

# Systematic reviews of clinical decision tools for acute abdominal pain

JLY Liu,<sup>1-3</sup> JC Wyatt,<sup>3-5\*</sup> JJ Deeks,<sup>1,6</sup> S Clamp,<sup>7</sup>  
J Keen,<sup>7</sup> P Verde,<sup>8,9</sup> C Ohmann,<sup>8</sup> J Wellwood,<sup>10</sup>  
M Dawes<sup>11,12</sup> and DG Altman<sup>1</sup>

<sup>1</sup> NHS/Cancer Research UK Centre for Statistics in Medicine, Wolfson College, Oxford University, UK

<sup>2</sup> Department of Public Health, Oxford University, UK

<sup>3</sup> Health Informatics Centre, University of Dundee, UK

<sup>4</sup> National Institute for Health and Clinical Excellence (NICE), London, UK

<sup>5</sup> Department of Primary Health Care, Oxford University, UK

<sup>6</sup> Department of Public Health and Epidemiology, University of Birmingham, UK

<sup>7</sup> Yorkshire Centre for Health Informatics, University of Leeds, UK

<sup>8</sup> Coordination Centre for Clinical Trials and Theoretical Surgery Unit, Heinrich-Heine University Düsseldorf, Germany

<sup>9</sup> Department of Statistics, University of Dortmund, Germany

<sup>10</sup> Department of Surgery, Whipps Cross Hospital, London, UK

<sup>11</sup> NHS R&D Centre for Evidence-Based Medicine, Oxford University, UK

<sup>12</sup> Department of Family Medicine, McGill University, Montreal, Canada

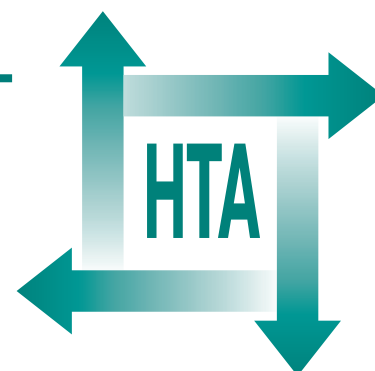
\* Corresponding author



## Executive summary

*Health Technology Assessment* 2006; Vol. 10: No. 47

Health Technology Assessment  
NHS R&D HTA Programme





## Executive summary

### Background

Making accurate decisions for patients with acute abdominal pain (AAP) is difficult. To avoid missing seriously ill patients, many undergo unnecessary surgery, with negative laparotomy rates of 25%. However, delays can lead to 20% perforation rates. Many conditions cause AAP and no single clinical finding or test is both specific and sensitive. Many decision tools (DTs) combining two or more findings have been developed to aid AAP management, but no consensus exists on their appropriateness for clinical use.

### Objectives

The study aimed to answer the following questions.

1. What are the diagnostic accuracies of DTs and doctors aided by DTs compared with those of unaided doctors?
2. What is the impact of providing doctors with an AAP DT on patient outcomes, clinical decisions and actions?
3. What factors are likely to determine the usage rates and usability of a DT?
4. What are the associated costs and likely cost-effectiveness of these DTs in routine use in the UK?

### Methods

#### Data sources

MEDLINE, EMBASE, CINAHL, INSPEC CENTRAL, SIGLE and HEALTH-CD were searched for empirical English-language studies. Searches were conducted to 1 July 2003.

#### Study selection (inclusion criteria)

For question 1, the criteria for eligible studies included:

- Unselected patients with AAP were recruited consecutively or randomly sampled from a primary or secondary care setting.
- Patients had previously undiagnosed AAP lasting for 7 days or less from onset.

- The study reported accuracies of AAP DTs, with or without comparisons to unaided doctors' decisions.
- An adequate reference standard was described.
- Sensitivity and specificity could be calculated.

For question 2, the criteria for eligible studies included:

- The study was a randomised controlled trial (RCT) or quasi-RCT.
- The patients were the same as for question 1.
- Evaluations were conducted of the impact of AAP DTs, compared with unaided doctors' decisions.
- The study reported some measure of impact on patient outcomes, clinical decisions or actions.

#### Data extraction

Data from each eligible study were extracted. For question 1, this included patient characteristics, type of DT, healthcare setting, and the accuracy of DTs and unaided doctors' decisions. For question 2, this included outcomes, clinical decisions and actions for patients of doctors aided or unaided by a DT. Potential sources of heterogeneity were extracted for both questions.

#### Data synthesis

For the accuracy review, meta-analysis was conducted. Among studies comparing diagnostic accuracies of DTs with unaided doctors, error rate ratios provided estimates of the differences between the false-negative and false-positive rates of the DT and unaided doctors' performance. Pooled error rate ratios and 95% confidence intervals (CIs) for false-negative rates and false-positive rates were computed. Metaregression was used to explore heterogeneity.

### Results

#### Question 1

Thirty-two studies from 27 articles, all based in secondary care, were eligible for the review of DT accuracies, while two were eligible for the review of the accuracy of hospital doctors aided by DTs. Sensitivities and specificities for DTs ranged from

53 to 99% and 30 to 99%, respectively. Those for unaided doctors ranged from 64 to 93% and 39 to 91%, respectively. Thirteen studies reported false-positive and false-negative rates for both DTs and unaided doctors, enabling a direct comparison of their performance. In random effects meta-analyses, DTs had significantly lower false-positive rates (error rate ratio 0.62, 95% CI 0.46 to 0.83) than unaided doctors. DTs may have higher false-negative rates than unaided doctors (error rate ratio 1.34, 95% CI 0.93 to 1.93). Significant heterogeneity was present.

Two studies compared the diagnostic accuracies of doctors aided by DTs to unaided doctors. In a multiarm cluster RCT ( $n = 5193$ ), the diagnostic accuracy of doctors not given access to DTs was not significantly worse (sensitivity 28.4% and specificity 96.0%) than that of three groups of aided doctors (sensitivities of 42.4–47.9%, and specificities of 95.5–96.5%, respectively). In an uncontrolled before-and-after study ( $n = 1484$ ), the sensitivities and specificities of aided and unaided doctors were 95.5% and 91.5% ( $p = 0.24$ ) and 78.1% and 86.4% ( $p < 0.001$ ), respectively.

The metaregression of DTs showed that:

- prospective test-set validation at the site of the tool's development was associated with considerably higher diagnostic accuracy than prospective test-set validation at an independent centre [relative diagnostic odds ratio (RDOR) 8.2; 95% CI 3.1 to 14.7]
- the earlier in the year the study was performed the higher the performance (RDOR 0.88, 0.83 to 0.92)
- when developers evaluated their own DT there was better performance than when independent evaluators carried out the study (RDOR = 3.0, 1.3 to 6.8)
- there was no evidence of association between other quality indicators and DT accuracy.

### Question 2

The one eligible study of the impact study review, a four-arm cluster randomised trial ( $n = 5193$ ), showed that hospital admission rates of patients by doctors not allocated to a DT (42.8%) were significantly higher than those by doctors allocated to three combinations of decision support (34.2–38.5%) ( $p < 0.001$ ). There was no evidence of a difference between perforation rates ( $p = 0.19$ ) and negative laparotomy rates in the four trial arms ( $p = 0.46$ ).

### Question 3

Usage rates of DTs by doctors in accident and emergency departments ranged from 10 to 77% in the six studies that reported them. Possible determinants of usability include the reasoning method used, the number of items used and the output format.

### Question 4

A deterministic cost-effectiveness comparison demonstrated that a paper checklist is likely to be 100–900 times more cost-effective than a computer-based DT, under stated assumptions.

## Conclusions

### Implications for healthcare

- With their significantly greater specificity and lower false-positive rates than doctors, DTs are potentially useful in confirming a diagnosis of acute appendicitis, but not in ruling it out.
- The clinical use of well-designed, condition-specific paper or computer-based structured checklists is promising as a way to improve impact on patient outcomes, subject to further research.

## Recommendations for research

This review uncovered important evidence gaps. The authors' research recommendations include the following:

- Primary research to compare paper-based checklists with computer-based tools exploring the type/format that maximises patient benefit.
- Empirical research to identify the determinants of successful DTs, to provide more evidence to support the development of clinically useful tools.
- More general systematic reviews (across a range of diseases or tools) to assess (1) factors that make DTs more acceptable to doctors and patients and (2) the relative clinical value of paper checklists versus computer-based tools.

## Publication

Liu JLY, Wyatt JC, Deeks JJ, Clamp S, Keen J, Verde P, *et al.* Systematic reviews of clinical decision tools for acute abdominal pain. *Health Technol Assess* 2006;**10**(47).

# 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 Health and 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, service-users groups and professional bodies such as Royal Colleges and NHS Trusts. Research suggestions are carefully considered by panels of independent experts (including service users) 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 conducting 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 short 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 by the HTA Programme as project number 97/38/03. The contractual start date was in May 2001. The draft report began editorial review in February 2006 and was accepted for publication in March 2006. 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  
Series Editors: Dr Aileen Clarke, Dr Peter Davidson, Dr Chris Hyde,  
Dr John Powell, Dr Rob Riemsma and Dr Ken Stein  
Managing Editors: Sally Bailey and Sarah Llewellyn Lloyd

ISSN 1366-5278

© Queen's Printer and Controller of HMSO 2006

This monograph 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 NCCHTA, Mailpoint 728, Boldrewood, University of Southampton, Southampton, SO16 7PX, UK.

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.