Tests for the diagnosis and investigation of urinary tract infection in children

Clinical effectiveness and cost-effectiveness of tests for the diagnosis and investigation of urinary tract infection in children: a systematic review and economic model

P Whiting,[1] M Westwood,[2*] L Bojke,[3] S Palmer,[3] G Richardson,[3] J Cooper,[4] I Watt,[5] J Glanville,[2] M Sculpher[3] and J Kleijnen[2]

[1] Department of Social Medicine, MRC HSRC, Bristol, UK
[2] Centre for Reviews and Dissemination, University of York, UK
[3] Centre for Health Economics, University of York, UK
[4] Radiology Department, York Hospital, UK
[5] Department of Health Sciences, University of York, UK

* Corresponding author

Executive summary

Health Technology Assessment 2006; Vol. 10: No. 36

Health Technology Assessment
NHS R&D HTA Programme

Executive summary

Background

Urinary tract infection (UTI) is one of the most common sources of infection in children under 5 years of age. It is important as it can cause troublesome and recurrent symptoms and may point to unsuspected anomalies of the urinary tract. In a small proportion of children UTI may lead to renal scarring. This outcome of infection is of concern as it is associated with future complications including poor renal growth, recurrent adult pyelonephritis, impaired glomerular function, early hypertension and end-stage renal disease. The aim of management should be prompt diagnosis, rapid treatment and the detection of any underlying cause that might predispose to further infection or lead to long-term renal damage.

Objectives

The aims of this review were:

1. to determine the diagnostic accuracy of tests for detecting UTI in children under 5 years of age
2. to evaluate the effectiveness of tests used to investigate further children with confirmed UTI
3. to evaluate the effectiveness of following up children with UTI
4. to evaluate the cost-effectiveness of diagnostic and imaging tests for the diagnosis and follow-up of UTI in children under 5
5. to develop a preliminary diagnostic algorithm for healthcare professionals. This should be based, as far as possible, on information derived from objectives 1–4.

Methods

A systematic review was undertaken according to published guidelines.

Data sources

Studies were identified through searches of electronic databases, Internet searches, handsearching, scanning reference lists of included papers and consultation with experts in the field.

Study selection

Two reviewers screened titles and abstracts for relevance. Full papers of potentially relevant studies were obtained and assessed for inclusion by one reviewer and checked by a second. Published and unpublished studies in any language were eligible for inclusion.

Data extraction

Data extraction and quality assessment were performed by one reviewer and checked by a second.

Data synthesis

Results were analysed according to test grouping: diagnosis of UTI and further investigation of UTI. Within these groups, data were analysed according to the clinical aim of studies, and specific tests or test combinations reported in the literature. For each test the range in sensitivity, specificity and likelihood ratios (of both positive and negative tests results), and diagnostic odds ratios were calculated. Individual study results were presented graphically in receiver operating characteristic (ROC) space. Heterogeneity of likelihood ratios was investigated using the Q statistic and through visual examination of forest plots of study results. Pooled estimates of positive and negative likelihood ratios were calculated. However, owing to the significant heterogeneity present in most tests, median likelihood ratios, together with their interquartile ranges, were also calculated and presented. Where sufficient data were available, heterogeneity was further investigated using regression analysis. The summary ROC model was extended to include covariates for study quality and other possible sources of heterogeneity.

Economic evaluations

The cost-effectiveness results from existing evaluations were synthesised through a narrative review with full tabulation of the results of the included studies. A separate cost-effectiveness model was developed using the best available evidence, in part derived from the results of the systematic review, to illustrate the potential cost-effectiveness of some alternative management strategies in a UK setting.

▶

Algorithm development

The results of the systematic review (objectives 1–3) were used to propose diagnostic algorithms for the diagnosis and further investigation of UTI in children. Economic analyses did not contribute directly to the development of these algorithms.

Results

Diagnosis of UTI

Clinical tests (six studies)

Very few studies of clinical tests for the diagnosis of UTI met the inclusion criteria. These examined a wide variety of clinical characteristics. No conclusions regarding the utility of the clinical examination in diagnosing UTI could be drawn from these studies.

Urine sampling (12 studies, 16 evaluations)

There was good agreement between culture of clean voided urine and suprapubic aspiration (SPA) urine samples. Only limited data were available on bag, pad and nappy samples. However, this did suggest that both bag and nappy/pad specimens may also be suitable alternatives to SPA.

Dipstick (38 studies, 106 evaluations)

It is difficult to draw conclusions about the overall accuracy of dipstick tests given the heterogeneity between studies in some areas, and the lack of data in others. There was insufficient information to make any judgement regarding the overall diagnostic accuracy of dipstick tests for protein or blood. The combination of a positive test for both nitrite and leucocyte esterase (LE) was found to be most accurate for ruling in disease, and a negative test for both nitrite and LE was found to be most accurate for ruling out disease. A test for the absence of urinary glucose was found to be considerably better than the other tests, for both ruling in and ruling out disease. However, only a limited number of studies of this test were included and these were conducted over 30 years ago.

Microscopy (39 studies, 101 evaluations)

Given the heterogeneity between studies within groups and the lack of data for combinations of tests, it is difficult to draw overall conclusions about the utility of microscopy techniques for the diagnosis of UTI. Microscopy positive for both pyuria and bacteriuria was found to be best for ruling in disease, and microscopy negative for both pyuria and bacteriuria was found to be best for ruling out disease.

Culture (nine studies)

There was considerable heterogeneity in studies of dipslide culture. The results suggested that this technique was less accurate for the diagnosis of UTI than either of the combinations of dipstick or microscopy tests outlined above.

Other tests (six studies)

Owing to the very small number of studies that looked at other tests there was insufficient information available to judge how useful any of them may be in the diagnosis of UTI.

Further investigation of UTI

Localisation of UTI (37 studies, 82 evaluations)

A limited number of studies of clinical and laboratory tests was identified that showed fairly poor accuracy for the localisation of UTI. Imaging techniques investigated included ultrasound, magnetic resonance imaging (MRI), computed tomography (CT), intravenous pyelography (IVP), cystography and various scintigraphic techniques. Scintigraphic techniques, generally regarded as the reference standard, were the only investigations able to localise UTI accurately.

Detection of reflux (34 studies, 57 evaluations)

Standard ultrasound techniques were found to have poor performance for the detection of reflux. Contrast-enhanced ultrasound techniques were accurate for both ruling in and for ruling out reflux. Other tests investigated were IVP, indirect voiding radionuclide cystography, *N*-acetyl-[β]-glucosaminidase/creatinine ratio, scintigraphy and a clinical risk scoring system. Although IVP and indirect voiding radionuclide cystography were both accurate for ruling in reflux, none of these tests was found to be useful for both ruling in and ruling out disease.

Prediction of scarring (four studies, nine evaluations)

The tests investigated were evaluated by one or two studies only; it is therefore not possible to draw conclusions regarding their utility in the prediction of renal scarring.

Detection of scarring (30 studies, 50 evaluations)

Static renal scintigraphy was found to have good diagnostic performance when evaluated using IVP as the reference standard. However, since renal scintigraphy itself, rather than IVP, is generally regarded as the appropriate reference standard, this evaluation is of limited value. Dynamic renal imaging using ^{99m}technetium-mercaptoacetyltriglycine ▶

was found to be reasonably comparable with ^{99m}technetium-dimercaptosuccinic acid scintigraphy. Ultrasound was found to be a reasonably good test for ruling in scarring, but less useful for ruling out disease. The association between the detection of reflux using micturating cystourethrography (MCUG) and the presence of scarring was found to be poor. Other tests investigated by a small number of studies were IVP, MRI, voiding radionuclide cystography and a combination of ultrasound and MCUG. IVP was found to have excellent specificity, but estimates of sensitivity showed considerable variation. Indirect voiding radionuclide cystography was found to be a poor test for the detection of scarring. The combination of ultrasound and MCUG was found to be a reasonable test for the detection of scarring, as was MRI. However, these were each investigated in only one study.

Multiple aims (eight studies, 17 evaluations)

Studies in this section used a wide variety of tests and combinations of tests as reference standards. The diagnostic accuracies reported by studies in this section were generally poor.

Effectiveness of follow-up (one study)

Only one study of the clinical effectiveness of imaging to investigate confirmed UTI was identified. This study was published as an abstract, and no additional data could be obtained. This study found that routine imaging of toilet-trained preschool and school-aged children with their first uncomplicated UTI led to higher rates of imaging, identification of reflux and prophylaxis than did selected imaging. However, it did not lead to a reduction in recurrent UTIs or renal scarring.

Economic evaluations

Only one study satisfied the inclusion criteria. The study was based on a comparison of a number of diagnostic strategies relating to UTI and the identification of urinary tract abnormalities and a model that linked evidence on diagnostic accuracy with that on therapeutic decisions and hence on health outcomes and costs. The review highlighted a number of potential limitations of this study for NHS decision-making. A separate decision-analytic model was therefore developed to provide a more reliable estimate of the optimal strategy regarding the diagnosis and further investigation of children under 5 with suspected UTI from the perspective of the NHS. The economic model found that the optimal diagnostic strategy for children presenting with symptoms suggestive of UTI depends on a number of key factors. These included the relevant subgroup of children concerned, in terms of gender and age, and the health service's maximum willingness to pay for an additional quality-adjusted life-year (QALY).

Conclusions

The results of the systematic review were used to derive an algorithm for the diagnosis of UTI in children under 5. This algorithm represents the conclusions of the review in terms of effective practice. There were insufficient data to propose an algorithm for the further investigation of UTI in children under 5; instead, the different imaging options are discussed and areas requiring further research are highlighted.

The quality assessment highlighted several areas that could be improved upon in future diagnostic accuracy studies. Future studies should follow the STARD guidelines for reporting of diagnostic accuracy studies.

Recommendations for research

The review highlighted the following specific areas requiring further research for the diagnosis of UTI:

- clinical signs and symptoms to select children to undergo testing for UTI
- urine sampling methods in younger children
- accuracy of clinical tests for the diagnosis of UTI
- accuracy of the glucose test, and its practical applicability
- handling of indeterminate nitrite and LE dipstick test results
- accuracy of microscopy in combination with a dipstick test
- usefulness of universal confirmatory culture
- usefulness of culture to determine antibiotic sensitivities in children with confirmed UTI.

Randomised controlled trials assessing the clinical effectiveness of all stages of the further investigation of UTI, for long-term renal outcomes, are urgently required. If the identification of reflux or renal scarring were found to be effective in any patient group, further,

well-designed diagnostic accuracy studies would be required to assess the potential of less invasive techniques to replace current reference standards. In the above case it would also be important to investigate options for minimising invasive testing by ruling out acute pyelonephritis. Non-invasive methods of localisation require further research addressed at this aim.

Publication

Whiting P, Westwood M, Bojke L, Palmer S, Richardson G, Cooper J, *et al*. Clinical effectiveness and cost-effectiveness of tests for the diagnosis and evaluation of urinary tract infection in children: a systematic review and economic model. *Health Technol Assess* 2006;**10**(36).

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 01/66/01. The contractual start date was in November 2002. The draft report began editorial review in August 2004 and was accepted for publication in January 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.

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.