Cervical screening programmes: can automation help?

Cervical screening programmes: can automation help? Evidence from systematic reviews, an economic analysis and a simulation modelling exercise applied to the UK

BH Willis,[1] P Barton,[2] P Pearmain,[3] S Bryan[2] and C Hyde[1*]

[1] CAIF, Department of Public Health and Epidemiology, University of Birmingham, UK
[2] Health Economics Facility, Health Services Management Centre, University of Birmingham, UK
[3] West Midlands Breast and Cervical Screening QA Reference Centre, University of Birmingham, UK

* Corresponding author

Executive summary

Health Technology Assessment 2005; Vol. 9: No. 13

Health Technology Assessment
NHS R&D HTA Programme

Executive summary

Background

Cervical cancer is a serious, but fortunately rare disease. Cervical screening programmes have undoubtedly contributed to reductions in incidence and mortality, but the cost, both financial and logistic, has been high. It has been hoped that technological advances, including automated image analysis of cervical smears, would help. However, the technology is expensive, the only currently commercially available automated image analysis devices costing in excess of £0.5 million each in 2001. The implied implementation cost for the NHS in England alone is conservatively estimated at £40 million. Inevitably, there has been concern about whether such costs can be justified.

Automated image analysis involves the translation of a cervical smear into a computerised image, which is then analysed to identify slides with cells likely to be abnormal. Increasingly, the location of abnormal cells on the slide is automatically recorded to facilitate review of the slide. Automated image analysis is incorporated into the existing manual screening procedure. In current devices the attention is on replacing the primary screening step. There has been considerable development of the technology since it was first introduced in the early 1990s. PAPNET and AutoPap have been the two main competing devices; however, commercial pressure has meant that the AutoPap Guided Screening (GS) System is now the only one available.

This assessment was completed in April 2002. Device and manufacturer names used in this report were correct at the time the report was written. Subsequently, the authors were informed that the device name of AutoPap had changed to FocalPoint slide profiler and the manufacturer from TriPath Imaging Inc to TriPath Care Technologies Inc. (Jackson AK, CellPath plc, UK: personal communication, 28 February 2002). While noting these changes, the authors have not altered manufacturer or device names used in the original version of the report.

Objectives

The overall objective of the project was to assess the immediate effects, the wider consequences and costs, and overall cost-effectiveness and cost-utility of introducing automated image analysis to a screening programme with characteristics similar to those currently operating in the UK.

Methods

A health technology assessment was undertaken. This had six interrelated components, each with its own specific objectives. Four systematic reviews of past reviews and health technology assessments, assessments of cost-effectiveness, assessments of clinical effectiveness and cost data, supplemented with a detailed survey for unpublished UK literature, fed into an attempt to model the cost-effectiveness of automated image analysis relative to manual screening alone. A discrete event simulation (DES) model of cervical screening was developed to overcome some anticipated limitations of other modelling approaches. All systematic reviews were carried out in accordance with recognised guidance. The searches for the systematic reviews covered all major electronic databases to the end of 2000. A special feature of the clinical effectiveness review was that studies assessing reproducibility, impact on process and impact on health outcomes were targeted in addition to studies assessing test performance.

Results

The predominant finding from the systematic reviews was the very limited amount of rigorously conducted primary research. For instance, concerning test performance, only two studies (approximately 13,000 slides) assessing impact on sensitivity and specificity of automated image analysis were included; even relaxing these criteria only allowed another five studies (approximately 51,000 slides) to be considered. The results of these studies were difficult to interpret, but debatably were most compatible with automated image analysis being equivalent in test performance to manual screening. Several studies provided information on reproducibility of assessments, which was often surprisingly poor. Two evaluations of impact on health outcomes were identified, and although they did not

contribute directly to the conclusions, they point to a type of evaluation that should be considered more often. Concerning process, there was evidence that automated image analysis does lead to reductions in average slide processing times. In the PRISMATIC trial this was reduced from 10.4 to 3.9 minutes using PAPNET, a statistically significant and practically important difference.

There are two important provisos to these findings. First, none of the included studies above refers to the only currently commercially available automated image analysis device, the AutoPap GS System. The majority of evaluations on test performance and impact on processing times have been performed on PAPNET. Second, detailed searches for UK unpublished literature on the test performance of automated image analysis revealed 13 studies, two of which appeared to be similar in quality to the studies included. This suggests that the findings are possibly highly susceptible to publication bias.

Concerning cost-effectiveness, although the DES model was developed, the authors were not satisfied with its validation. Given the possibility of equivalence of test performance, a cost-minimisation analysis was also used. This tentatively suggested that the AutoPap GS system may be efficient. The key proviso is that credible data become available that the AutoPap GS system has test performance and processing times equivalent to those obtained for PAPNET.

Conclusions

As in previous health technology assessments on this subject, the conclusion is that the available evidence on test performance, impact on process and cost-effectiveness is still insufficient to recommend implementation of automated image analysis systems. The priority for action remains further research. An important difference is that previously the insufficiency of evidence was general. Now, a general case for automated image analysis has probably just been made, but is specifically absent for the single device currently commercially available. The findings with respect to other and in many cases older automated image analysis devices need to be confirmed for the AutoPap GS System.

Implications for research on automated image analysis

The areas of greatest priority are:

- 'clinical effectiveness' of the AutoPap GS System relative to existing cervical screening programmes
- further development of the DES model presented in this report, particularly its validation
- further assessment of the cost-effectiveness of the introduction of automation alongside other approaches, including non-technological, to improving cervical screening
- further research on the effectiveness and costs of these other approaches.

Public research funding bodies should consider taking a greater lead in future research to ensure its independence and methodological rigour.

Implications for methodological research

There are many areas that may be pursued, in particular:

- research on the advantages and disadvantages of different research designs assessing the test performance of screening or diagnostic tests, especially two-armed designs
- research on the conduct of systematic reviews of dimensions of the impact of screening and diagnostic tests, other than test performance, especially their reproducibility and impact on process
- further research on publication bias, especially the role and conduct of detailed surveys for unpublished literature.

Publication

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).

NHS R&D HTA Programme

The research findings from the NHS R&D Health Technology Assessment (HTA) Programme directly influence key decision-making bodies such as the National Institute for Clinical Excellence (NICE) and the National Screening Committee (NSC) who rely on HTA outputs to help raise standards of care. HTA findings also help to improve the quality of the service in the NHS indirectly in that they form a key component of the 'National Knowledge Service' that is being developed to improve the evidence of clinical practice throughout the NHS.

The HTA Programme was set up in 1993. Its role is to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined to include all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care, rather than settings of care.

The HTA programme commissions research only on topics where it has identified key gaps in the evidence needed by the NHS. Suggestions for topics are actively sought from people working in the NHS, the public, consumer groups and professional bodies such as Royal Colleges and NHS Trusts.

Research suggestions are carefully considered by panels of independent experts (including consumers) whose advice results in a ranked list of recommended research priorities. The HTA Programme then commissions the research team best suited to undertake the work, in the manner most appropriate to find the relevant answers. Some projects may take only months, others need several years to answer the research questions adequately. They may involve synthesising existing evidence or designing a trial to produce new evidence where none currently exists.

Additionally, through its Technology Assessment Report (TAR) call-off contract, the HTA Programme is able to commission bespoke reports, principally for NICE, but also for other policy customers, such as a National Clinical Director. TARs bring together evidence on key aspects of the use of specific technologies and usually have to be completed within a limited time period.

Criteria for inclusion in the HTA monograph series
Reports are published in the HTA monograph series if (1) they have resulted from work commissioned 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 monograph was commissioned and funded by the HTA Programme on behalf of NICE as project number 98/38/01. 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, NICE or the Department of Health.

ISSN 1366-5278

Published by Gray Publishing, Tunbridge Wells, Kent, on behalf of NCCHTA.
Printed on acid-free paper in the UK by St Edmundsbury Press Ltd, Bury St Edmunds, Suffolk.