of reporting trials 2010 checklist of information to include when reporting a randomised trial

Section/Topic	Item number	Checklist item	Reported on page number
Title and abstract			
	1a	Identification as a randomised trial in the title	i
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	iii, iv
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	3, 4, 13
	2b	Specific objectives or hypotheses	13, 14
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	14, 15, 25, 26
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	18, 19
Participants	4a	Eligibility criteria for participants	15
	4b	Settings and locations where the data were collected	15, 16
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	18–22
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	13, 14, 37
	6b	Any changes to trial outcomes after the trial commenced, with reasons	N/A
Sample size	7a	How sample size was determined	25, 26
	7b	When applicable, explanation of any interim analyses and stopping guidelines	N/A
Randomisation			
Sequence generation	8a	Method used to generate the random allocation sequence	14, 15
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	14, 15
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	14, 15
Implementation	10	Who generated the random allocation sequence, who enrolled participants and who assigned participants to interventions	15
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	20
	11b	If relevant, description of the similarity of interventions	N/A
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	26–28
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	28

Section/Topic	Item number	Checklist item	Reported on page number
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment and were analysed for the primary outcome	42
	13b	For each group, losses and exclusions after randomisation, together with reasons	42
Recruitment	14a	Dates defining the periods of recruitment and follow-up	39
	14b	Why the trial ended or was stopped	N/A
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	N/A
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	41, 42
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	55–58
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	55–58
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	57–86
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	N/A
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision and, if relevant, multiplicity of analyses	87, 88
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	88
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	89–91, 94, 95
Other information			
Registration	23	Registration number and name of trial registry	iv
Protocol	24	Where the full trial protocol can be accessed, if available	151
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	iv

N/A, not applicable.

Health Technology Assessment programme

Director,
Professor Tom Walley, CBE,
 Director, NIHR HTA programme, Professor of Clinical Pharmacology,
 University of Liverpool

Deputy Director,
Professor Hywel Williams,
 Professor of Dermato-Epidemiology,
 Centre of Evidence-Based Dermatology,
 University of Nottingham

Prioritisation Group

Members

<p>Chair, Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool</p> <p>Professor Imti Choonara, Professor in Child Health, Academic Division of Child Health, University of Nottingham Chair – Pharmaceuticals Panel</p> <p>Dr Bob Coates, Consultant Advisor – Disease Prevention Panel</p> <p>Dr Andrew Cook, Consultant Advisor – Intervention Procedures Panel</p> <p>Dr Peter Davidson, Director of NETSCC, Health Technology Assessment</p>	<p>Dr Nick Hicks, Consultant Adviser – Diagnostic Technologies and Screening Panel, Consultant Advisor – Psychological and Community Therapies Panel</p> <p>Ms Susan Hird, Consultant Advisor, External Devices and Physical Therapies Panel</p> <p>Professor Sallie Lamb, Director, Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick Chair – HTA Clinical Evaluation and Trials Board</p> <p>Professor Jonathan Michaels, Professor of Vascular Surgery, Sheffield Vascular Institute, University of Sheffield Chair – Interventional Procedures Panel</p>	<p>Professor Ruairidh Milne, Director – External Relations</p> <p>Dr John Pounsford, Consultant Physician, Directorate of Medical Services, North Bristol NHS Trust Chair – External Devices and Physical Therapies Panel</p> <p>Dr Vaughan Thomas, Consultant Advisor – Pharmaceuticals Panel, Clinical Lead – Clinical Evaluation Trials Prioritisation Group</p> <p>Professor Margaret Thorogood, Professor of Epidemiology, Health Sciences Research Institute, University of Warwick Chair – Disease Prevention Panel</p>	<p>Professor Lindsay Turnbull, Professor of Radiology, Centre for the MR Investigations, University of Hull Chair – Diagnostic Technologies and Screening Panel</p> <p>Professor Scott Weich, Professor of Psychiatry, Health Sciences Research Institute, University of Warwick Chair – Psychological and Community Therapies Panel</p> <p>Professor Hywel Williams, Director of Nottingham Clinical Trials Unit, Centre of Evidence-Based Dermatology, University of Nottingham Chair – HTA Commissioning Board Deputy HTA Programme Director</p>
---	---	--	--

HTA Commissioning Board

<p>Chair, Professor Hywel Williams, Professor of Dermato-Epidemiology, Centre of Evidence-Based Dermatology, University of Nottingham</p>	<p>Deputy Chair, Professor Andrew Farmer, Professor of General Practice, Department of Primary Health Care, University of Oxford Programme Director,</p>	<p>Professor Tom Walley, CBE, Professor of Clinical Pharmacology, Director, NIHR HTA programme, University of Liverpool</p>
---	---	---

Members

<p>Professor Ann Ashburn, Professor of Rehabilitation and Head of Research, Southampton General Hospital</p> <p>Professor Deborah Ashby, Professor of Medical Statistics and Clinical Trials, Queen Mary, Department of Epidemiology and Public Health, Imperial College London</p> <p>Professor Peter Brocklehurst, Director, National Perinatal Epidemiology Unit, University of Oxford</p> <p>Professor John Cairns, Professor of Health Economics, London School of Hygiene and Tropical Medicine</p>	<p>Professor Peter Croft, Director of Primary Care Sciences Research Centre, Keele University</p> <p>Professor Jenny Donovan, Professor of Social Medicine, University of Bristol</p> <p>Professor Jonathan Green, Professor and Acting Head of Department, Child and Adolescent Psychiatry, University of Manchester Medical School</p> <p>Professor John W Gregory, Professor in Paediatric Endocrinology, Department of Child Health, Wales School of Medicine, Cardiff University</p>	<p>Professor Steve Halligan, Professor of Gastrointestinal Radiology, University College Hospital, London</p> <p>Professor Freddie Hamdy, Professor of Urology, Head of Nuffield Department of Surgery, University of Oxford</p> <p>Professor Allan House, Professor of Liaison Psychiatry, University of Leeds</p> <p>Dr Martin J Landray, Reader in Epidemiology, Honorary Consultant Physician, Clinical Trial Service Unit, University of Oxford</p>	<p>Professor Stephen Morris, Professor of Health Economics, University College London, Research Department of Epidemiology and Public Health, University College London</p> <p>Professor E Andrea Nelson, Professor of Wound Healing and Director of Research, School of Healthcare, University of Leeds</p> <p>Professor John David Norris, Chair in Clinical Trials and Biostatistics, Robertson Centre for Biostatistics, University of Glasgow</p> <p>Dr Rafael Perera, Lecturer in Medical Statistics, Department of Primary Health Care, University of Oxford</p>
---	---	--	---

HTA Commissioning Board *(continued)*

Professor James Raftery,
Chair of NETSCC and Director of
the Wessex Institute, University of
Southampton

Professor Barney Reeves,
Professorial Research Fellow
in Health Services Research,
Department of Clinical Science,
University of Bristol

Professor Marion Walker,
Professor in Stroke Rehabilitation,
Associate Director UK Stroke
Research Network, University of
Nottingham

Dr Duncan Young,
Senior Clinical Lecturer and
Consultant, Nuffield Department
of Anaesthetics, University of
Oxford

Professor Martin Underwood,
Warwick Medical School,
University of Warwick

Observers

Dr Morven Roberts,
Clinical Trials Manager, Health
Services and Public Health
Services Board, Medical Research
Council

HTA Clinical Evaluation and Trials Board

Chair,
Professor Sallie Lamb,
Director,
Warwick Clinical Trials Unit,
Warwick Medical School,
University of Warwick and Professor of
Rehabilitation,
Nuffield Department of Orthopaedic,
Rheumatology and Musculoskeletal Sciences,
University of Oxford

Deputy Chair,
Professor Jenny Hewison,
Professor of the Psychology of Health Care,
Leeds Institute of Health Sciences,
University of Leeds

Programme Director,
Professor Tom Walley, CBE,
Director, NIHR HTA programme, Professor of
Clinical Pharmacology, University of Liverpool

Members

Professor Keith Abrams,
Professor of Medical Statistics,
Department of Health Sciences,
University of Leicester

Professor Martin Bland,
Professor of Health Statistics,
Department of Health Sciences,
University of York

Professor Jane Blazeby,
Professor of Surgery and
Consultant Upper GI Surgeon,
Department of Social Medicine,
University of Bristol

Professor Julia M Brown,
Director, Clinical Trials Research
Unit, University of Leeds

Professor Alistair Burns,
Professor of Old Age Psychiatry,
Psychiatry Research Group, School
of Community-Based Medicine,
The University of Manchester &
National Clinical Director for
Dementia, Department of Health

Dr Jennifer Burr,
Director, Centre for Healthcare
Randomised trials (CHART),
University of Aberdeen

Professor Linda Davies,
Professor of Health Economics,
Health Sciences Research Group,
University of Manchester

Professor Simon Gilbody,
Prof of Psych Medicine and Health
Services Research, Department of
Health Sciences, University of York

Professor Steven Goodacre,
Professor and Consultant in
Emergency Medicine, School of
Health and Related Research,
University of Sheffield

Professor Dyfrig Hughes,
Professor of Pharmacoeconomics,
Centre for Economics and Policy
in Health, Institute of Medical
and Social Care Research, Bangor
University

Professor Paul Jones,
Professor of Respiratory Medicine,
Department of Cardiac and
Vascular Science, St George's
Hospital Medical School,
University of London

Professor Khalid Khan,
Professor of Women's Health and
Clinical Epidemiology, Barts and
the London School of Medicine,
Queen Mary, University of London

Professor Richard J McManus,
Professor of Primary Care
Cardiovascular Research, Primary
Care Clinical Sciences Building,
University of Birmingham

Professor Helen Rodgers,
Professor of Stroke Care, Institute
for Ageing and Health, Newcastle
University

Professor Ken Stein,
Professor of Public Health,
Peninsula Technology Assessment
Group, Peninsula College
of Medicine and Dentistry,
Universities of Exeter and
Plymouth

Professor Jonathan Sterne,
Professor of Medical Statistics
and Epidemiology, Department
of Social Medicine, University of
Bristol

Mr Andy Vail,
Senior Lecturer, Health Sciences
Research Group, University of
Manchester

Professor Clare Wilkinson,
Professor of General Practice and
Director of Research North Wales
Clinical School, Department of
Primary Care and Public Health,
Cardiff University

Dr Ian B Wilkinson,
Senior Lecturer and Honorary
Consultant, Clinical Pharmacology
Unit, Department of Medicine,
University of Cambridge

Observers

Ms Kate Law,
Director of Clinical Trials,
Cancer Research UK

Dr Morven Roberts,
Clinical Trials Manager, Health
Services and Public Health
Services Board, Medical Research
Council

Diagnostic Technologies and Screening Panel

Members

<p>Chair, Professor Lindsay Wilson Turnbull, Scientific Director of the Centre for Magnetic Resonance Investigations and YCR Professor of Radiology, Hull Royal Infirmary</p> <p>Professor Judith E Adams, Consultant Radiologist, Manchester Royal Infirmary, Central Manchester & Manchester Children's University Hospitals NHS Trust, and Professor of Diagnostic Radiology, University of Manchester</p> <p>Mr Angus S Arunkalaivanan, Honorary Senior Lecturer, University of Birmingham and Consultant Urogynaecologist and Obstetrician, City Hospital, Birmingham</p>	<p>Dr Stephanie Dancer, Consultant Microbiologist, Hairmyres Hospital, East Kilbride</p> <p>Dr Diane Eccles, Professor of Cancer Genetics, Wessex Clinical Genetics Service, Princess Anne Hospital</p> <p>Dr Trevor Friedman, Consultant Liaison Psychiatrist, Brandon Unit, Leicester General Hospital</p> <p>Dr Ron Gray, Consultant, National Perinatal Epidemiology Unit, Institute of Health Sciences, University of Oxford</p> <p>Professor Paul D Griffiths, Professor of Radiology, Academic Unit of Radiology, University of Sheffield</p>	<p>Mr Martin Hooper, Service User Representative</p> <p>Professor Anthony Robert Kendrick, Associate Dean for Clinical Research and Professor of Primary Medical Care, University of Southampton</p> <p>Dr Anne Mackie, Director of Programmes, UK National Screening Committee, London</p> <p>Mr David Mathew, Service User Representative</p> <p>Dr Michael Millar, Consultant Senior Lecturer in Microbiology, Department of Pathology & Microbiology, Barts and The London NHS Trust, Royal London Hospital</p>	<p>Mrs Una Rennard, Service User Representative</p> <p>Dr Stuart Smellie, Consultant in Clinical Pathology, Bishop Auckland General Hospital</p> <p>Ms Jane Smith, Consultant Ultrasound Practitioner, Leeds Teaching Hospital NHS Trust, Leeds</p> <p>Dr Allison Streetly, Programme Director, NHS Sickle Cell and Thalassaemia Screening Programme, King's College School of Medicine</p> <p>Dr Alan J Williams, Consultant Physician, General and Respiratory Medicine, The Royal Bournemouth Hospital</p>
---	--	---	---

Observers

<p>Dr Tim Elliott, Team Leader, Cancer Screening, Department of Health</p> <p>Dr Catherine Moody, Programme Manager, Medical Research Council</p>	<p>Professor Julietta Patrick, Director, NHS Cancer Screening Programme, Sheffield</p> <p>Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health</p>	<p>Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool</p>	<p>Dr Ursula Wells, Principal Research Officer, Policy Research Programme, Department of Health</p>
---	--	---	---

Disease Prevention Panel

Members

<p>Chair, Professor Margaret Thorogood, Professor of Epidemiology, University of Warwick Medical School, Coventry</p> <p>Dr Robert Cook, Clinical Programmes Director, Bazian Ltd, London</p> <p>Dr Colin Greaves, Senior Research Fellow, Peninsula Medical School (Primary Care)</p> <p>Mr Michael Head, Service User Representative</p>	<p>Professor Cathy Jackson, Professor of Primary Care Medicine, Bute Medical School, University of St Andrews</p> <p>Dr Russell Jago, Senior Lecturer in Exercise, Nutrition and Health, Centre for Sport, Exercise and Health, University of Bristol</p> <p>Dr Julie Mytton, Consultant in Child Public Health, NHS Bristol</p>	<p>Professor Irwin Nazareth, Professor of Primary Care and Director, Department of Primary Care and Population Sciences, University College London</p> <p>Dr Richard Richards, Assistant Director of Public Health, Derbyshire Country Primary Care Trust</p> <p>Professor Ian Roberts, Professor of Epidemiology and Public Health, London School of Hygiene & Tropical Medicine</p>	<p>Dr Kenneth Robertson, Consultant Paediatrician, Royal Hospital for Sick Children, Glasgow</p> <p>Dr Catherine Swann, Associate Director, Centre for Public Health Excellence, NICE</p> <p>Professor Carol Tannahill, Glasgow Centre for Population Health</p> <p>Mrs Jean Thurston, Service User Representative</p> <p>Professor David Weller, Head, School of Clinical Science and Community Health, University of Edinburgh</p>
--	--	---	--

Observers

<p>Ms Christine McGuire, Research & Development, Department of Health</p>	<p>Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health</p>	<p>Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool</p>
---	---	---

External Devices and Physical Therapies Panel

Members

Chair, Dr John Pounsford, Consultant Physician North Bristol NHS Trust	Dr Dawn Carnes, Senior Research Fellow, Barts and the London School of Medicine and Dentistry	Professor Christine Norton, Professor of Clinical Nursing Innovation, Bucks New University and Imperial College Healthcare NHS Trust	Dr Pippa Tyrrell, Senior Lecturer/Consultant, Salford Royal Foundation Hospitals' Trust and University of Manchester
Deputy Chair, Professor E Andrea Nelson, Reader in Wound Healing and Director of Research, University of Leeds	Dr Emma Clark, Clinician Scientist Fellow & Cons. Rheumatologist, University of Bristol	Dr Lorraine Pinnigton, Associate Professor in Rehabilitation, University of Nottingham	Dr Sarah Tyson, Senior Research Fellow & Associate Head of School, University of Salford
Professor Bipin Bhakta, Charterhouse Professor in Rehabilitation Medicine, University of Leeds	Mrs Anthea De Barton-Watson, Service User Representative	Dr Kate Radford, Senior Lecturer (Research), University of Central Lancashire	Dr Nefyn Williams, Clinical Senior Lecturer, Cardiff University
Mrs Penny Calder, Service User Representative	Professor Nadine Foster, Professor of Musculoskeletal Health in Primary Care Arthritis Research, Keele University	Mr Jim Reece, Service User Representative	
	Dr Shaheen Hamdy, Clinical Senior Lecturer and Consultant Physician, University of Manchester	Professor Maria Stokes, Professor of Neuromusculoskeletal Rehabilitation, University of Southampton	

Observers

Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health	Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool	Dr Ursula Wells, Principal Research Officer, Policy Research Programme, Department of Health
--	--	---

Interventional Procedures Panel

Members

Chair, Professor Jonathan Michaels, Professor of Vascular Surgery, University of Sheffield	Ms Leonie Cooke, Service User Representative	Dr John Holden, General Practitioner, Garswood Surgery, Wigan	Dr Jane Montgomery, Consultant in Anaesthetics and Critical Care, South Devon Healthcare NHS Foundation Trust
Deputy Chair, Mr Michael Thomas, Consultant Colorectal Surgeon, Bristol Royal Infirmary	Mr Seumas Eckford, Consultant in Obstetrics & Gynaecology, North Devon District Hospital	Professor Nicholas James, Professor of Clinical Oncology, School of Cancer Sciences, University of Birmingham	Professor Jon Moss, Consultant Interventional Radiologist, North Glasgow Hospitals University NHS Trust
Mrs Isabel Boyer, Service User Representative	Professor Sam Eljamel, Consultant Neurosurgeon, Ninewells Hospital and Medical School, Dundee	Dr Fiona Lecky, Senior Lecturer/Honorary Consultant in Emergency Medicine, University of Manchester/Salford Royal Hospitals NHS Foundation Trust	Dr Simon Padley, Consultant Radiologist, Chelsea & Westminster Hospital
Mr David P Britt, Service User Representative	Dr Adele Fielding, Senior Lecturer and Honorary Consultant in Haematology, University College London Medical School	Dr Nadim Malik, Consultant Cardiologist/Honorary Lecturer, University of Manchester	Dr Ashish Paul, Medical Director, Bedfordshire PCT
Mr Sankaran Chandra Sekharan, Consultant Surgeon, Breast Surgery, Colchester Hospital University NHS Foundation Trust	Dr Matthew Hatton, Consultant in Clinical Oncology, Sheffield Teaching Hospital Foundation Trust	Mr Hisham Mehanna, Consultant & Honorary Associate Professor, University Hospitals Coventry & Warwickshire NHS Trust	Dr Sarah Purdy, Consultant Senior Lecturer, University of Bristol
Professor Nicholas Clarke, Consultant Orthopaedic Surgeon, Southampton University Hospitals NHS Trust			Professor Yit Chiun Yang, Consultant Ophthalmologist, Royal Wolverhampton Hospitals NHS Trust

Observers

Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health	Dr Morven Roberts, Clinical Trials Manager, Health Services and Public Health Services Board, Medical Research Council	Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool	Dr Ursula Wells, Principal Research Officer, Policy Research Programme, Department of Health
--	--	--	---

Pharmaceuticals Panel

Members

Chair, Professor Imti Choonara, Professor in Child Health, University of Nottingham	Dr James Gray, Consultant Microbiologist, Department of Microbiology, Birmingham Children's Hospital NHS Foundation Trust	Dr Dyfrig Hughes, Reader in Pharmacoeconomics and Deputy Director, Centre for Economics and Policy in Health, IMSCar, Bangor University	Dr Gillian Shepherd, Director, Health and Clinical Excellence, Merck Serono Ltd
Deputy Chair, Dr Yoon K Loke, Senior Lecturer in Clinical Pharmacology, University of East Anglia	Ms Kylie Gyertson, Oncology and Haematology Clinical Trials Manager, Guy's and St Thomas' NHS Foundation Trust London	Dr Maria Kouimtzi, Pharmacy and Informatics Director, Global Clinical Solutions, Wiley-Blackwell	Mrs Katrina Simister, Assistant Director New Medicines, National Prescribing Centre, Liverpool
Dr Martin Ashton-Key, Medical Advisor, National Commissioning Group, NHS London	Dr Jurjees Hasan, Consultant in Medical Oncology, The Christie, Manchester	Professor Femi Oyebo, Consultant Psychiatrist and Head of Department, University of Birmingham	Professor Donald Singer, Professor of Clinical Pharmacology and Therapeutics, Clinical Sciences Research Institute, CSB, University of Warwick Medical School
Mr John Chapman, Service User Representative	Dr Carl Heneghan, Deputy Director Centre for Evidence-Based Medicine and Clinical Lecturer, Department of Primary Health Care, University of Oxford	Dr Andrew Prentice, Senior Lecturer and Consultant Obstetrician and Gynaecologist, The Rosie Hospital, University of Cambridge	Mr David Symes, Service User Representative
Dr Peter Elton, Director of Public Health, Bury Primary Care Trust		Ms Amanda Roberts, Service User Representative	Dr Arnold Zermansky, General Practitioner, Senior Research Fellow, Pharmacy Practice and Medicines Management Group, Leeds University
Dr Ben Goldacre, Research Fellow, Division of Psychological Medicine and Psychiatry, King's College London		Dr Martin Shelly, General Practitioner, Silver Lane Surgery, Leeds	

Observers

Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health	Dr Heike Weber, Programme Manager, Medical Research Council	Dr Ursula Wells, Principal Research Officer, Policy Research Programme, Department of Health
Mr Simon Reeve, Head of Clinical and Cost- Effectiveness, Medicines, Pharmacy and Industry Group, Department of Health	Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool	

Psychological and Community Therapies Panel

Members

Chair, Professor Scott Weich, Professor of Psychiatry, University of Warwick, Coventry	Mrs Val Carlill, Service User Representative	Dr Jeremy J Murphy, Consultant Physician and Cardiologist, County Durham and Darlington Foundation Trust	Dr Paul Ramchandani, Senior Research Fellow/Cons. Child Psychiatrist, University of Oxford
Deputy Chair, Dr Howard Ring, Consultant & University Lecturer in Psychiatry, University of Cambridge	Dr Steve Cunningham, Consultant Respiratory Paediatrician, Lothian Health Board	Dr Richard Neal, Clinical Senior Lecturer in General Practice, Cardiff University	Dr Karen Roberts, Nurse/Consultant, Dunston Hill Hospital, Tyne and Wear
Professor Jane Barlow, Professor of Public Health in the Early Years, Health Sciences Research Institute, Warwick Medical School	Dr Anne Hesketh, Senior Clinical Lecturer in Speech and Language Therapy, University of Manchester	Mr John Needham, Service User Representative	Dr Karim Saad, Consultant in Old Age Psychiatry, Coventry and Warwickshire Partnership Trust
Dr Sabyasachi Bhaumik, Consultant Psychiatrist, Leicestershire Partnership NHS Trust	Dr Peter Langdon, Senior Clinical Lecturer, School of Medicine, Health Policy and Practice, University of East Anglia	Ms Mary Nettle, Mental Health User Consultant	Dr Lesley Stockton, Lecturer, School of Health Sciences, University of Liverpool
	Dr Yann Lefevre, GP Partner, Burrage Road Surgery, London	Professor John Potter, Professor of Ageing and Stroke Medicine, University of East Anglia	Dr Simon Wright, GP Partner, Walkden Medical Centre, Manchester
		Dr Greta Rait, Senior Clinical Lecturer and General Practitioner, University College London	

Observers

Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health	Dr Morven Roberts, Clinical Trials Manager, Health Services and Public Health Services Board, Medical Research Council	Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool	Dr Ursula Wells, Principal Research Officer, Policy Research Programme, Department of Health
--	--	--	---

Expert Advisory Network

Members

Professor Douglas Altman,
Professor of Statistics in Medicine,
Centre for Statistics in Medicine,
University of Oxford

Professor John Bond,
Professor of Social Gerontology
& Health Services Research,
University of Newcastle upon Tyne

Professor Andrew Bradbury,
Professor of Vascular Surgery,
Solihull Hospital, Birmingham

Mr Shaun Brogan,
Chief Executive, Ridgeway
Primary Care Group, Aylesbury

Mrs Stella Burnside OBE,
Chief Executive, Regulation and
Improvement Authority, Belfast

Ms Tracy Bury,
Project Manager, World
Confederation of Physical Therapy,
London

Professor Iain T Cameron,
Professor of Obstetrics and
Gynaecology and Head of the
School of Medicine, University of
Southampton

Professor Bruce Campbell,
Consultant Vascular & General
Surgeon, Royal Devon & Exeter
Hospital, Wonford

Dr Christine Clark,
Medical Writer and Consultant
Pharmacist, Rossendale

Professor Collette Clifford,
Professor of Nursing and Head
of Research, The Medical School,
University of Birmingham

Professor Barry Cookson,
Director, Laboratory of Hospital
Infection, Public Health
Laboratory Service, London

Dr Carl Counsell,
Clinical Senior Lecturer in
Neurology, University of Aberdeen

Professor Howard Cuckle,
Professor of Reproductive
Epidemiology, Department
of Paediatrics, Obstetrics &
Gynaecology, University of Leeds

Professor Carol Dezateaux,
Professor of Paediatric
Epidemiology, Institute of Child
Health, London

Mr John Dunning,
Consultant Cardiothoracic
Surgeon, Papworth Hospital NHS
Trust, Cambridge

Mr Jonathan Earnshaw,
Consultant Vascular Surgeon,
Gloucestershire Royal Hospital,
Gloucester

Professor Martin Eccles,
Professor of Clinical Effectiveness,
Centre for Health Services
Research, University of Newcastle
upon Tyne

Professor Pam Enderby,
Dean of Faculty of Medicine,
Institute of General Practice
and Primary Care, University of
Sheffield

Professor Gene Feder,
Professor of Primary Care
Research & Development, Centre
for Health Sciences, Barts and The
London School of Medicine and
Dentistry

Mr Leonard R Fenwick,
Chief Executive, Freeman
Hospital, Newcastle upon Tyne

Mrs Gillian Fletcher,
Antenatal Teacher and Tutor and
President, National Childbirth
Trust, Henfield

Professor Jayne Franklyn,
Professor of Medicine, University
of Birmingham

Mr Tam Fry,
Honorary Chairman, Child
Growth Foundation, London

Professor Fiona Gilbert,
Consultant Radiologist and NCRN
Member, University of Aberdeen

Professor Paul Gregg,
Professor of Orthopaedic Surgical
Science, South Tees Hospital NHS
Trust

Bec Hanley,
Co-director, TwoCan Associates,
West Sussex

Dr Maryann L Hardy,
Senior Lecturer, University of
Bradford

Mrs Sharon Hart,
Healthcare Management
Consultant, Reading

Professor Robert E Hawkins,
CRC Professor and Director of
Medical Oncology, Christie CRC
Research Centre, Christie Hospital
NHS Trust, Manchester

Professor Richard Hobbs,
Head of Department of Primary
Care & General Practice,
University of Birmingham

Professor Alan Horwich,
Dean and Section Chairman,
The Institute of Cancer Research,
London

Professor Allen Hutchinson,
Director of Public Health and
Deputy Dean of ScHARR,
University of Sheffield

Professor Peter Jones,
Professor of Psychiatry, University
of Cambridge, Cambridge

Professor Stan Kaye,
Cancer Research UK Professor of
Medical Oncology, Royal Marsden
Hospital and Institute of Cancer
Research, Surrey

Dr Duncan Keeley,
General Practitioner (Dr Burch &
Ptnrs), The Health Centre, Thame

Dr Donna Lamping,
Research Degrees Programme
Director and Reader in
Psychology, Health Services
Research Unit, London School of
Hygiene and Tropical Medicine,
London

Professor James Lindesay,
Professor of Psychiatry for the
Elderly, University of Leicester

Professor Julian Little,
Professor of Human Genome
Epidemiology, University of
Ottawa

Professor Alistaire McGuire,
Professor of Health Economics,
London School of Economics

Professor Neill McIntosh,
Edward Clark Professor of Child
Life and Health, University of
Edinburgh

Professor Rajan Madhok,
Consultant in Public Health, South
Manchester Primary Care Trust

Professor Sir Alexander Markham,
Director, Molecular Medicine
Unit, St James's University
Hospital, Leeds

Dr Peter Moore,
Freelance Science Writer, Ashted

Dr Andrew Mortimore,
Public Health Director,
Southampton City Primary Care
Trust

Dr Sue Moss,
Associate Director, Cancer
Screening Evaluation Unit,
Institute of Cancer Research,
Sutton

Professor Miranda Mugford,
Professor of Health Economics
and Group Co-ordinator,
University of East Anglia

Professor Jim Neilson,
Head of School of Reproductive
& Developmental Medicine
and Professor of Obstetrics
and Gynaecology, University of
Liverpool

Mrs Julietta Patnick,
Director, NHS Cancer Screening
Programmes, Sheffield

Professor Robert Peveler,
Professor of Liaison Psychiatry,
Royal South Hants Hospital,
Southampton

Professor Chris Price,
Director of Clinical Research,
Bayer Diagnostics Europe, Stoke
Poges

Professor William Rosenberg,
Professor of Hepatology and
Consultant Physician, University
of Southampton

Professor Peter Sandercock,
Professor of Medical Neurology,
Department of Clinical
Neurosciences, University of
Edinburgh

Dr Philip Shackley,
Senior Lecturer in Health
Economics, Sheffield Vascular
Institute, University of Sheffield

Dr Eamonn Sheridan,
Consultant in Clinical Genetics, St
James's University Hospital, Leeds

Dr Margaret Somerville,
Director of Public Health
Learning, Peninsula Medical
School, University of Plymouth

Professor Sarah Stewart-Brown,
Professor of Public Health,
Division of Health in the
Community, University of
Warwick, Coventry

Dr Nick Summerton,
GP Appraiser and Codirector,
Research Network, Yorkshire
Clinical Consultant, Primary Care
and Public Health, University of
Oxford

Professor Ala Szczepura,
Professor of Health Service
Research, Centre for Health
Services Studies, University of
Warwick, Coventry

Dr Ross Taylor,
Senior Lecturer, University of
Aberdeen

Dr Richard Tiner,
Medical Director, Medical
Department, Association of the
British Pharmaceutical Industry

Mrs Joan Webster,
Consumer Member, Southern
Derbyshire Community Health
Council

Professor Martin Whittle,
Clinical Co-director, National
Co-ordinating Centre for Women's
and Children's Health, Lymington

Feedback

The HTA programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (www.hta.ac.uk) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.