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

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, M Westwood, L Bojke, S Palmer, G Richardson, J Cooper, I Watt, J Glanville, M Sculpher and J Kleijnen

October 2006

Health Technology Assessment
NHS R&D HTA Programme

How to obtain copies of this and other HTA Programme reports.
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs 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 monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

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

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

Declared competing interests of authors: none

Published October 2006

This report should be referenced as follows:

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 investigation of urinary tract infection in children: a systematic review and economic model. *Health Technol Assess* 2006;**10**(36).

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

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.

Abstract

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

Objectives: To determine the diagnostic accuracy of tests for detecting urinary tract infection (UTI) in children under 5 years of age and to evaluate the effectiveness of tests used to investigate further children with confirmed UTI. Also, to evaluate the effectiveness of following up children with UTI and the cost-effectiveness of diagnostic and imaging tests for the diagnosis and follow-up of UTI in children under 5. An additional objective was to develop a preliminary diagnostic algorithm for healthcare professionals.

Data sources: Electronic databases were searched up to the end of 2002/early 2003. Consultation with experts in the field.

Review methods: A systematic review was undertaken using published guidelines and results were analysed according to test grouping: diagnosis of UTI and further investigation of UTI. The cost-effectiveness results from existing evaluations were synthesised. 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. The results of the systematic review 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: The studies included in the review provided very little data on the accuracy of clinical investigations for the diagnosis of UTI, and criteria for clinical suspicion of UTI were not further defined. The majority of studies included in the review found that clean voided midstream urine (CVU) samples had similar accuracy to suprapubic aspiration (SPA) samples when cultured with the advantage of being a non-invasive collection method that can be used in the GP's surgery. Pad, nappy or bag specimens may be appropriate methods for obtaining a urine sample in non-toilet-trained children, although only limited data were available. Although the glucose test was reported to have the highest accuracy in terms of both ruling in and ruling out disease, only a limited number of studies of this test were included and these were conducted over 30 years ago. Dipstick tests are easy to perform in the GP's surgery, give an immediate result and are relatively cheap. The results of the systematic review showed that a dipstick for leucocyte esterase (LE) and nitrite, where both test results are interpreted in combination, was a good test both for ruling in (both positive) and ruling out (both negative) a UTI. A dipstick positive for either LE or nitrite and negative for the other provides inconclusive diagnostic information and further testing is therefore required in these patients. Microscopy is more time consuming and expensive to perform than a dipstick test, but potentially quicker and cheaper than culture. As with dipstick tests, a combination of microscopy for pyuria and bacteriuria can be used accurately to rule in and rule out a UTI. An indeterminate test result is again obtained if microscopy is positive for either pyuria or bacteriuria, and negative for the other. Confirmatory culture is required in these patients. In patients considered to have a UTI, further culture to determine antibiotic sensitivities may be an option to inform treatment decisions. Only one study satisfied the inclusion criteria of the economic review and 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.

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. The quality assessment highlighted several areas that could be improved upon in future diagnostic accuracy studies.

Contents

Glossary and list of abbreviations

Technical terms and abbreviations are used throughout this report. The meaning is usually clear from the context, but a glossary is provided for the non-specialist reader. In some cases, usage differs in the literature, but the term has a constant meaning throughout this review.

Glossary

This section summarises the measures of diagnostic test performance used in the review, and how these are calculated.

		UTI	
		Present	Absent
Test result	**+**	**a**	**b**
	–	**c**	**d**

True positives (TP) Correct positive test result: a = number of diseased persons with a positive test result.

True negatives (TN) Correct negative test result: d = number of non-diseased persons with a negative test result.

False positives (FP) Incorrect positive test result: b = number of non-diseased persons with a positive test result.

False negatives (FN) Incorrect negative test result: c = number of diseased persons with a negative test result.

Sensitivity $a/(a + c)$ = proportion of people with the target disorder who have a positive test result.

Specificity $d/(b + d)$ = proportion of people without the target disorder who have a negative test result.

Likelihood ratio (LR): positive (LR +ve), negative (LR –ve) Describes how many times a person with disease is more likely to receive a particular test result than a person without disease. A likelihood ratio of a positive test result is usually a number greater than 1, a likelihood ratio of a negative test result usually lies between 0 and 1.

$$\begin{aligned} LR+ &= \{a/(a + c)\}/\{b/(b + d)\} \\ &= \text{Sensitivity}/(1 - \text{Specificity}) \\ LR- &= \{c/(a + c)\}/\{d/(b + d)\} \\ &= (1 - \text{Sensitivity})/\text{Specificity} \end{aligned}$$

Diagnostic odds ratio (DOR) Used as an overall (single indicator) measure of the diagnostic accuracy of a diagnostic test. It is calculated as the odds of positivity among diseased persons, divided by the odds of positivity among non-diseased persons. When a test provides no diagnostic evidence then the DOR is 1.0.

$$\begin{aligned} DOR &= \{a/c\}/\{b/d\} \\ &= \{\text{Sensitivity}/(1 - \text{Specificity})\}/ \\ &\quad \{(1 - \text{Sensitivity})/\text{Specificity}\} \\ &= LR\ +\text{ve}/LR\ -\text{ve} = ad/bc \end{aligned}$$

Predictive value Positive predictive value (PPV): the probability of disease among all persons with a positive test result.

$$PPV = a/(a + b)$$

Negative predictive value (NPV): the probability of non-disease among all persons with a negative test result.

$$NPV = d(c + d)$$

Predictive values depend on disease prevalence; the more common a disease, the more likely it is that a positive test result is right and a negative result is wrong.

Receiver operating characteristic (ROC) curve An ROC curve represents the relationship between the true-positive fraction (sensitivity) and false-positive fraction (1 – specificity). It displays the trade-offs between sensitivity and specificity as a result of varying the cut-off value for positivity in case of a continuous test result.

continued

Glossary *continued*

Receiver operating characteristic (ROC) space The ROC space is the graphical area available for plotting the sensitivity vs (1 – specificity) for a binary classifier system.

Summary ROC (SROC) curve The SROC approach models test accuracy, defined by the log of the diagnostic odds ratio [D = logit(sensitivity) – logit(1 – specificity)], as a function of test threshold [S = logit(sensitivity) + logit(1 – specificity)]. S relates to the positivity threshold: it has a value of 0 in studies where sensitivity equals specificity, it is positive in studies where sensitivity is higher than specificity, and negative when specificity is higher than sensitivity. For a set of primary studies, the following linear regression model is fitted:

$$D = \alpha + \beta S$$

where D is the log odds ratio in each study, α is the intercept, which is the expected log odds ratio when $S = 0$, and β is the coefficient of S, indicating whether the log diagnostic odds ratio varies with the threshold.
The estimated SROC curve can be plotted by computing the expected sensitivity for each value of 1 – specificity across the range of the observed values. The expected sensitivity is given by:

$$\text{Sensitivity} = [1 + e^{-\alpha(1-\beta)} \cdot V^{(1+\beta)(1-\beta)}]^{-1}$$

where V = Specificity/(1 – Specificity).

List of abbreviations

ACB	antibody-coated bacteria
APN	acute pyelonephritis
CCT	controlled clinical trial
cfu	colony-forming units
CI	confidence interval
CR	creatinine
CRD	Centre for Reviews and Dissemination
CRP	C-reactive protein
CT	computed tomography
CVU	clean voided urine
DMSA	dimercaptosuccinic acid
DOR	diagnostic odds ratio
DTPA	diethylenetriamine pentaacetic acid
ERPF	effective renal plasma flow
ESR	erythrocyte sedimentation rate
ESRD	end-stage renal disease
FN	false negative
FP	false positive
FPR	false-positive rate
hpf	high-power field
IQR	interquartile range
IVP	intravenous pyelography
IVU	intravenous urography
KUB	kidney, ureter and bladder
LE	leucocyte esterase
LR	likelihood ratio
β_2M	β_2-Macroglobulin
MAG3	mercaptoacetyltriglycine
MCUG	micturating cystourethrography
α_1-MG	α_1-microglobulin
MRI	magnetic resonance imaging

continued

List of abbreviations *continued*

NA	not applicable
NAG	N-acetyl-[β]-glucosaminidase
NICE	National Institute for Health and Clinical Excellence
NMB	net monetary benefit
NPV	negative predictive value
NR	not reported
NS	not stated
oif	oil immersion fields
PCT	procalcitonin
PPV	positive predictive value
PRS	progressive renal scarring
QALY	quality-adjusted life-year
QUADAS	Quality Assessment of Diagnostic Studies
RCT	randomised controlled trial
RDOR	relative diagnostic odds ratio
ROC	receiver operating characteristic
SD	standard deviation
SE	standard error
SPA	suprapubic aspiration
SPECT	single-photon emission computed tomography
SR	systematic review
SROC	summary receiver operating characteristic
STIR	short T1 inversion recovery
T1-W	T1-weighted
^{99m}Tc	technetium-99m
TN	true negative
TP	true positive
TPR	true-positive rate
TTC	triphenyl-tetrazolium chloride
UA	urinalysis
US	ultrasound
UTI	urinary tract infection
VCUG	voiding cystourethrography
VUR	vesicoureteral reflux
WBC	white blood cells

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 at the end of the table.

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.

Chapter 1
Background

Epidemiology

The normal urinary tract is sterile. A urinary tract infection (UTI) is a microbial infection of the urethra, bladder, ureters or kidneys.[1] Infection is most commonly caused by Gram-negative aerobic bacteria.[2] *Escherichia coli* accounts for about 75–80% of community-acquired infections,[2–5] and the remainder are caused by *Proteus* spp. (more common in boys),[3–5] *Klebsiella* spp., *Pseudomonas* spp.,[4] and Gram-positive *Enterococcus* spp.[5] Occasionally, infection may be caused by other Gram-positive bacteria including *Staphylococcus* spp.[4]

Infection is important as it can cause troublesome and recurrent symptoms and may point to unsuspected anomalies of the urinary tract,[4] such as reflux. This occurs when urine passes from the bladder back into the ureter [vesicoureteral reflux (VUR)] or back further towards the kidney (vesicoureteric reflux).[1] The most common factor leading to UTI is urinary stasis.[4] This can result from reflux, bladder dysfunction, habitually infrequent or incomplete voiding, stones, outflow obstruction or constipation.[4] When examining children with UTI, it is important that those with complications, abnormalities of the kidneys, reflux or bladder dysfunction are identified.[6]

In a small proportion of children, especially in those less than 2 months of age, UTI may lead to renal scarring.[4,7] This outcome of infection is of concern as it is associated with future complications including poor renal growth, recurrent adult pyelonephritis (infection leading to inflammation of the kidney and its pelvis, beginning in the interstitium and rapidly extending to involve the tubules, glomeruli and blood vessels),[1] impaired glomerular function, early hypertension and end-stage renal disease (ESRD).[3] As renal scarring is symptomless, it is an important factor to screen for following UTI.[8] When pregnancy occurs later in life, scarring is associated with the risk of acute pyelonephritis, pre-eclampsia, operative delivery and induced delivery.[9]

Management

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.[4] Evidence-based guidelines propose that the management of UTI in children can be divided into four phases, as illustrated in *Figure 1*.

Phase 1: recognising a child at risk

The first phase involves identifying children at a high risk of UTI based on age and clinical factors.[10] If these children are not identified, the potential benefits of diagnosis and subsequent treatment are lost.[10] This phase involves clinical examination of the child with possible UTI infection.

Phase 2: diagnosing UTI

Children who are misdiagnosed either fail to receive appropriate treatment or receive unnecessary treatment and evaluation for urinary tract abnormalities.[10] As a result, they do not receive treatment for the real cause of their symptoms. It is therefore important that an accurate diagnosis is made as soon as possible.

Phase 3: short-term treatment

Guidelines recommend that a child with presumed UTI be given antibiotics pending the results of culture.[11] However, empirical treatment without establishing the diagnosis could hide other serious infection and delay appropriate investigation.[4] There is little evidence on the effect of giving early empirical treatment versus awaiting the results of microscopy or culture. There is some evidence from retrospective studies, which suggest that prompt treatment with antibiotics may reverse acute changes and prevent or limit complications such as renal scarring; however, the evidence is inconclusive.[3,12]

Antibiotics commonly used to treat UTI include trimethoprim, nitrofurantoin or cephalosporins.[12] Treatment duration is 5–7 days of antibiotics for acute infection.[13] Longer duration has not been shown to be any more effective, while shorter courses may increase recurrence and resistance.[12] Hospital admission is indicated in any child who is systemically ill or who is at significant risk of becoming seriously ill because of their age or the presence of urinary tract abnormalities.[14] Current UK recommendations state that after treatment of

the acute infection, prophylactic antibiotic therapy should be given in low dose at least until investigation of the urinary tract is complete.[13]

Phase 4: imaging evaluation of a child with UTI

Current UK recommendations state that all children should be investigated after their first confirmed infection.[13] This may demonstrate an abnormality requiring surgery, long-term antibiotic prophylaxis or, if the kidneys are scarred, lifelong follow-up. However, some studies have shown no evidence of benefit from routine diagnostic imaging of all children with a first UTI,[15] although there is indirect evidence that subgroups at increased risk of morbidity may benefit from investigation.[3] The objectives of imaging in children with their first UTI are to identify those at risk of scarring or reflux nephropathy, to detect reflux nephropathy or scars that have already occurred, and to identify vesicoureteric reflux, while minimising radiation exposure, morbidity and cost.[16] The performance of current tests used to evaluate children with UTI, together with evidence of long-term effectiveness of follow-up, needs to be assessed.

As shown in *Figure 1*, there are three phases in the management of UTI that involve diagnosis. A summary of currently used tests is provided in the following sections.

Diagnosis of UTI

Clinical tests

A clinical examination is the first step in the diagnosis of UTI. In the very young, symptoms of UTI are generally non-specific and are easily missed.[4,17] Many guidelines recommend that the presence of UTI should be considered in infants and young children aged 2 months to 2 years of age with unexplained fever.[11,12,18] However, the evidence on which this is based appears to be limited. Common clinical symptoms in children aged less than 2 years include feeding disorders, slow weight gain, vomiting, diarrhoea,[4] sepsis and failure to thrive.[11] Between 1 and 5 years of age, fever, general malaise, frequency, abdominal discomfort and delayed bladder control are common presenting features.[4] Dysuria (painful or difficult urination) in this age group may be a symptom of UTI or may be due to external

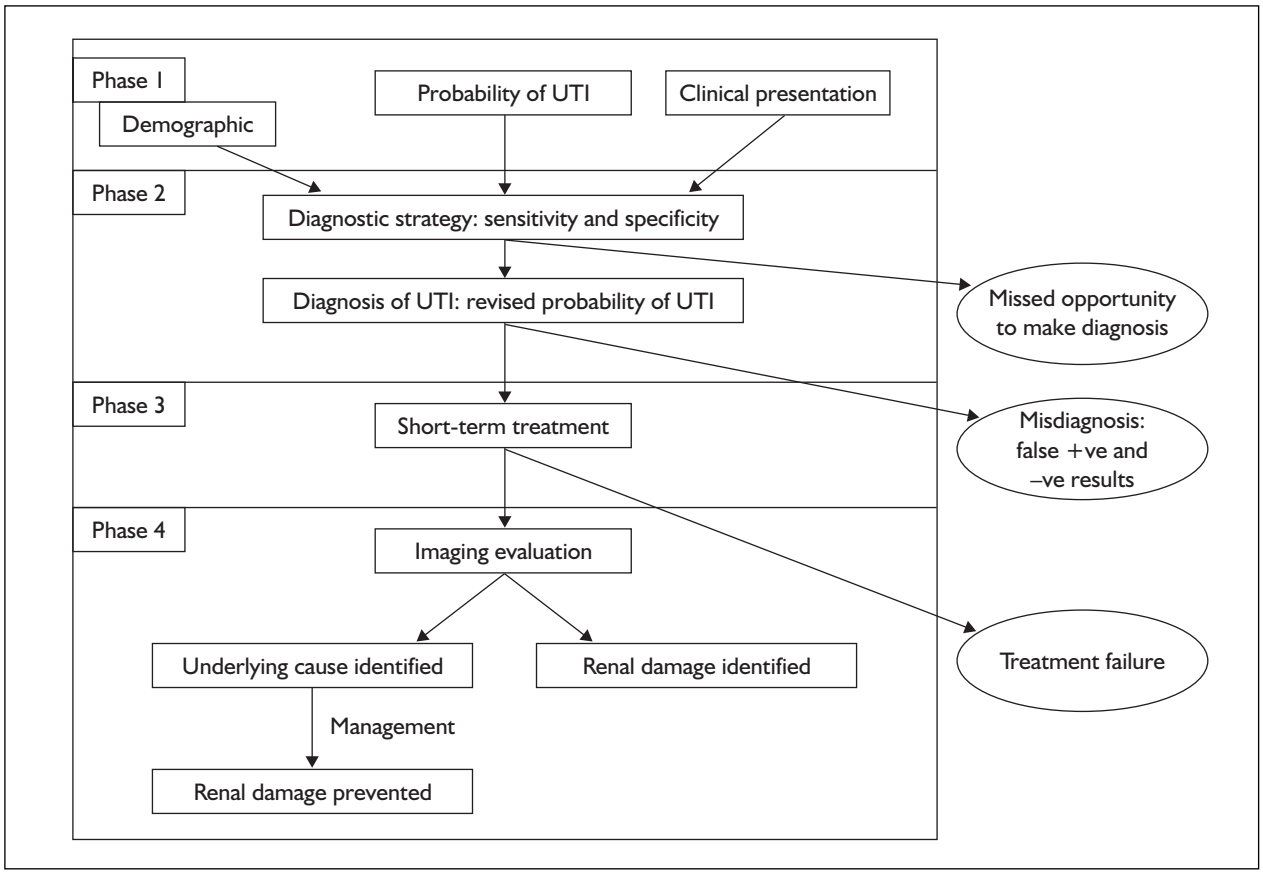

FIGURE 1 *Stages involved in the management of children with UTI*

irritation.[4] However, different articles report different symptoms of UTI in children. Thus, there is a need to identify which clinical symptoms are most consistently associated with UTI, and to estimate the test characteristics of a clinical diagnosis using various symptoms of UTI.

Urine tests

Urine sampling

All of the tests commonly used for the diagnosis of UTI are carried out on urine samples. The best sample with the least risk of contamination is obtained by suprapubic aspiration (SPA).[10] However, as this is an invasive procedure it is generally resisted.[10] Other methods of urine collection from children under 2 years of age include transurethral catheterisation or urine bags or pads. These methods are more susceptible to contamination, which may reduce the performance of diagnostic tests. In older children, a clean voided midstream urine (CVU) specimen is believed to be the preferred method of urine collection.[19]

Dipstick tests

Dipstick tests generally involve dipping the reactive section of a dry phase chemistry reagent strip briefly into urine and then comparing the colour change.[20] Analytes commonly tested for include leucocyte esterase (LE), nitrite, blood and protein.[10] Tests can be carried out individually or in combination; combinations may be serial (test results must be positive for the combination to be positive) or parallel (a positive result on any one test defines a positive result).[10]

Microscopy

Microscopic examination of urine samples for leucocytes or bacteria[10] is considerably more time consuming and labour intensive than the dipstick method.[21] Both unstained and Gram-stained samples are used for the detection of bacteria. Specimens may be centrifuged before examination.[10] When identifying leucocytes, it is important that the test is done shortly after the specimen is collected.[10] As with dipstick testing, tests can be carried out individually or in combination.

Culture

The reference standard test for UTI in children under 2 years of age is considered to be any bacterial growth on a culture of urine obtained by SPA.[10] The standard culturing technique involves streaking urine on enrichment and selective media. More recently dipslide and rapid culture methods have been developed. The dipslide is a miniature culture plate, which is immersed in the urine immediately after voiding.[13]

Other tests

A wide range of other tests has been used in the diagnosis of UTI. These include colorimetric tests, headspace gas analysis, impedance, bioluminescence and chemical luminescence, immunological tests, enzyme tests, bacterial oxygen consumption and turbidimetry. However, these are not in widespread use.

Further investigation of UTI

Alternative strategies for further investigation of UTI can be considered by their cost, invasiveness and test characteristics.[10] The potential yield of these investigations will be affected by the accuracy of the initial diagnosis of UTI.[10] It is therefore important that UTI is diagnosed accurately before further testing is undertaken. Follow-up testing from confirmed UTI may be used to localise infection; infections of the lower urinary tract (cystitis) where there is no renal involvement require no further investigation. Where renal involvement is suspected, a variety of imaging tests may be used to elucidate potential causes of infection and to quantify any renal damage present.

No single imaging investigation provides a full assessment of the whole urinary tract, and there is lack of agreement on the role and the optimum timing of each test in children of different age, gender and clinical presentation.[4]

Ultrasonography

Ultrasonography is commonly used as the preliminary investigation for children with confirmed UTI because of its widespread availability, relatively low cost and absence of side-effects.[22] Standard ultrasound techniques are non-invasive and generally preferred by patients and parents. Outcome is dependent on the skill of the examiner,[22] and techniques and equipment vary significantly;[23] therefore, these factors should be considered when evaluating the efficacy of ultrasonography. Preliminary ultrasound examination is generally used to provide a structural overview of the renal anatomy, and may be used to rule out hydronephrosis, abscess or calculus.[6] It is also used to detect malformations such as duplex kidneys.[6] Recently developed, contrast-enhanced cystosonography techniques have been used to evaluate VUR. These techniques involve the introduction of an inert, microbubble contrast material into the bladder by catheter and the use of ultrasound to follow filling and voiding. This method had the advantage over

micturating cystourethrography (MCUG) of not involving ionising radiation. However, examination time is longer than that for MCUG, the child still has to be catheterised and contrast media are expensive.[6]

Planar imaging

Planar imaging involves the injection of a contrast medium (usually iodine), either intravenously or through a catheter, after which the patient is X-rayed to provide an image of urinary tract/renal anatomy. There are several different forms of this type of imaging.

Intravenous urography (IVU)/intravenous pyelography (IVP)/excretory urography/kidney, ureter and bladder imaging (KUB)

This method has traditionally been used for imaging children with UTI, and was historically regarded as the reference standard for detection of renal scarring.[22] It has now been largely superseded by technetium-99m (^{99m}Tc)-dimercaptosuccinic acid (DMSA) scintigraphy (see below), but is still used in some areas, particularly for presurgical imaging.[23] It provides a precise anatomical image of the kidneys, ureter and bladder, and can readily identify some urinary tract abnormalities.[19] IVU requires higher doses of radiation than DMSA and carries a risk of reaction to contrast medium.[19]

Cystography

Cystography is similar to IVU, but it is only used to give an image of the bladder. In this technique, contrast medium is injected intravenously.

MCUG

This technique, called voiding cystourethrography (VCUG) in the USA, is used to detect grades of VUR. It also provides an assessment of the size and shape of the bladder, and a means of detecting posterior urethral anomalies in boys.[19] MCUG has the disadvantage of involving a relatively large dose of radiation and requiring catheterisation of the bladder, which can cause discomfort;[19] it is considered an invasive procedure and is distressing for some parents and almost all children over 1 year of age. There is a small risk, particularly in the very young, of introducing organisms during this procedure, thus causing a UTI.

Nuclear medicine

Nuclear medicine techniques involve the intravenous injection of a radioactive contrast medium and detection using an electronic gamma camera. Several imaging tests of this type are used in the evaluation of children with UTI.

DMSA scintigraphy/static renal scintigraphy/renal cortical scintigraphy

This technique provides information on renal structure and function, and on the presence or absence and extent of cortical scarring following UTI.[16] It is generally regarded as the reference standard technique for evaluation of renal scarring. The contrast medium is injected intravenously and is taken up by cells in the kidney. Imaging generally occurs 2–3 hours after injection to allow renal clearance of excess activity, and to maximise visualisation. In general, an integrated gamma camera and computer systems are used for the acquisition of planar images for either a fixed time interval or fixed total counts.[16] This procedure needs a degree of cooperation and some children require mild sedation. The timing of DMSA is important; if the examination is done at the time of the UTI then all defects seen do not necessarily represent permanent scars. The procedure is therefore generally conducted at least 6 weeks after the UTI.

Radionuclide cystography/isotope cystography

This technique represents an alternative method for the evaluation of reflux, which results in significantly less exposure to radiation in comparison with conventional radiographic techniques (MCUG). There are two methods of radionuclide cystography: direct and indirect. The direct method requires catheterisation of the bladder and instillation of radionuclide and fluid for maximum distension of the bladder, allowing imaging during filling and voiding, and after voiding. Indirect radionuclide cystography does not require bladder catheterisation, but requires intravenous injection of the radiopharmaceutical for the evaluation of renal function and urine drainage as well as the detection of VUR.[17] As this method provides renal imaging as well as assessment of reflux, it may have the potential to replace two separate examinations. The indirect method requires a cooperative, potty-trained child.

Other tests

Additional tests have been used to investigate UTI further, particularly for the localisation of infection. These include clinical history, numerous biochemical analytes [e.g. C-reactive protein (CRP), procalcitonin (PCT), β_2-macroglobulin (β^2M)], erythrocyte sedimentation rate (ESR) and immunofluorescent detection of bacteria.

Clinical applications of tests used in the further investigation of UTI

There are four main clinical applications, seen in the literature, of tests for the further investigation

of UTI: localisation of UTI, detection of reflux, prediction of scarring and detection of scarring. These are broadly encompassed by the term 'imaging investigations' used in phase 4 of *Figure 1*, although non-imaging can also be used to localise infection. *Table 1* summarises which tests have been evaluated and used for these different clinical applications.

TABLE 1 *Summary of tests evaluated and used for different clinical applications*

Aim	Main tests used or evaluated	Current reference standard
Localisation of UTI	Clinical features Laboratory-based Ultrasound DMSA	Acute DMSA
Detection of reflux	Ultrasound Cystosonography Radionuclide cystography MCUG	MCUG
Prediction of scarring	Clinical features Ultrasound MCUG DMSA	Follow-up DMSA
Detection of scarring	Ultrasound IVP Radionuclide cystography DMSA	Follow-up DMSA

Chapter 2

Research questions

The objectives of this review are:

1. to determine the diagnostic accuracy of tests (including clinical examination) and different methods of urine sampling 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 who manage these patients, which could be evaluated in future primary research.

Chapter 3
Review methods

An advisory panel was established. In addition to providing subject-specific input during the review, members of the panel were invited to offer comment on the protocol and draft report. Details of advisory panel members can be found in Appendix 1. The systematic review was undertaken in accordance with the Centre for Reviews and Dissemination (CRD) guidelines for undertaking systematic reviews[24] and published guidelines on the meta-analysis of diagnostic tests.[25,26]

Search strategy

A database of published and unpublished literature was assembled from systematic searches of electronic sources, handsearching and consultation with experts in the field. The database was built using the Endnote software package.

The following databases were searched for diagnostic test evaluations: MEDLINE (1966 to October 2002), PreMEDLINE (1966 to November 2002), BIOSIS (1985 to December 2002), Pascal (1973 to January 2003), LILACS (25 February 2003), The Cochrane Library (Issue 2002/4), Science Citation Index (1980 to January 2003), BL Inside Conferences (1993 to January 2003), SIGLE (1980 to June 2002), Dissertation Abstracts (1861–2002), NTIS (1970–2002), Greylit (28 February 2003), NHS EED (December 2002), EMBASE (1980 to November 2002), National Research Register (Issue 2003/1) and Controlled Clinical Trials (24 February 2003). A series of search strategies was explored then a strategy was chosen as capturing relevant records while excluding large numbers of irrelevant records. This strategy is shown in a version that will run in the Ovid interfaces of MEDLINE. The strategy was appropriately adapted to run on other databases with different interfaces and search options. This is shown at the start of Appendix 2; the full strategies and descriptions of all searches undertaken are also presented in Appendix 2.

A search of the Internet using Altavista was conducted on 28 February 2003 and a selection of the results was scanned for further studies. In addition, handsearches of the following key journals were performed for the period 1998–2003: *Pediatric Nephrology*, *Journal of Urology*, *Archives of Disease in Childhood*, *Pediatrics*, *BMJ*, *Journal of Pediatrics*, *Pediatric Radiology*, *Pediatric Infectious Disease Journal*, *Journal of Nuclear Medicine*, and *British Journal of Urology*. Reference lists of included papers were also scanned to identify further possibly relevant studies.

To identify economic evaluations, the Endnote database was searched using the following text words: cost or costs or economic or costly or costing or prices or price or pricing or expenditure or money or budget or preference. In addition, a search of NHS EED was undertaken using the following search terms: urinary(w)tract(w)infection$ OR UTI OR cystitis OR pyelonephritis. To identify information to populate the decision model, a sequence of search strategies was required. The approach was based on research undertaken for the HTA-funded project 'A review of guidelines for good practice in decision-analytic modelling in health technology assessment' (project reference 02/32/01). The strategies are given in full in Appendix 3. Because of the number of searches required, only MEDLINE and the publication 'Morbidity statistics from general practice' were searched.

Update searches of all databases were conducted in May 2004.

Some data used to populate the economic models were derived from additional searches outside the scope of the systematic review.

Inclusion/exclusion criteria

Two reviewers screened titles and abstracts for relevance independently, and any disagreements were resolved by consensus. Full papers of potentially relevant studies were obtained and assessed for inclusion by one reviewer and checked by a second. There were separate inclusion criteria for the three sections of the systematic review of effectiveness component of the project (objectives 1–3) and for the review of existing cost-effectiveness literature.

Diagnosis of UTI

Study design: randomised controlled trial (RCTs), controlled clinical trials (CCTs) or diagnostic cohort studies with at least 20 participants. Diagnostic case–control studies were not eligible for inclusion.

Population: children aged less than 5 years with suspected UTI. As very few studies were identified that were conducted exclusively in children aged less than 5 years, studies of children of all ages were also included as long as they included at least some children aged less than 5 years. Studies conducted in adults and children were only included if data for children were reported separately.

Index tests: any test that aimed to diagnose UTI. Studies that compared urine sampling methods were also eligible for inclusion. Tests not available in the UK were excluded, as were tests used specifically to screen for schistosomiasis.

Reference standard: culture or culture combined with other tests.

Outcome measures: studies had to report sufficient information to construct a 2×2 table. Studies that reported patient-centred outcomes were also eligible for inclusion.

Further investigation of UTI

Study design: RCTs, CCTs or diagnostic cohort studies with at least 20 participants. Diagnostic case–control studies were not eligible for inclusion.

Population: children aged less than 5 years, at least some of whom had to have confirmed UTI. Studies conducted exclusively in children with known urinary tract problems, for example with reflux, were excluded. As very few studies were identified that were conducted exclusively in children aged less than 5 years, studies of children of all ages were also included as long as they included at least some children aged less than 5 years. Studies conducted in adults and children were only included if data for children were reported separately.

Index tests: any investigation used to investigate UTIs further.

Reference standard: studies with any reported reference standard were eligible for inclusion.

Outcome measures: studies had to report sufficient information to construct a 2×2 table. Studies that reported patient-centred outcomes were also eligible for inclusion.

Effectiveness of follow-up

Study design: RCTs, CCTs, cohort studies, case–control studies or cross-sectional studies.

Population: children who had suspected or confirmed UTI when aged less than 5 years. As very few studies were identified that were conducted exclusively in children aged less than 5 years, studies of children of all ages were also included as long as they included at least some children aged less than 5 years. Studies conducted in adults and children were only included if data for children were reported separately.

Intervention: diagnostic testing or imaging evaluation.

Outcome measures: long-term follow-up relating to the incidence of renal disease, recurrent infection or any other reported outcome.

Economic evaluations

Study design: full economic evaluations in which both costs and health outcomes were estimated and at least two options compared.

Population: children who had suspected or confirmed UTI. Studies conducted in adults and children were only included if data for children were reported separately.

Intervention: diagnostic testing or imaging evaluation.

Outcome measures: costs from a health service or broader (e.g. societal) perspective; appropriate health-related outcomes as defined in the sections 'Diagnosis of UTI', 'Further investigation of UTI' and 'Effectiveness of follow-up', above.

Data extraction

Data extraction forms were developed using Microsoft Access. These were piloted on a small selection of studies. Separate forms were developed for diagnostic accuracy studies and controlled trials. Data extraction was performed by one reviewer and checked by a second. Foreign-language papers were extracted by one reviewer, accompanied by a speaker of that language, and the data were entered directly into the Access database. A second reviewer did not check foreign-language studies. The following information was extracted for all studies: study details [identifier, author, year, study design, language, objective, country, setting (primary or secondary care, and if

secondary care, whether it was in a teaching hospital)], participant details (number of children, number of boys and girls, mean age, age range and inclusion criteria/spectrum composition), study withdrawals and any reported adverse events relating to any of the tests performed. Data specific to the type of study were also extracted. Economic studies were summarised by the health economics team.

Diagnosis of UTI

Data were extracted on urine sampling methods, index test details (test evaluated, details of test performance, definition of a positive test result), reference standard details (reference standard, details of reference standard execution, definition of a positive reference standard result) and results (data to construct a 2×2 table).

Further investigation of UTI

Data were extracted on patient spectrum (confirmed UTI/mixed population with some UTI), aim of further investigation (localisation of UTI, to detect reflux or to detect/predict scarring) index test details (test evaluated, details of test performance, definition of a positive test result, time from infection to performance of index test), reference standard details (reference standard, details of reference standard execution, definition of a positive reference standard result, time from infection to reference standard execution) and results (data to construct a 2×2 table). Where studies provided $n \times n$ data on staging disease, for example vesicoureteral reflux, these were extracted separately. These data were also dichotomised and extracted as 2×2 table data. For studies of imaging techniques, data were extracted on both the number of false positives and the number of 'true false positives'. This distinction was made as some studies classified certain findings as positive that were not consistent with the aim of the study. For example, studies looking at ultrasound for the detection of scarring might classify a horseshoe kidney as positive when this was not positive for scarring. For such studies these findings were classified as negative; in this example, if the kidney was not positive for scarring it was classified as negative. For such studies, the original 2×2 data presented in the article were reported in brackets together with the recalculated data, in the results tables.

Effectiveness of follow-up

As only one study was identified, a narrative summary is presented. No formal data extraction was performed.

Economic evaluations

As only one study was identified, a narrative summary is presented.

Quality assessment

Quality assessment forms were developed using Microsoft Access for the different study designs included in the review. Quality assessment was carried out by one reviewer and checked by a second.

Diagnostic accuracy studies

Included diagnostic evaluation studies (for both diagnosis and further investigation of UTI) were assessed for methodological quality using the Quality Assessment of Diagnostic Studies (QUADAS) tool.[27,28] The QUADAS tool, together with details on how studies were scored, is provided in Appendix 4.

RCTs/CCTs

No full reports of RCTs/CCTs were identified. The only RCT of the effectiveness imaging in UTI was reported as an abstract.[29] The reviewers were unable to obtain additional data on this study. Therefore, no quality assessment of this study could be conducted.

Economic evaluations

The quality of the cost-effectiveness studies was assessed according to a checklist updated from that developed by Drummond and colleagues.[30] This checklist reflects the criteria for economic evaluation detailed in the methodological guidance developed by the National Institute for Health and Clinical Excellence (NICE).[31]

Statistical analysis

The analyses of data identified by the systematic review component of the project are described below. Data on diagnostic accuracy, clinical effectiveness and cost-effectiveness were analysed or summarised separately. The economic models were developed using additional data to that derived from the systematic review, and the explicit methods used in the modelling component of the project are described in Chapter 6.

Diagnosis/further investigation of UTI

Results were analysed according to test grouping: diagnosis of UTI (clinical, urine sampling, dipstick, microscopy, culture, other and test

combinations) and further investigation of UTI (localisation of UTI, detection of reflux, prediction of scarring, detection of scarring and imaging studies with multiple aims). Within these groups, tests were examined according to the specific tests or test combinations reported in the literature. Combinations of tests were analysed as test combinations, where appropriate.

For each test, or test combination, the range in sensitivity, specificity and likelihood ratios (of both positive and negative tests results), and diagnostic odds ratios (DORs) were calculated. These were presented in tables. To account for 0 cells in the 2×2 tables, 0.5 was added to every cell for all 2×2 tables, as recommended by Moses and co-workers.[36] Individual studies results were presented graphically using summary receiver operating characteristic (SROC) curves. These were estimated using the following equation:

$$Sensitivity = \frac{1}{1 + \frac{1}{e^{\frac{a}{1-b}} \times \left(\frac{1 - \text{Specificity}}{\text{Specificity}}\right)^{\frac{1+b}{1-b}}}}$$

a and b were calculated using the following regression equation:

$$D = a + bS$$
$$D = \{\text{logit (TPR)} - \text{logit (FPR)}\} = \log (\text{DOR})$$
$$S = \{\text{logit (TPR)} + \text{logit (FPR)}\}$$

$$\text{Logit (TPR)} = \ln(\text{TPR}/(1 - \text{TPR}))$$
$$\text{Logit (FPR)} = \ln(\text{FPR}/(1 - \text{FPR}))$$

where TPR is the true-positive rate and FPR is the false-positive rate.

This was estimated by regressing D against S, weighting according to sample size, for each study. β provides an estimate of the extent to which D is dependent on the threshold used. If β is 0 (when the line is symmetrical with respect to the line TPR = 1 – FPR), or not significantly different from 0, then the DOR is not affected by the threshold used. When this was the case the DOR was pooled according to standard methods for pooling odds ratios.[33] In such cases the following equation was used to calculate the SROC curves:

$$Sensitivity = \frac{1}{1 + \frac{1}{\text{DOR}_T \times \left(\frac{1 - \text{Specificity}}{\text{Specificity}}\right)}}$$

Likelihood ratios were selected as the measure of test performance for further analysis as physicians more easily interpret these measures than sensitivity and specificity. Heterogeneity of likelihood ratios was investigated using the Q statistic[34] and through visual examination of forest plots of study results.[35] 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 (minimum of ten studies), heterogeneity was further investigated using regression analysis. The SROC model,[36] as outlined above, was extended to include the covariates presented below.[37] A multivariate linear regression analysis was conducted, again weighted by sample size.

The following items were investigated as possible sources of heterogeneity, where data were available:

- factors affecting the reference standard: dependent on test grouping
- factors affecting the index test: dependent on test
- age: <2 years, <5 years, <12 years and <18 years
- country: grouped according to region
- QUADAS items.

Initially, univariate analysis was performed with items included individually in the model. Items that showed a significant association at the 5% significance level with D were investigated further using stepwise multivariate models. In this approach, all items found to be significant in the univariate models were entered into the multivariate model and then dropped in a stepwise fashion, with the least significant item dropped first. The final model was achieved when all items remaining in the model showed a significant association with D at the 5% level.

Effectiveness of follow-up

As only one study was identified, a narrative discussion is presented.

Economic evaluations

As only one study was identified, a narrative discussion is presented. In addition to a critical review of the cost-effectiveness literature relating to alternative diagnostic tests and imaging strategies, a decision-analytic model was

developed to estimate the cost-effectiveness of alternative tests and strategies. Full details of the modelling are detailed in Chapter 6.

Development of an algorithm for the diagnosis and further investigation of UTI in children

The results of the systematic review of effectiveness (objectives 1–3) were used to inform the consideration of preliminary diagnostic algorithms for the diagnosis and further investigation of UTI in children. The diagnosis and further investigation of UTI in children were considered separately. The economic modelling component of the project did not contribute directly to this process.

Chapter 4

Studies included in and studies excluded from the review

Table 40 in Appendix 5 lists all studies included in the review that addressed objective 1 of the review, 'to determine the diagnostic accuracy of tests (including clinical examination) and different methods of urine sampling for detecting UTI in children under 5 years of age'. Studies are reported along with all tests for which they provided data sets.[38–113]

Table 41 in Appendix 6 lists all studies included in the review that assessed the diagnostic accuracy of one or more tests used in the further investigation of UTI.[114–218] The clinical objective or target condition of the test(s) is recorded, along with the tests for which each study provided data sets.

Only one study was identified that addressed the clinical effectiveness of investigating confirmed UTI.[219]

Table 42 in Appendix 7 lists all studies for which full copies were obtained and which were subsequently excluded from the review. The reasons for exclusion are summarised. The inclusion criteria relevant to diagnostic accuracy studies are summarised in the table. Studies were excluded if they were not diagnostic cohort studies evaluating tests for the diagnosis or further investigation of UTI, did not include at least some children under 5 years of age, included fewer than 20 participants, did not report sufficient data for the construction of 2×2 tables, did not report a reference standard (further investigation) or used a reference standard that did not include culture (diagnosis). As many excluded studies failed to meet more than one of these criteria, inclusion criteria were assessed in a sequential manner (left to right in *Table 42*). All criteria, of those assessed, that were failed by an excluded study are highlighted in bold.

Chapter 5
Results of the review

Results of the literature searches

The literature searches identified over 10,000 references. These were screened for relevance and 1044 references were considered to be potentially relevant. Copies of four of these articles could not be obtained.[220–223] A Bulgarian article appeared to meet relevance criteria, but as no translator was available this paper could not be assessed for inclusion.[224] *Figure 2* shows the flow of studies through the review process and the number of studies excluded according to each of the inclusion criteria. Appendix 7 summarises the studies excluded from the review.

A total of 187 studies met all of the inclusion criteria; 80 examined the diagnostic accuracy of tests for UTI, 106 examined the diagnostic accuracy of tests used in the further investigation of UTI, and one study examined the effectiveness of follow-up. Four studies that met the inclusion criteria based on their English abstracts could not be extracted as they were published in languages for which translators could not be found. Three that met the inclusion criteria for UTI diagnosis were in Czech,[225] Croatian[226] and Danish.[227] One study that met the inclusion criteria for the further investigation of UTI was in Finnish.[228] Thus, 77 studies were included for the diagnosis of UTI, 105 studies for the further investigation of UTI and one for the effectiveness of follow-up. Studies on the diagnosis of UTI included 272 test evaluations and those on the further investigation included 215 test evaluations. Thus, a total of 487 test evaluations was included in the review.

In total, 26 non-English language papers were included in this review: four French,[73,150,175,229] four German,[100,115,181,195] four Italian,[55,107,119,122] one Polish[166] and 13 Spanish.[48–50,60–62,74,88,89,97,144,146,212]

Where insufficient details were reported, authors were contacted to provide further information. For example, authors were contacted if the study was published as an abstract, or if it appeared that 2×2 table data should be available for the study, but were not extractable from the published report. In total, 39 authors were contacted requesting clarification or further details of data reported in published articles or abstracts, or details of studies entered on research registers. Thirteen replies were received, of which seven provided additional data for this review.

Study quality is discussed in detail in the next section. Detailed tables of study quality for each study are included in this section. Study quality is also referred to throughout the results section, but readers requiring more information on the quality of an individual study should refer to the next section.

Quality assessment

For quality assessment, studies were divided into three groups rather than the two used in other sections of the review. The grouping of 'diagnosis of UTI' was retained, and the section on further investigation of UTI was split into 'localisation of UTI' and 'further investigation of UTI'. This was because the studies of localisation of UTI included clinical and laboratory tests as well as imaging tests, while the other studies on further investigation of UTI were almost exclusively studies of imaging evaluations. It was felt that the quality issues might differ because of these differences in tests.

Tests for the diagnosis of UTI

The 77 studies included in this category were of reasonable quality (*Table 2*). The median number of the 14 items included in the QUADAS tool fulfilled by these studies was 8 (range 5–13). *Figure 3* shows the number of quality items fulfilled by these studies. Over 80% of studies fulfilled the criteria for avoidance of incorporation bias, differential and partial verification bias, disease progression bias, and the use of an appropriate reference standard. Less than 40% of studies included an appropriate spectrum of patients and only just over 50% provided sufficient details on how children were selected for inclusion in the study. Almost 80% of studies reported sufficient details on how the test was performed to permit replication of the test, while just under 60% of studies provided sufficient details of reference standard execution. Studies failed to report sufficient details on clinical review bias, diagnostic

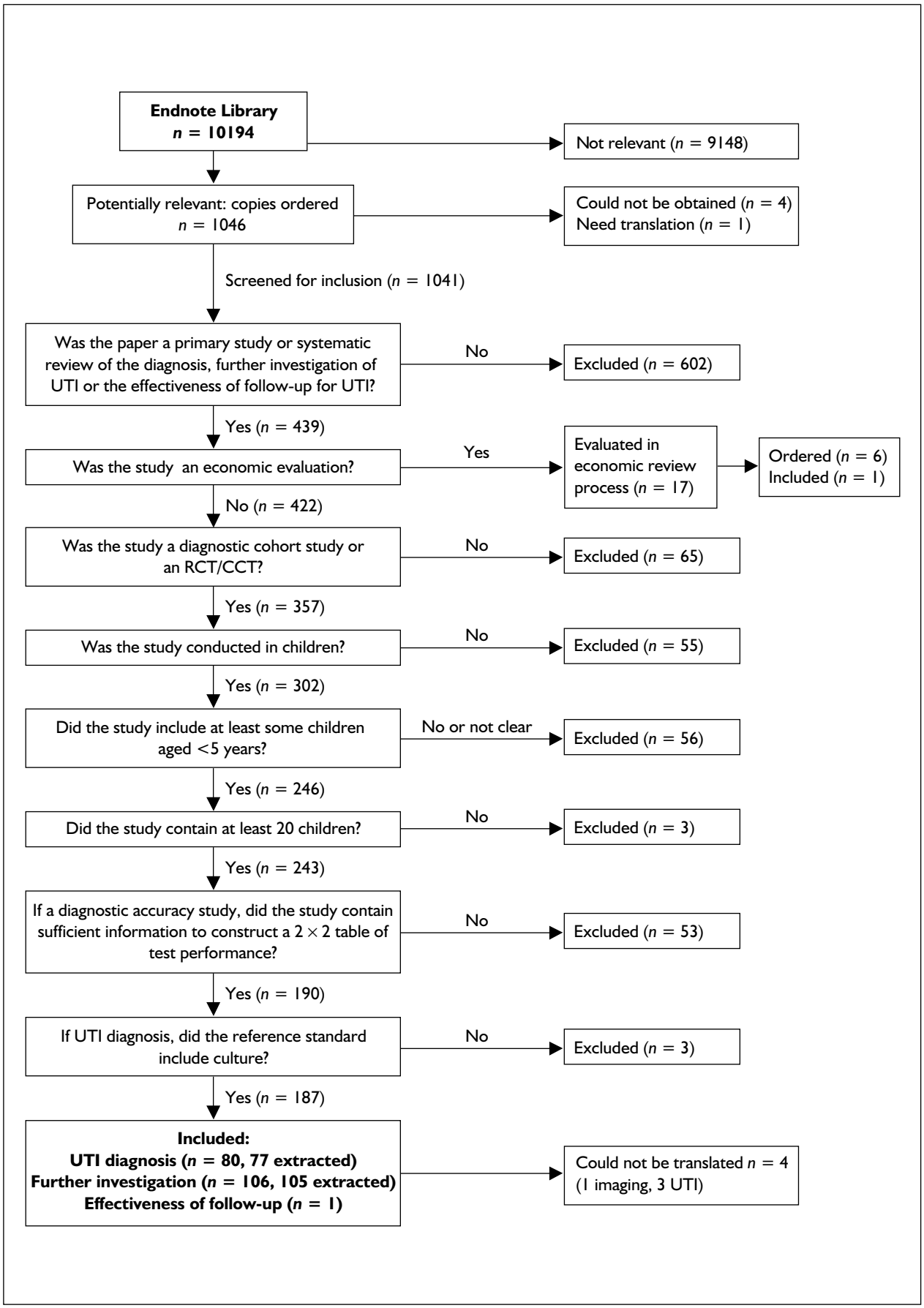

FIGURE 2 *Flowchart of studies through the review process. SR, systematic review.*

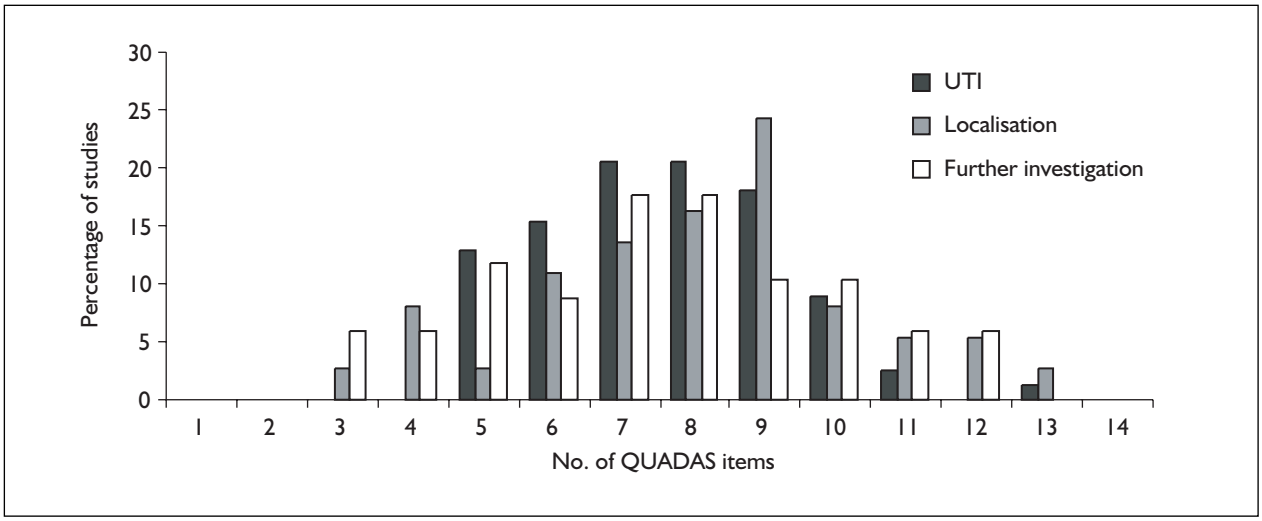

FIGURE 3 *Number of quality items fulfilled by studies*

review bias and test review bias to judge whether these were avoided. Study withdrawals and handling of uninterpretable results were also poorly reported. *Figure 4* illustrates the number of studies that answered 'yes', 'no' and 'not stated' to each of the 14 QUADAS items.

Tests for localisation of UTI

The 37 studies included in this section (*Table 3*) showed a similar pattern of quality to those included for the diagnosis of UTI. The median number of QUADAS items fulfilled by these studies was 8, with a range of 3–13. The distribution of the number of QUADAS items fulfilled is shown in *Figure 3*. Incorporation bias and differential verification bias were avoided in all studies, and all but one study also reported that all patients had received verification by the reference standard. The time delay between the index test and reference standard was more of a problem with these studies than with those for the diagnosis of UTI: just over 60% of studies reported that the time between the index test and reference standard was short enough that the disease state was unlikely to have changed between tests. The use of an appropriate reference standard was also a problem in some of these studies. Just over 60% used an appropriate reference standard, around 10% did not use an appropriate reference standard and in the remainder of studies it was not clear whether the reference standard used was appropriate. Spectrum composition and reporting of details of how children were selected for inclusion in the study was better in these studies than in the studies of the diagnosis of UTI. Over 50% of studies used an appropriate spectrum of patients and around 70% provided sufficient details on how children were selected for the study. Only around half of studies provided sufficient details of how the index test and reference standard were performed to allow replication of these tests. Almost 40% of studies provided information indicating that test and diagnostic review bias had been avoided, in the remainder of studies this information was not reported. Almost 90% of studies did not provide details on whether clinical information was available when test results were interpreted. Few details were provided on how uninterpretable results were handled, with around 70% of studies failing to provide this information. Study withdrawals were also poorly reported, with almost 50% of studies not reporting any details of study withdrawals and over 20% of studies not accounting for withdrawals.

Further investigation of UTI

The quality of the 68 studies of the further investigation of UTI (*Table 4*) was similar to that of studies for the diagnosis and localisation of UTI. The median quality score was 7.5 (range 3–12). The distribution of the number of quality items fulfilled by these studies is shown in *Figure 4*. None of the quality criteria was fulfilled by all studies. Around half of studies reported that diagnostic and test review bias had been avoided; the remaining studies did not report whether the index test and reference standard were interpreted blind to the results of the other test. Studies also provided very little information on whether clinical information was available when test results were interpreted, how uninterpretable results were handled, and whether there were any withdrawals from the study and if so whether all withdrawals were accounted for. Four items were fulfilled by more than 60% of studies: the avoidance of

TABLE 2 *Quality assessment results for studies of the diagnosis of UTI*

Study	Appropriate spectrum composition	Selection criteria reported	Appropriate reference standard	Disease progression bias avoided	Partial verification bias avoided	Differential verification bias avoided	Incorporation bias avoided	Test execution details reported	Reference execution details reported	Test review bias avoided	Diagnostic review bias avoided	Clinical review bias avoided	Uninterpretable results reported	Withdrawals accounted for
Ahmad, 1991[38]	N	N	Y	NS	Y	Y	Y	Y	Y	NS	NS	NS	Y	NS
Anad, 2001[39]	N	Y	Y	Y	Y	Y	Y	Y	N	NS	NS	NS	N	N
Armengol, 2001[230]	N	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS
Armengol, 2000[40]	NS	N	Y	Y	Y	Y	Y	N	N	NS	NS	NS	NS	NS
Aronson, 1973[41]	N	N	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS
Arslan, 2002[42]	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS
Bachur, 2001[43]	N	Y	Y	NS	NS	Y	Y	Y	N	NS	NS	NS	Y	Y
Baum, 1972[44]	NS	N	Y	Y	Y	Y	Y	N	N	NS	NS	NS	NS	NS
Boreland, 1986[45]	NS	N	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS
Braude, 1967[46]	Y	N	Y	NS	Y	Y	Y	N	Y	NS	NS	NS	Y	Y
Bulloch, 2000[47]	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	Y	NS	NS
Caballero, 2001[48]	NS	N	Y	Y	Y	Y	Y	Y	N	NS	NS	NS	Y	Y
Cervilla, 2001[49]	N	Y	Y	Y	N	Y	Y	N	N	NS	NS	NS	NS	NS
Cid, 1992[50]	Y	Y	Y	Y	Y	Y	Y	Y	N	NS	NS	NS	NS	NS
Cohen, 1997[51]	Y	Y	Y	NS	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS
Craver, 1997[52]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS
Dayan, 2002[53]	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	N
Dayan, 2000[54]	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	Y	Y
Demi, 1993[55]	N	N	Y	NS	N	Y	Y	Y	N	Y	N	NS	NS	NS
Dosa, 1973[56]	N	N	Y	Y	Y	N	Y	Y	N	NS	NS	NS	NS	NS
Farrell, 2002[57]	N	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	N
Feasey, 1999[58]	N	N	Y	Y	Y	Y	Y	N	N	NS	NS	NS	Y	Y
Fennell, 1977[59]	N	N	NS	NS	Y	N	Y	Y	Y	Y	NS	NS	Y	N
Fernandez, 2000[60]	Y	N	Y	Y	Y	N	Y	N	Y	NS	NS	NS	NS	NS
Fernandez, 1996[61]	N	Y	Y	Y	Y	Y	Y	N	N	NS	NS	NS	Y	Y
Giraldez, 1998[62]	Y	Y	Y	Y	Y	Y	Y	N	N	NS	NS	NS	NS	NS
Godard, 1979[63]	N	N	NS	N	Y	N	Y	N	N	Y	NS	NS	Y	Y
Gorelick, 2000[64]	N	Y	Y	Y	Y	Y	Y	Y	N	NS	NS	NS	NS	N
Hardy, 1976[65]	N	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	Y	NS
Hiraoka, 1995[66]	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS
Hoberman, 1996[67]	N	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS

continued

TABLE 2 *Quality assessment results for studies of the diagnosis of UTI (cont'd)*

Study	Appropriate spectrum composition	Selection criteria reported	Appropriate reference standard	Disease progression bias avoided	Partial verification bias avoided	Differential verification bias avoided	Incorporation bias avoided	Test execution details reported	Reference execution details reported	Test review bias avoided	Diagnostic review bias avoided	Clinical review bias avoided	Uninterpretable results reported	Withdrawals accounted for
Hoberman, 1994[69]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	Y	Y
Hoberman, 1993[68]	NS	N	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS
Holland, 1968[70]	NS	N	Y	Y	Y	Y	Y	N	N	NS	NS	NS	NS	NS
Kohler, 1970[71]	N	N	Y	Y	Y	N	Y	Y	N	NS	NS	NS	NS	NS
Kunin, 1977[72]	N	Y	Y	Y	Y	Y	Y	Y	N	NS	NS	NS	NS	NS
Labbe, 1982[73]	Y	N	Y	NS	Y	Y	Y	Y	Y	N	N	Y	N	N
Lagos, 1994[74]	N	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS
Lejeune, 1991[75]	NS	N	Y	Y	Y	Y	Y	Y	N	NS	NS	NS	NS	NS
Lin, 2000[77]	N	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	Y
Lin, 2000[76]	N	Y	Y	Y	Y	Y	Y	N	Y	NS	NS	NS	NS	Y
Liptak, 1993[78]	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	Y	NS
Littlewood, 1977[79]	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS
Lockhart, 1995[80]	Y	Y	Y	Y	Y	Y	Y	N	Y	NS	NS	NS	NS	Y
Lohr, 1993[81]	N	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	Y	NS	NS
Manson, 1985[82]	N	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	Y	Y
Marret, 1995[83]	NS	N	Y	Y	Y	N	Y	Y	N	NS	NS	NS	NS	NS
Marsik, 1986[84]	NS	N	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS
Matthai, 1995[85]	Y	N	Y	Y	Y	Y	Y	N	Y	NS	NS	NS	NS	NS
Mongeau, 1972[86]	N	N	Y	Y	Y	Y	Y	Y	N	NS	NS	NS	NS	NS
Morton, 1982[87]	Y	N	Y	Y	Y	Y	Y	N	Y	NS	NS	NS	NS	N
Navarrete, 1966[88]	N	N	Y	Y	Y	Y	Y	N	N	NS	NS	NS	NS	NS
Ordonez, 1994[89]	Y	N	Y	Y	Y	Y	Y	Y	N	NS	NS	NS	NS	NS
Palmer, 1997[90]	N	N	Y	Y	Y	Y	Y	Y	N	NS	NS	NS	NS	NS
Parmington, 1989[91]	Y	Y	Y	N	Y	Y	Y	N	N	NS	NS	NS	NS	NS
Pryles, 1965[92]	NS	N	Y	Y	Y	Y	Y	Y	N	NS	NS	NS	NS	NS
Purwar, 1972[93]	NS	N	Y	N	Y	Y	Y	Y	N	NS	NS	NS	NS	NS
Pylkkanen, 1979[94]	Y	N	Y	Y	Y	Y	Y	N	N	NS	NS	NS	NS	NS
Ramage, 1999[95]	N	N	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS
Rich, 1976[96]	N	N	Y	N	Y	N	Y	Y	Y	NS	NS	NS	Y	Y
Santos, 1982[97]	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS
Saxena, 1975[98]	Y	N	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS

continued

TABLE 2 *Quality assessment results for studies of the diagnosis of UTI (cont'd)*

Study	Appropriate spectrum composition	Selection criteria reported	Appropriate reference standard	Disease progression bias avoided	Partial verification bias avoided	Differential verification bias avoided	Incorporation bias avoided	Test execution details reported	Reference execution details reported	Test review bias avoided	Diagnostic review bias avoided	Clinical review bias avoided	Uninterpretable results reported	Withdrawals accounted for
Schersten, 1968[99]	N	N	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	Y	N
Schreiter, 1971[100]	N	N	Y	Y	Y	Y	Y	N	Y	NS	NS	NS	NS	NS
Sharief, 1998[101]	N	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	NS	Y	NS
Shaw, 1998[102]	Y	N	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	N
Shaw, 1991[103]	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS
Struthers, 2003[104]	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	NS	NS
Tahirovic, 1988[105]	NS	N	Y	Y	Y	Y	Y	Y	N	NS	NS	NS	NS	NS
Todd, 1974[106]	N	N	NS	NS	Y	Y	NS	Y	N	Y	NS	Y	NS	NS
Vangone, 1985[107]	NS	N	Y	Y	Y	Y	Y	Y	N	Y	NS	NS	NS	NS
Vickers, 1991[108]	N	Y	Y	Y	Y	Y	N	Y	Y	NS	NS	NS	Y	Y
Waisman, 1999[109]	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS
Wammanda, 2000[110]	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS
Weinberg, 1991[111]	N	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	Y	NS	NS
Wiggelinkhuizen, 1988[112]	N	N	Y	NS	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS
Woodward, 1993[113]	N	Y	Y	NS	Y	Y	Y	Y	N	NS	NS	NS	Y	NS

N, no; NS, not stated; Y, yes.

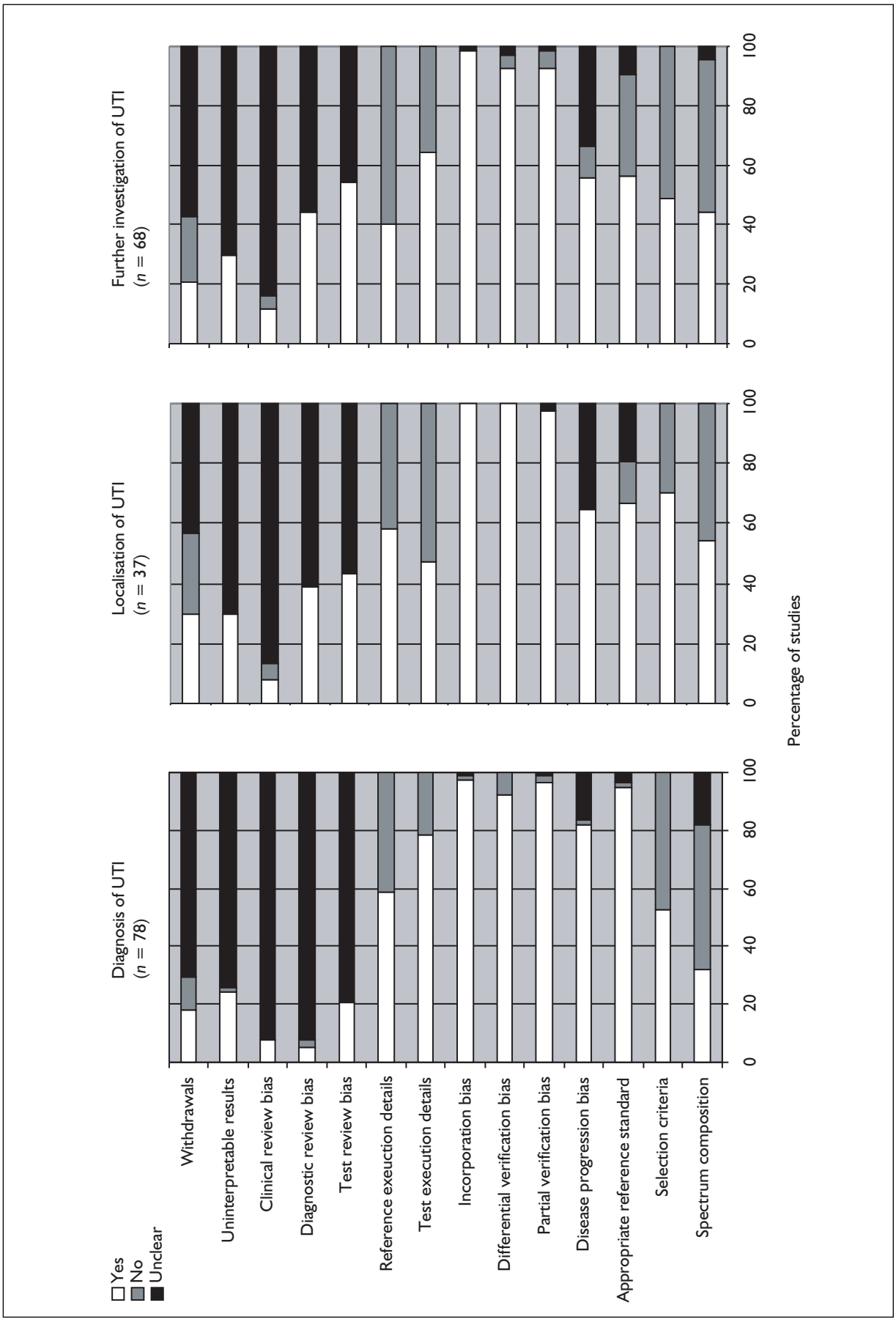

FIGURE 4 *Proportion of studies rated as yes, no or not clear for each of the QUADAS items, separately for diagnosis, localisation and further investigation of UTI*

TABLE 3 *Quality assessment results for studies of the localisation of UTI*

Study	Appropriate spectrum composition	Selection criteria reported	Appropriate reference standard	Disease progression bias avoided	Partial verification bias avoided	Differential verification bias avoided	Incorporation bias avoided	Test execution details reported	Reference execution details reported	Test review bias avoided	Diagnostic review bias avoided	Clinical review bias avoided	Uninterpretable results reported	Withdrawals accounted for
Andrich, 1992[116]	N	N	NS	NS	Y	Y	Y	N	N	NS	NS	Y	NS	N
Barnett, 1978[118]	N	N	N	NS	NS	Y	Y	Y	N	NS	NS	NS	NS	N
Benador, 1994[121]	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS
Biggi, 2001[125]	Y	Y	NS	NS	Y	Y	Y	N	Y	Y	NS	NS	NS	N
Bircan, 1995[126]	Y	N	Y	Y	Y	Y	Y	N	Y	Y	Y	NS	Y	Y
Boudailliez, 1998[127]	N	Y	Y	Y	Y	Y	Y	N	N	NS	NS	NS	NS	NS
Buyan, 1993[129]	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS
Capa Kaya, 2001[130]	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS
Dacher, 1996[134]	N	Y	N	Y	Y	Y	Y	Y	Y	NS	Y	N	NS	NS
Everaert, 1998[140]	N	Y	Y	NS	Y	Y	Y	Y	Y	NS	NS	NS	Y	NS
Fretzayas, 2000[143]	Y	Y	Y	Y	Y	Y	Y	N	Y	NS	Y	NS	NS	NS
Gervaix, 2001[145]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	Y	NS
Girona, 1995[146]	Y	Y	Y	NS	Y	Y	Y	N	Y	NS	NS	NS	NS	N
Guermazi, 1993[148]	N	N	Y	Y	Y	Y	Y	N	N	NS	NS	NS	NS	N
Hajjar, 2002[150]	N	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS
Hellerstein, 1978[153]	N	N	N	Y	Y	Y	Y	Y	Y	Y	NS	NS	Y	NS
Hitzel, 2002[155]	N	Y	NS	NS	Y	Y	Y	Y	Y	Y	Y	NS	NS	Y
Hitzel, 2000[177]	N	Y	NS	NS	Y	Y	Y	N	N	Y	Y	NS	NS	Y
Ilyas, 2002[156]	N	Y	Y	Y	Y	Y	Y	N	Y	NS	NS	NS	Y	Y
Jakobsson, 1992[157]	N	Y	Y	Y	Y	Y	Y	Y for US, N for MCUG	N	NS	NS	NS	NS	N
Jantausch, 1994[158]	Y	Y	Y	NS	Y	Y	Y	Y	N	NS	NS	NS	Y	Y
Jequier, 1998[160]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	Y	Y
Krzemien, 2002[166]	Y	Y	Y	Y	Y	Y	Y	N	N	NS	NS	NS	NS	NS
La Cava, 1990[200]	Y	N	NS	Y	Y	Y	Y	N	N	NS	Y	Y	Y	NS
Landau, 1994[168]	N	Y	Y	NS	Y	Y	Y	Y	Y	NS	NS	Y	NS	Y
Landau, 1994[167]	N	N	Y	NS	Y	Y	Y	N	N	NS	NS	NS	NS	NS
Lavocat, 1997[169]	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	Y
Lonergan, 1998[173]	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	NS	Y	Y
Montplaisir, 1981[182]	Y	N	Y	NS	Y	Y	Y	Y	N	Y	NS	NS	NS	NS

continued

TABLE 3 *Quality assessment results for studies of the localisation of UTI (cont'd)*

Study	Appropriate spectrum composition	Selection criteria reported	Appropriate reference standard	Disease progression bias avoided	Partial verification bias avoided	Differential verification bias avoided	Incorporation bias avoided	Test execution details reported	Reference execution details reported	Test review bias avoided	Diagnostic review bias avoided	Clinical review bias avoided	Uninterpretable results reported	Withdrawals accounted for
Morin, 1999[183]	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y for DMSA, NS for MCUG	NS	NS	NS
Pylkkanen, 1978[190]	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	N
Sfakianakis, 1989[201]	Y	N	NS	NS	Y	Y	Y	N	N	NS	NS	NS	NS	N
Smolkin, 2002[204]	Y	Y	NS	Y	Y	Y	Y	N	N	Y	NS	NS	Y	Y
Sreenarasimhalah, 1995[205]	Y	Y	Y	Y	Y	Y	Y	N	Y	NS	NS	NS	NS	N
Stokland, 1996[208]	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	NS	Y	Y
Traisman, 1986[211]	N	Y	N	NS	Y	Y	Y	N	N	Y	Y	N	NS	N
Verboven, 1990[215]	Y	Y	Y for DMSA, N for clinical	Y	Y	Y	Y	N	Y for clinical, N for DMSA	NS	NS	NS	NS	N

US, ultrasound.

TABLE 4 *Quality assessment results for studies for the further investigation of UTI*

Study	Appropriate spectrum composition	Selection criteria reported	Appropriate reference standard	Disease progression bias avoided	Partial verification bias avoided	Differential verification bias avoided	Incorporation bias avoided	Test execution details reported	Reference execution details reported	Test review bias avoided	Diagnostic review bias avoided	Clinical review bias avoided	Uninterpretable results reported	Withdrawals accounted for
Alon, 1986[114]	Y	Y	N	Y	Y	Y	Y	N	N	Y	Y	N	Y	NS
Alzen, 1994[115]	N	N	Y	Y	Y	Y	Y	N	N	NS	NS	NS	NS	NS
Bagni, 1997[117]	Y	N	Y	Y	Y	Y	NS	Y	Y	NS	NS	NS	NS	N
Baronciani, 1986[119]	Y	N	Y	NS	Y	Y	Y	Y	N	Y	Y	Y	NS	N
Barry, 1998[120]	NS	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	Y	Y
Benigno, 1986[122]	Y	N	Y	Y	Y	Y	Y	N	N	Y	NS	NS	NS	NS
Bergius, 1990[123]	N	N	Y	Y	Y	Y	Y	Y	N	NS	NS	NS	NS	NS
Berrocal, 2001[124]	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	Y	Y
Bower, 1985[128]	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y
Cavanagh, 1983[131]	Y	Y	N	Y	Y	Y	Y	N	N	NS	Y	NS	NS	NS
Chan, 1999[132]	N	N	Y	NS	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS
Clarke, 1990[133]	Y	N	N	NS	Y	Y	Y	N	N	NS	NS	NS	NS	NS
De Sadeleer, 1994[135]	N	N	Y for MCUG, N for DMSA	Y	Y	Y	Y	Y	Y	Y	NS	Y	NS	NS
Ditchfield, 1994[136]	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	NS	NS	NS
Drachman, 1984[137]	Y	Y	N	Y	Y	Y	Y	N	N	NS	NS	NS	NS	NS
Elison, 1992[138]	NS	N	N	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS
Evans, 1999[139]	N	Y	Y	Y	Y	Y	Y	Y	N	NS	NS	NS	NS	NS
Farnsworth, 1991[141]	N	N	N	N	Y	Y	Y	Y	Y	NS	NS	NS	NS	N
Foresman, 2001[142]	N	Y	Y	NS	N	Y	Y	Y	N	Y	NS	NS	NS	N
Frutos, 2000[144]	N	N	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	Y	Y
Gordon, 1992[147]	Y	N	Y	NS	Y	Y	Y	Y	Y	Y	Y	NS	Y	Y
Haberlick, 1997[149]	N	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	NS	NS	Y
Hanbury, 1989[151]	N	N	N	N	N	Y	Y	N	N	NS	NS	NS	NS	Y
Hedman, 1978[152]	Y	N	Y	Y	Y	Y	Y	N	N	NS	NS	NS	NS	NS
Hellstrom, 1989[154]	Y	Y	NS	NS	Y	NS	Y	Y	Y	NS	Y	NS	Y	Y
Jequier, 1985[159]	Y	N	N	Y	Y	N	Y	N	N	Y	NS	NS	NS	N
Johnson, 1985[161]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	N
Johnson, 1990[162]	Y	N	N	NS	Y	N	Y	Y	N	NS	NS	NS	NS	NS
Kenda, 1989[163]	Y	N	NS	NS	Y	Y	Y	N	N	NS	NS	NS	NS	NS

continued

TABLE 4 *Quality assessment results for studies for the further investigation of UTI (cont'd)*

Study	Appropriate spectrum composition	Selection criteria reported	Appropriate reference standard	Disease progression bias avoided	Partial verification bias avoided	Differential verification bias avoided	Incorporation bias avoided	Test execution details reported	Reference execution details reported	Test review bias avoided	Diagnostic review bias avoided	Clinical review bias avoided	Uninterpretable results reported	Withdrawals accounted for
Kenda, 2000[164]	NS	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS
Kessler, 1982[165]	N	N	Y	NS	Y	Y	Y	Y	N	Y	Y	NS	NS	NS
Leonidas, 1985[170]	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	NS	Y	Y
LeQuesne, 1986[171]	N	N	NS	NS	Y	Y	Y	N	N	Y	Y	NS	NS	NS
Lindsell, 1986[172]	N	N	N	Y	Y	Y	Y	Y	Y	N	Y	NS	NS	NS
MacKenzie, 1994[174]	Y	Y	N	Y	Y	Y	Y	Y	Y	NS	Y	NS	NS	N
Mage, 1989[175]	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	NS	Y	NS
Mahant, 2002[176]	N	N	Y	NS	Y	Y	Y	Y	Y	NS	Y	NS	NS	Y
McLorie, 1989[178]	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	Y	NS
Mentzel, 2002[179]	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	Y	NS
Merrick, 1980[180]	N	N	N	NS	Y	Y	Y	N	N	NS	NS	NS	NS	NS
Misselwitz, 1971[181]	N	N	N	NS	Y	Y	Y	Y	N	NS	NS	NS	NS	NS
Mucci, 1994[184]	N	N	NS	NS	Y	Y	Y	Y	N	NS	NS	NS	NS	NS
Muensterer, 2002[185]	N	N	Y	Y	Y	Y	Y	Y	N	NS	NS	NS	Y	NS
Oostenbrink, 2000[186]	Y	Y	Y	NS	Y	Y	Y	Y	N	NS	NS	NS	NS	NS
Piaggio, 2003[187]	N	N	Y	NS	Y	Y	Y	Y	N	NS	NS	NS	Y	NS
Pickworth, 1992[188]	N	N	Y for DMSA, NA for MAG3	N	NS	Y	Y	N	N	NS	NS	Y	NS	N
Piepsz, 1992[189]	N	Y	N	NS	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS
Radmayr, 2002[191]	N	Y	NS	Y	Y	Y	Y	Y	N	Y	Y	NS	Y	Y
Redman, 1984[192]	N	N	N	NS	Y	Y	Y	Y	N	NS	NS	NS	NS	NS
Rehling, 1989[193]	Y	N	N	Y	Y	Y	Y	Y	N	Y	Y	NS	Y	N
Rickwood, 1992[194]	Y	Y	NS	Y	Y	N	Y	Y	N	Y	Y	NS	Y	NS
Rohden, 1995[195]	N	N	Y	NS	Y	Y	Y	Y	N	NS	NS	NS	NS	NS
Rossleigh, 1990[196]	Y	N	N	N	Y	Y	Y	Y	N	NS	NS	NS	NS	NS
Salih, 1994[197]	N	N	Y	Y	Y	Y	Y	Y	N	Y	Y	NS	NS	NS
Scherz, 1994[198]	N	N	Y	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	N
Schneider, 1984[199]	N	N	Y	NS	Y	Y	Y	Y	N	NS	NS	NS	NS	NS
Siamplis, 1996[202]	N	Y	Y	Y	Y	Y	Y	Y	N	NS	NS	NS	NS	Y
Smellie, 1995[203]	Y	N	Y	Y	Y	Y	Y	N	N	NS	NS	NS	NS	NS

continued

TABLE 4 *Quality assessment results for studies for the further investigation of UTI (cont'd)*

Study	Appropriate spectrum composition	Selection criteria reported	Appropriate reference standard	Disease progression bias avoided	Partial verification bias avoided	Differential verification bias avoided	Incorporation bias avoided	Test execution details reported	Reference execution details reported	Test review bias avoided	Diagnostic review bias avoided	Clinical review bias avoided	Uninterpretable results reported	Withdrawals accounted for
Stokland, 1998[209]	Y	Y	Y	N	Y	Y	Y	Y for DMSA, no for other tests	Y	Y	Y	NS	Y	N
Stokland, 1996[207]	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N
Stokland, 1994[206]	N	N	N	N	Y	Y	Y	Y	N	Y	Y	NS	Y	NS
Tan, 1988[210]	Y	Y	Y	NS	N	NS	Y	Y	N	Y	NS	NS	NS	N
Trave, 1997[212]	Y	N	N for DMSA, Y for MCUG	Y	Y	Y	Y	N	Y for DMSA, N for MCUG	NS	NS	NS	NS	N
Valentini, 2001[213]	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NS	Y	Y
Verber, 1988[214]	Y	N	Y for MCUG, N for IVU	NS	N	Y	Y	Y	Y	NS	NS	NS	NS	N
Volti, 1991[216]	Y	Y	N	Y	Y	Y	Y	Y	Y	NS	NS	NS	NS	NS
Whitear, 1990[217]	N	Y	N	NS	Y	Y	Y	N	N	Y	Y	Y	NS	NS
Wujanto, 1987[218]	N	Y	N	NS	Y	Y	Y	Y	N	Y	Y	NS	Y	NS

MAG3, mercaptoacetyltriglycine.

incorporation bias, differential verification bias, partial verification bias and the provision of sufficient details of test execution. Fewer studies, around 40%, provided appropriate details of how the reference standard was performed. Less than 50% of studies included an appropriate spectrum of patients or provided details on how children were selected for inclusion in the study, and less than 60% of studies used an appropriate reference standard. The possibility of disease progression bias was also a problem in some of these studies, with less than 60% reporting a sufficiently short interval between the index test and reference standard to make it unlikely that the disease state would have changed between tests.

Accuracy of tests used to diagnose UTI

A total of 77 studies that evaluated the accuracy of tests for the diagnosis of UTI met the inclusion criteria. These studies included a total of 272 test evaluations.

Clinical tests

Six studies looked at the accuracy of various clinical features for the diagnosis of UTI[47,49,64,74,78,104] (*Table 5*). Only one of these studies included an appropriate spectrum of patients. The reporting of blinding of the investigators to other test results and clinical data (avoidance of review bias) was poor, and only half of the studies adequately described the conduct of the index test and the reference standard. Studies investigating clinical features can be divided into two groups: those that aim to identify children who may have a UTI and who may then go on to receive tests to diagnose the UTI, and those that aim to diagnose the UTI.

Two studies aimed to identify children who may have a UTI. One study looked at the diagnostic accuracy of a temperature of over 38.1 °C for the diagnosis of UTI. Sensitivity was less than 50% and specificity was only 56%.[49] The children included in this study were selected on the basis of having either a temperature of over 38.1 °C or other symptoms of UTI. A second study looked at a combination of clinical symptoms and included children aged less than 2 years with fever of unknown source.[64] The clinical features investigated were: age less than 12 months old, white race, temperature of at least 39 °C, fever for 2 days or more and absence of another source of fever on examination. The presence of any two of these five clinical features or symptoms was used to define a positive result. This study had a sensitivity of 95% and a specificity of 31%, suggesting that this combination of clinical symptoms may be useful for ruling out disease in children aged less than 2 years with fever of unknown source. Neither of these studies included an appropriate spectrum of patients or provided details on blinding, uninterpretable results or study withdrawals. Verification bias may have been a problem in the study of temperature.[49]

Two studies assessed urine clarity: UTI was stated to be present if the urine was cloudy. Both studies reported a reasonable performance for this test, with sensitivities of around 90% and specificities of 82% and 72%. It is difficult to generalise from these results as neither study included an appropriate spectrum of patients. However, both studies scored well on other quality criteria. One study assessed the diagnostic accuracy of a urine odour questionnaire completed by parents; both sensitivity and specificity were less than 50%.[104] This study was of good quality and included an appropriate patient spectrum. An additional study looked at the presence of any of the following symptoms for the diagnosis of UTI: urinary odour and cloudiness, fever, malaise, haematuria and pain.[78] This study was performed in a high-risk group consisting of children with neurogenic bladder and spinal cord injury.[78] The results are therefore unlikely to be generalisable to children presenting in GP surgeries. This study reported poor accuracy, with a sensitivity of 56% and a specificity of 76%.

Summary

Very few different clinical tests were investigated in the included studies. The only test for which there was some evidence of reasonable accuracy was urine clarity. All other tests investigated showed poor diagnostic performance.

Urine sampling

Twelve studies, with a total of 16 different test evaluations, compared the diagnostic accuracy of different methods of urine sampling[38,41,46,51,54,57,58,61,65,87,94,95] (*Table 6*). These studies compared the results of culture from urine obtained by different sampling methods. Only around half of these studies included an appropriate spectrum of patients and provided an adequate description of patient selection. No study reported whether or not those interpreting test results were blinded to the sampling method and clinical data (avoidance of review bias). Adequate description of the index test and reference standard was a problem in around 25% of studies.

TABLE 5 *Results of studies of clinical features*

Study	Test	Definition of positive result	Reference standard; details; definition of positive result	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Bulloch, 2000[47]	Urine clarity; visual inspection	Writing could not be read through tube	Culture; catheter $>10^4$, CVU $\geq 10^5$ cfu ml^{-1}	26	23	3	107	89.7	82.3	34.6	4.9	0.14
Lagos, 1994[74]	Urine clarity; visual inspection	Cloudy	Culture and microscopy: $\geq 10^5$ cfu ml^{-1} and >10 cells mm^{-3}	317	181	31	461	91.1	71.8	25.6	3.2	0.13
Struthers, 2003[104]	Urine odour questionnaire, determined by parents	One or more positive replies	Culture $\geq 10^5$ cfu ml^{-1}	3	54	4	49	42.9	47.6	0.7	0.8	1.18
Liptak, 1993[78]	Clinical: urinary odour and cloudiness, fever, malaise, haematuria, pain	Presence of any symptoms	Culture $\geq 10^5$ cfu ml^{-1}	58	54	46	171	55.8	76	4.0	2.3	0.58
Gorelick, 2000[64]	Clinical decision rule: <12 months old, white race, temperature of ≥39 °C, fever for 2 days or more, absence of another source of fever on examination	Presence of at least 2/5 variables	Culture $\geq 10^4$ cfu ml^{-1}	60	970	3	436	95.2	31.0	7.8	1.4	0.18
Cervilla, 2001[49]	Temperature	>38.1°C	Culture; SPA $\geq 10^2$, catheter $\geq 10^4$, CVU $\geq 10^5$ cfu ml^{-1}	20	26	23	33	46.5	55.9	1.1	1.1	0.96

a, true positives; b, false positives; c, false negatives; d, false negatives (see Glossary).
cfu, colony-forming units.

TABLE 6 *Results of studies of urine sampling*

Study; group	Index test positive result	Reference standard positive result	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
CVU vs SPA: culture											
Aronson, 1973[41]: *age <18 months*	$>10^5$ cfu ml^{-1}	$\geq 10^2$ cfu ml^{-1}	20	27	4	35	83.3	56.5	5.9	1.9	0.30
Aronson, 1973[41]: *age: 3–12 years*	$>10^5$ cfu ml^{-1}	$\geq 10^2$ cfu ml^{-1}	10	4	4	16	71.4	80	8.6	3.6	0.36
Hardy, 1976[65]	$\geq 10^4$ cfu ml^{-1}	$\geq 10^4$ cfu ml^{-1}	3	1	1	25	75	96.2	39.7	19.5	0.26
Morton, 1982[87]	$>10^5$ cfu ml^{-1}	Any growth	5	0	0	25	100	100	561	47.7	0.08
Pylkkanen, 1979[94]	$\geq 10^4$ cfu ml^{-1}	Not clear	179	18	20	55	89.9	75.3	26.3	3.6	0.13
Pylkkanen, 1979[94]	$\geq 10^5$ cfu ml^{-1}	Not clear	161	9	38	64	80.9	87.7	28.5	6.6	0.22
Ramage, 1999[95]	10^5 cfu ml^{-1}	Any growth	16	2	2	38	88.9	95	101.6	17.8	0.12
Bag/CVU vs catheter: culture and microscopy											
Braude, 1967[46]: *age <5 years*	Not clear	$\geq 10^4$ cfu ml^{-1} and ≥ 10 WBC mm^{-3}	17	4	4	26	81.0	86.7	22.9	5.5	0.24
Braude, 1967[46]: whole sample	Not clear	$\geq 10^4$ cfu ml^{-1} and ≥ 10 WBC mm^{-3}	23	7	7	31	76.7	81.6	13.2	3.9	0.30
Bag vs SPA: culture											
Fernandez, 1996[61]	$>10^5$ cfu ml^{-1}	$\geq 10^3$ cfu ml^{-1}	13	4	0	31	100	88.6	189	7.7	0.04
Hardy, 1976[65]	$\geq 10^4$ cfu ml^{-1}	$\geq 10^4$ cfu ml^{-1}	2	2	2	24	50	92.3	9.8	5.4	0.55
Pad/nappy vs bag: culture											
Farrell, 2002[57]	Bacteria present	Bacteria present	2	0	3	15	40	100	22.1	13.3	0.60
Feasey, 1999[58]	$\geq 10^5$ cfu ml^{-1}	$\geq 10^5$ cfu ml^{-1}	2	0	0	44	100	100	445	75.0	0.17
Ahmad, 1991[38]	$>10^5$ cfu ml^{-1}	$\geq 10^5$ cfu ml^{-1}	6	0	0	39	100	100	1027	74.3	0.07
Pad/nappy vs SPA: culture											
Cohen, 1997[51]	$>10^5$ cfu ml^{-1}	$\geq 10^5$ cfu ml^{-1}	5	2	0	31	100	93.9	138.6	12.5	0.09
Early- vs late-stream catheter sample: culture											
Dayan, 2000[54]	$\geq 50{,}000$ cfu ml^{-1}	$\geq 50{,}000$ cfu ml^{-1}	6	4	0	76	100.0	95.0	221.0	16.7	0.08

WBC, white blood cells.

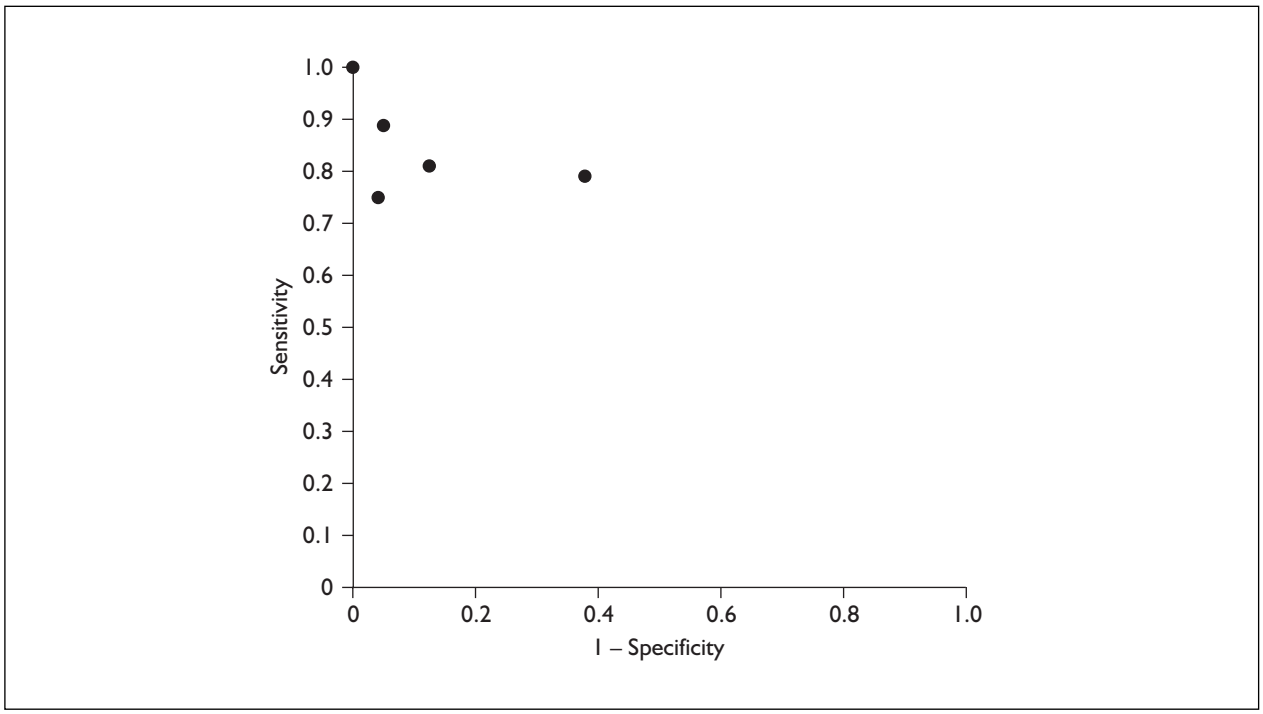

FIGURE 5 *Urine sampling (accuracy of CVU using SPA as the reference standard): study sensitivity and 1 – specificity plotted in ROC space*

CVU samples

Five studies reporting seven data sets assessed the diagnostic accuracy of a clean catch urine sample, using an SPA urine sample as the reference standard.[41,65,87,94,95] Four studies used culture alone of both urine samples, whereas one used a combination of culture and microscopy of the two samples.[41] Two studies each provided two estimates of test performance. One stratified results by age. The results of this study were combined for this review to give an overall estimate of test performance.[41] The other evaluated test performance at two different cut-off points.[94] The results for a cut-off point of over 10^5 cfu ml^{-1} were used in this review for further analysis as this was the cut-off point used in most of the other studies.

Sensitivity ranged from 75% (specificity of 96%) to 100% (specificity of 100%) and specificity ranged from 57% (sensitivity 83%) to 100% (sensitivity 100%). The positive likelihood ratios (LRs) ranged from 1.9 (LR– 0.30) to 47.7 (LR– 0.08). Negative likelihood ratios ranged from 0.08 (LR+ 47.7) to 0.36 (LR+ 3.57). *Figure 5* shows estimates of sensitivity and 1 – specificity from these studies plotted in a receiver operating characteristic (ROC) space. Although there was considerable heterogeneity in study results, all studies were clustered towards the upper left-hand corner of the graph, suggesting that acceptable diagnostic performance is obtained from CVU samples. Looking at the graph, one study appeared to be an outlier.[41] All studies were of reasonable quality and the one outlying study did not differ in quality from the others. The outlying study used culture and microscopy of the two samples, whereas all other studies used culture alone.

There was considerable heterogeneity in positive likelihood ratios ($p < 0.0001$). However, the negative likelihood ratios were statistically homogeneous ($p = 0.504$). The pooled positive likelihood ratio was 7.7 [95% confidence interval (CI) 2.5 to 23.5] and the pooled negative likelihood ratio was 0.23 (95% CI 0.18, 0.30). The median positive likelihood ratio was 17.8 [interquartile range (IQR) 6.6–19.5]. The median negative likelihood ratio was 0.22 (IQR 0.12–0.26). There were insufficient studies to allow further investigation of heterogeneity.

Bag specimens

One study[46] provided separate estimates of test performance for two age groups and compared culture and microscopy results from bag specimens to those obtained from catheter specimens. Both estimated sensitivity and specificity to be around 80%. The appropriateness of a catheter specimen as the reference standard is questionable, and this, along with the small amount of data available, means that these results are of limited value.

Two studies compared culture of urine bag specimens to culture of SPA samples.[61,65] There were considerable differences in the results from these studies, with one reporting a sensitivity of 100% and the other a sensitivity of 50%. Both studies found specificity to be around 90%. Overall, there were insufficient data to draw any conclusions regarding the appropriateness of using urine samples obtained from bags.

Pad/nappy specimens

Four studies examined the diagnostic accuracy of pad/nappy specimens. Three compared culture of pad/nappy specimens to culture of bag specimens.[38,57,58] Given the fact that bag specimens are unlikely to be the best method of urine sample collection, the results of these studies are of limited value.

One study compared culture of a pad/nappy specimen to culture of SPA samples.[51] This study reported a sensitivity of 100% and specificity of 94%, suggesting excellent agreement between the two sampling methods. However, the limited data (three of the cells in the 2×2 table contained five or fewer children) make it impossible to draw firm conclusions from this finding.

Early- versus late-stream samples

One study compared the results of culture from early catheter samples to those from late catheter samples.[54] The agreement between the two samples was very good, with an estimated sensitivity of 100% and specificity of 95%. This study also contained limited data, with three of the cells in the 2×2 table containing six or fewer children, again making firm conclusions impossible.

Summary

The only type of urine sampling for which a reasonable amount of data was available was the comparison of CVU samples to SPA samples. These data showed that when both samples are cultured the agreement between the two methods was good. The one outlying study showed a poor performance of CVU. The reasons for this are unclear, but may be related to the fact that this study used a combination of culture and microscopy of the two urine samples, whereas all other studies used culture alone.

Dipstick tests

A total of 38 studies reporting 106 data sets evaluated dipstick tests for the diagnosis of UTI.[39,40,45,47,49,52,53,55,56,59,60,62,70–75,78,81,83–85,89–91,99,101–103,105,106,109–113,230] These studies assessed the utility of dipstick tests for nitrite, LE, protein, glucose and blood alone and in combination.

Nitrite

Twenty-three studies reporting 27 data sets examined nitrite dipstick tests.[45,47,53,55,59,60,62,70,72–75,78,81,84,89,91,101,105,110–112,230] More than half (17/23) of these studies did not include an appropriate spectrum of patients, and patient selection criteria were often inadequately described. Avoidance of test review bias was poorly reported; 14 out of the 23 studies did not report whether those interpreting results were blinded to the results of other tests. Differential verification bias was a problem in three studies. One study presented the results from visual interpretation and a subset of results as read by a machine.[45] Only those results that were interpreted visually were included in the analysis below, as these were more comparable to other studies in the group, and were available for the whole sample. One study presented results for the total population as well as separate results for infants aged less than 1 year.[101] The results for the total population were used in all further analyses to prevent data from the same participants being used twice. A third study incubated samples that were initially negative for 4 hours. The samples that remained negative were then incubated with sodium nitrite for a further 4 hours. Results were presented for each stage of this process.[105] Only the results obtained initially without incubation were included in further analyses, as these were considered more comparable to other studies in the group (all other nitrite dipstick tests were performed without incubation). The final study to report more than one test evaluation examined dipsticks from two different manufacturers. Results were highly consistent (there were two more true-negative results, with one test out of a total of 809 negative results). Only one set of results was further analysed to prevent duplication of almost identical data from the same participants.[112] It was decided to use the results from the Multistix™ (Ames) test as it was also used in several other studies.

Culture was used as the reference standard in all but two studies. In these two studies, a combined reference standard of culture and microscopy was used.[74,75] In the majority of those studies where culture was the reference standard (15/21), the cut-off point defining presence of a UTI was 10^5 cfu ml^{-1}. Of the remaining six studies where culture was the reference standard, one study used 10^4 cfu ml^{-1} as the cut-off point,[230] one used 50,000 cfu ml^{-1},[60] and the others used more than one urine sampling technique (and,

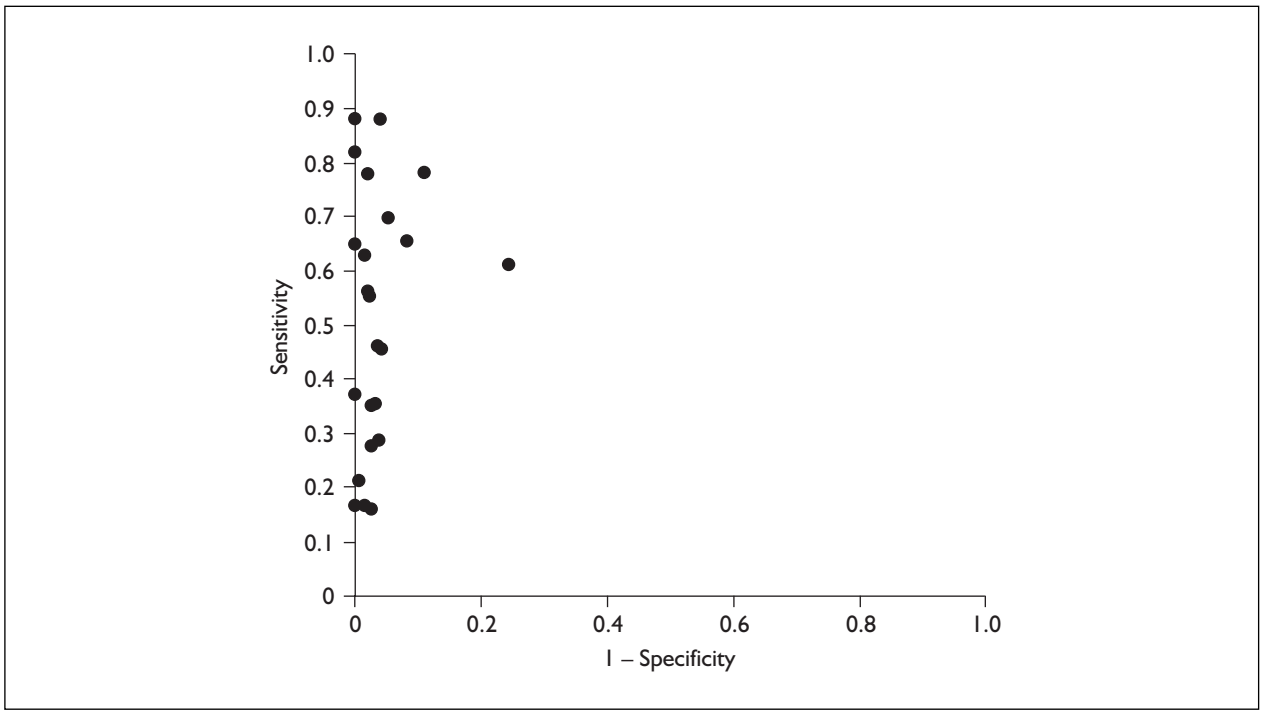

FIGURE 6 *Nitrite: study sensitivity and 1 – specificity plotted in ROC space*

correspondingly, more than one cut-off point) per study (see *Table 11*).[47,53,81,84]

In general, studies found relatively poor sensitivity with high specificity. Sensitivity ranged from 16.2% (specificity 97.6%) to 88.1% (specificity 100%). Specificity ranged from 75.6% (sensitivity 61.1%) to 100% (sensitivity 16.7–88.1%); all but two estimates of specificity were above 90%. There was considerable heterogeneity in terms of likelihood ratios ($p < 0.001$). Positive likelihood ratios ranged from 2.5 (LR– 0.51) to 439.6 (LR– 0.63). Negative likelihood ratios ranged from 0.12 (LR+ 157) to 0.86 (LR+ 6.7). The pooled positive likelihood ratio was 15.9 (95% CI 10.7 to 23.7) and the pooled negative likelihood ratio was 0.51 (95% CI 0.43 to 0.60). However, owing to significant heterogeneity these should be interpreted with extreme caution. *Figure 6* shows estimates of sensitivity and 1 – specificity plotted in ROC space. The median positive likelihood ratio was 15.1 (IQR: 10.8–47.7) and the median negative likelihood ratio was 0.5 (IQR 0.33–0.70).

A regression analysis was carried out to investigate possible explanations for the observed heterogeneity. The regression model $D = \alpha + \beta S$ was extended to include variables for quality items, age and region. The results of the univariate regression analysis are shown in *Table* 7. Two items, the presence of clinical review bias and region, were significant in this analysis. These items were included in a multivariate analysis, but only clinical review bias remained significant. The DOR was around 4.5 times higher in studies that avoided clinical review bias than in studies in which clinical review bias may have been a problem. The adjusted r^2 value increased from 0.09 to 0.40 when clinical review bias was included in the model, suggesting that this variable accounted for some but not all of the observed heterogeneity. However, this association may reflect the quality of reporting, as clinical review bias was scored as 'not clear' rather than 'no' in the majority of studies.

Leucocyte esterase

Fourteen studies reporting a total of 16 data sets examined the diagnostic accuracy of dipstick tests for LE.[47,53,55,62,74,75,78,81,84,89,101,111,112,230] The majority of these studies (12/14) did not use an appropriate spectrum of patients. Around one-third of studies did not provide an adequate description of the criteria used to select patients or of the reference standard used to confirm diagnosis. Approximately half of the studies did not report sufficient information to assess the avoidance of review bias, and partial verification bias was a problem in one study. One study reported results for the whole population and

TABLE 7 *Results of the regression analysis for nitrite dipstick*

Variable		β	RDOR	*p*-Value	Adjusted r^2
Spectrum composition		–0.5	0.6	0.548	0.11
Selection criteria		0.1	1.1	0.896	0.09
Reference standard		–0.9	0.4	0.75	0.09
Time		–0.5	0.6	0.378	0.12
Partial verification		1	2.7	0.484	0.11
Differential verification		0.4	1.5	0.782	0.09
Incorporation				Dropped	
Test details		0.2	1.2	0.797	0.09
Reference standard details		0.8	2.2	0.19	0.16
Test bias		0.1	1.1	0.896	0.09
Review bias		–0.6	0.5	0.601	0.10
Clinical review bias		1.5	4.5	0.004	0.40
Uninterpretable results		–0.7	0.5	0.47	0.11
Withdrawals				Dropped	
Age:	< 2 years			Reference	0.17
	< 5 years	1.8	6.0	0.213	
	<12 years	0.8	2.2	0.548	
	<18 years	1.3	3.7	0.076	
Region:	North America			Reference	0.29
	Europe	–1.2	0.3	0.04	
	Other	–1.2	0.3	0.05	

provided separate results for children less than 1 year of age.[101] As the results were very similar for both analyses, only the results for the whole group were included in this analysis. Another study evaluated dipstick tests from two different manufacturers and reported very similar results for the two products.[112] Only the results for one of the tests, Multistix (Ames) were included in the present analysis to avoid duplication of data. Twelve studies used culture as the reference standard and two used a combination of culture and microscopy.[74,75]

In general, specificity was higher than sensitivity. Sensitivity ranged from 37.5% (specificity 96.4%) to 100% (specificity 92%). Specificity ranged from 69.3% (sensitivity 93.5%) to 97.8% (sensitivity 70%). Positive likelihood ratios ranged from 2.6 (LR– 0.39) to 32.2 (LR– 0.31). Negative likelihood ratios ranged from 0.02 (LR+ 12.5) to 0.66 (LR+ 6.97). There was considerable heterogeneity in both positive and negative likelihood ratios ($p < 0.001$). The pooled positive likelihood ratio was 5.5 (95% CI 4.1 to 7.3) and the pooled negative likelihood ratio was 0.26 (95% CI 0.18 to 0.36). These should be interpreted with extreme caution owing to the presence of significant heterogeneity. *Figure 7* shows estimates of sensitivity and 1 – specificity plotted in ROC space. The median positive likelihood ratio was 5.0 (IQR 3.3–12.3) and the median negative likelihood ratio was 0.23 (IQR 0.16–0.30).

A regression analysis was carried out to investigate possible explanations for the observed heterogeneity. *S* was not significant in the regression model $D = \alpha + \beta S$. Standard metaregression analysis using *D* as the dependent variable was therefore performed. The following variables were investigated: quality, age and region. The results of the univariate regression analysis are shown in *Table 8*. None of the items investigated showed a significant association with the DOR.

Protein

Only two studies with three data sets examined the diagnostic accuracy of protein dipsticks.[45,75] One used culture as the reference standard[45] and the other used a combination of culture and microscopy.[75] Neither of these studies used an appropriate spectrum of patients or adequately reported the criteria used to select patients. Both studies reported insufficient information to assess the avoidance of review bias. Both studies found that protein dipstick was a poor test for the diagnosis of UTI. Estimates of sensitivity were particularly low, ranging from 8.1% (specificity 95.1%) to 53.3% (specificity 83.9%).

Glucose

Four studies with five data sets examined the diagnostic accuracy of biochemical test strips for glucose.[56,71,99,106] These test strips are not currently commercially available. Currently

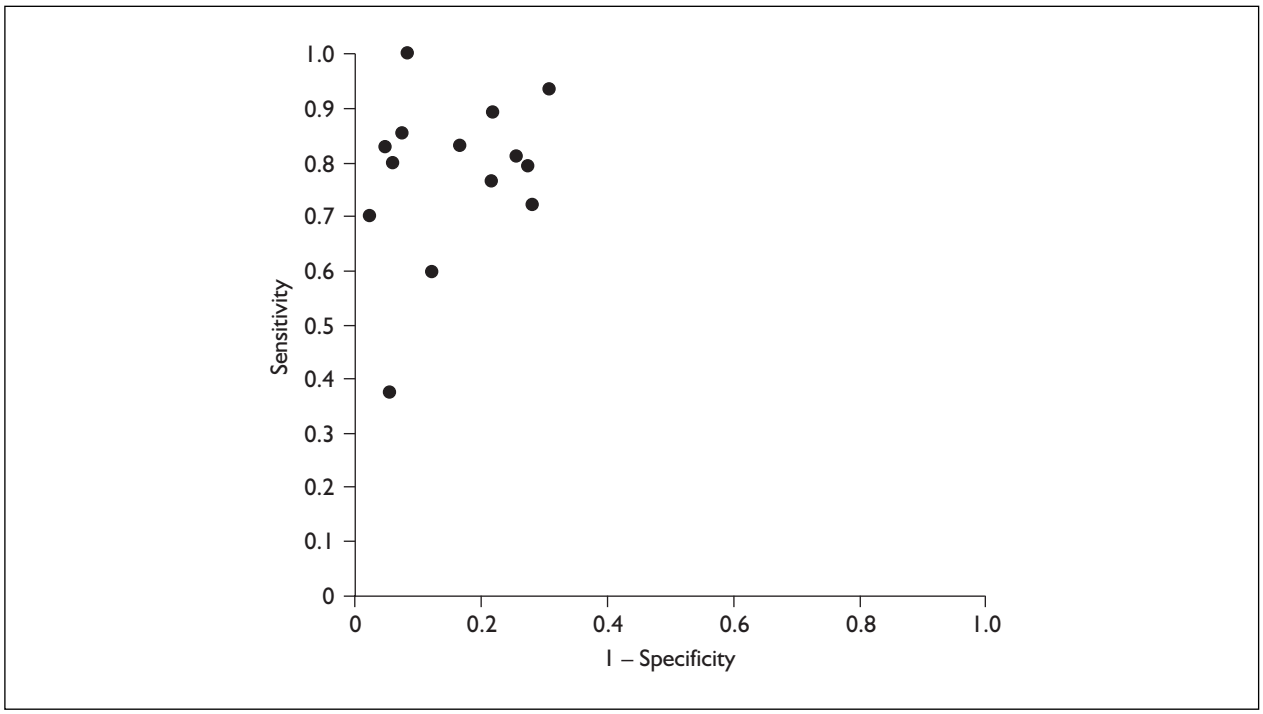

FIGURE 7 *LE: study sensitivity and 1 – specificity plotted in ROC space*

TABLE 8 *Results of the regression analysis for LE dipstick*

Variable		β	RDOR	*p*-Value	Adjusted r^2
Spectrum composition		0	1.0	0.979	–0.08
Selection criteria		0.4	1.5	0.377	–0.01
Reference standard				Dropped	
Time		0	1.0	0.955	–0.08
Partial verification		0.9	2.5	0.44	–0.03
Differential verification				Dropped	
Incorporation				Dropped	
Test details		–3	0.05	0.255	0.03
Reference standard details		0.6	1.8	0.364	–0.01
Test bias		–0.2	0.8	0.729	–0.07
Review bias		–0.2	0.8	0.775	–0.08
Clinical review bias		0.4	1.5	0.386	–0.01
Uninterpretable results		–1	0.4	0.203	–0.06
Withdrawals				Dropped	
Age:	< 2 years			Reference	0.00
	< 5 years	1.5	4.5	0.252	
	< 12 years	3.2	24.5	0.247	
	< 18 years	0.2	1.2	0.782	
Region:	North America			Reference	–0.04
	Europe	–0.8	0.4	0.268	
	Other	0	1.0	0.939	

available dipstick tests for glucose are designed to detect abnormally high urinary glucose levels (typically >100 mg dl^{-1}), using a colour change at this threshold. The test strips evaluated by studies in this section differentiated between normal urinary glucose in an overnight fasting sample and low levels caused by the presence of bacteriuria. A positive test was therefore denoted as the absence of colour change, the threshold for which was 2–3 mg dl^{-1}. None of these studies used an appropriate spectrum of patients or provided an adequate description of patient selection criteria. Three studies did not adequately describe the reference standard used to confirm diagnosis.

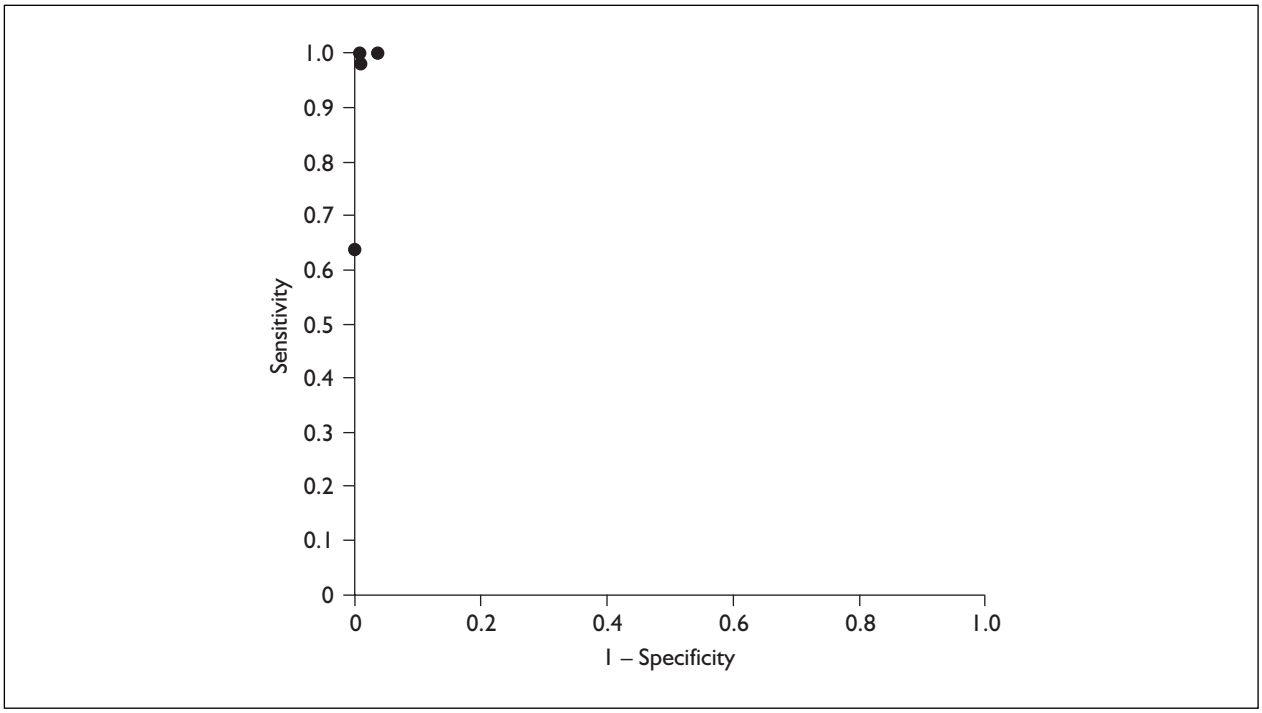

FIGURE 8 *Glucose: study sensitivity and 1 – specificity plotted in ROC space*

One study provided estimates of test performance for two different groups of children.[56] The first group of children were healthy newborns, whereas the second group consisted of children with a suspected UTI. Only the results from the second group of children are included in the analysis as these are more similar to the children included in the other studies and are also more similar to the children in whom the test will be used in practice. All studies reported very high estimates of specificity, ranging from 96.4 to 100%. Sensitivity was also very high (over 98%) in three studies, but was only 64% in the fourth study.[56]

Positive likelihood ratios ranged from 27.8 (LR– 0.07) to 166.2 (LR– 0.02) with a pooled estimate of 66.3 (95% CI 20.0 to 219.6). Negative likelihood ratios ranged from 0.02 (LR+ 166.2 and 113.7) to 0.36 (LR+ 32.5), with a pooled estimate of 0.07 (95% CI 0.01 to 0.83). However, there was significant heterogeneity in both positive and negative likelihood ratios ($p<0.001$) and so the pooled estimates should be interpreted with extreme caution. *Figure 8* shows estimates of sensitivity and 1 – specificity plotted in ROC space. The median positive likelihood ratio was 73.1 (IQR 31.3–126.8) and the median negative likelihood ration was 0.05 (IQR 0.02–0.15).

The only study to include an appropriate spectrum of patients was the study that reported a much lower sensitivity than the other studies.[56] Two of the other studies were conducted in screening settings[71,99] and one was conducted in a high-risk group.[106] The study reporting a low sensitivity was conducted in children aged less than 1 year, whereas the other three studies were conducted in children 3 years of age or more. The studies were all conducted between 1968 and 1974, suggesting that this is a test that is not currently in use for the diagnosis of UTI. However, given the promising results from studies further research in this area may prove useful.

Blood

One study investigated the diagnostic accuracy of dipstick tests for blood.[45] The patient selection criteria were inadequately reported, and insufficient information was provided to determine the appropriateness of the spectrum of patients. Insufficient details were reported to assess the avoidance of review bias. This study reported two estimates of test performance; one for visual examination of the results and one for a subset of patients whose results were analysed using an automated device. This study suggested that dipstick for blood is not a useful tool for the diagnosis of UTI in children, with estimated sensitivities of 25.4% and 53.3% for visual and automated examination, respectively, and specificities of around 85%.

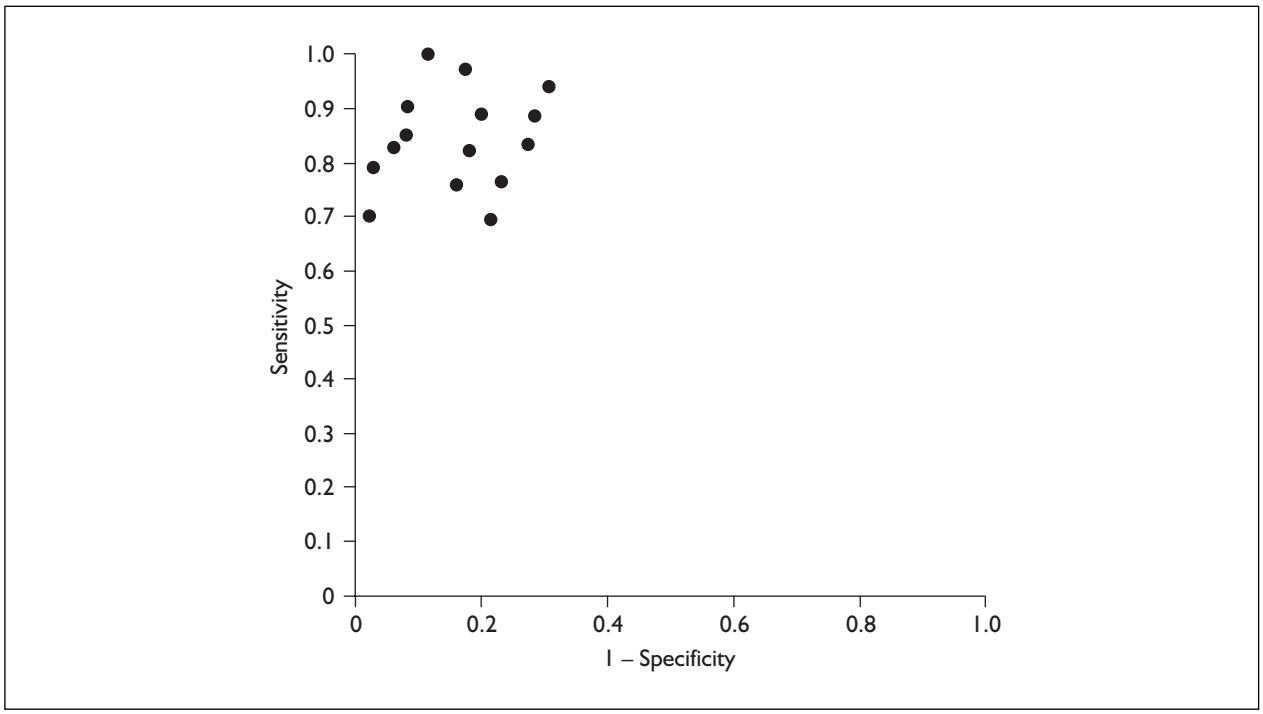

FIGURE 9 *LE or nitrite: study sensitivity and 1 – specificity plotted in ROC space*

Combinations of two dipstick tests

LE or nitrite positive

Fifteen studies including a total of 20 data sets examined the combination of LE and nitrite tests, where a positive result from either was taken as a positive test result.[39,47,53,74,78,81,84,101–103,109,111–113,230] The majority of these studies (11/15) did not use an appropriate spectrum of patients, and two did not adequately describe the criteria used to select patients. Around two-thirds of studies did not report whether or not those interpreting test results were blinded to the results of other tests (avoidance of test review bias). One study reported results for the whole population and for a subgroup of children aged less than 1 year.[101] As results were very similar for both groups, only the results for the total population were included in the analysis. A second study also reported results for the whole population and for a subgroup of children aged less than 2 years. As with the previous study, only results for the total population were included in the analysis.[103] This second study also reported results for two different cut-off points for LE in the total population: at least trace LE and at least small LE. As most other studies classed at least trace LE as positive only this test evaluation was included in the analysis. A second study by the same author also examined two cut-off points for LE: at least trace and at least moderate LE.[102] As before, only the results for at least trace LE were included in the analysis. A further study used dipstick tests produced by two different manufacturers.[112] This study reported very similar results for the two products. Only the results for the Multistix test are included in the analysis as this test was most commonly used by other studies in this section.

The use of a combination test, where either a positive LE or a positive nitrite denoted a positive result, appeared to increase sensitivity and decrease specificity compared with the results for the individual dipstick tests. Sensitivity ranged from 69.4% (specificity 78.5%) to 100% (specificity 88.4%). Specificity ranged from 69.2% (sensitivity 94.1%) to 97.8% (sensitivity 70%). The likelihood ratios showed considerable heterogeneity ($p < 0.001$). Positive likelihood ratios ranged from 3.0 (LR– 0.23) to 32.2 (LR– 3.1). Negative likelihood ratios ranged from 0.03 (LR + 5.6) to 0.39 (LR+ 3.2). The pooled positive likelihood ratio was 6.1 (95% CI 4.3 to 8.6) and the pooled negative likelihood ratio was 0.20 (95% CI 0.16 to 0.26). *Figure 9* shows sensitivity and 1 – specificity plotted in ROC space.

The median positive likelihood ratio was 4.7 (IQR 3.2–10.6) and the median negative likelihood ratio was 0.2 (IQR 0.1–0.3).

A regression analysis was carried out to investigate possible explanations for the observed heterogeneity. *S* was not significant in the

TABLE 9 *Results of the regression analysis for positive nitrite or LE dipstick*

Variable		β	RDOR	*p*-Value	Adjusted r^2
Spectrum composition		1.2	3.3	0.016	0.32
Selection criteria		–0.9	0.4	0.073	0.17
Reference standard				Dropped	
Time		0.1	1.1	0.879	–0.07
Partial verification				Dropped	
Differential verification				Dropped	
Incorporation				Dropped	
Test details				Dropped	
Reference standard details		1.4	4.1	0.107	0.13
Test bias		–0.7	0.5	0.291	0.01
Review bias		–1	0.4	0.408	–0.02
Clinical review bias		–0.1	0.9	0.86	–0.07
Uninterpretable results		–0.9	0.4	0.333	0.00
Withdrawals				Dropped	
Region:	North America			Reference	0.12
	Europe	–1.5	0.2	0.088	
	Other	–0.6	0.5	0.364	

regression model $D = \alpha + \beta S$. Therefore, standard metaregression analysis using D as the dependent variable was performed. The following variables were investigated: quality and region. The results of the univariate regression analysis are shown in *Table 9*. The only item to show a significant association with D was spectrum composition. The DOR was around three times higher if an appropriate spectrum of patients was included. The estimate of r^2 increased from around 0.01 to 0.32, suggesting that differences in spectrum composition may have accounted for a reasonable proportion, but not all, of the observed heterogeneity.

LE and nitrite positive

Nine studies that included a total of 12 data sets examined the combination of nitrite and LE dipstick results, where a positive test from both tests was taken as a positive test result.[53,62,75,83,84,101,103,112,113] The majority of these studies (7/9) either did not include an appropriate spectrum of patients or reported insufficient information to determine appropriateness; four studies inadequately reported the criteria used to select patients. The reference standard used to confirm diagnosis was poorly described in five studies. Less than half of the studies reported sufficient information to assess the avoidance of test review bias. Two studies presented results for the whole population and also for an age-stratified sample (aged <1 year and aged <2 years).[101,231] To avoid counting the same data twice, only the results for the whole population were included in the analysis. One study evaluated dipstick tests from two different manufacturers.[112] As this study reported almost identical results for the two products, only the results for the most commonly studied test, Multistix, were included in the analysis.

Sensitivity was generally low, ranging from 30% (specificity 100%) to 89.2% (specificity 97.6%). Specificity was higher, with all but one estimate being over 90%, and ranged from 89.2% (sensitivity 87%) to 100% (sensitivity 30–88%). Negative likelihood ratios were statistically heterogeneous ($p<0.001$). Positive likelihood ratios were also heterogeneous, but to a lesser extent ($p = 0.037$). Positive likelihood ratios ranged from 8.0 (LR– 0.15) to 197.1 (LR– 0.17). Negative likelihood ratios ranged from 0.11 (LR+ 36.7) to 0.7 (LR+ 107.7). The pooled positive likelihood ratio was 28.2 (95% CI 15.5 to 43.4) and the pooled negative likelihood ratio was 0.37 (95% CI 0.26 to 0.52). *Figure 10* shows sensitivity and 1 – specificity plotted in ROC space. The median positive likelihood ratio was 36.7 (IQR 20.4–52.1) and the median negative likelihood ratio was 0.2 (IQR 0.1–0.6).

Protein and LE positive

One study examined the diagnostic accuracy of a positive LE and a positive protein dipstick.[75] Insufficient information was reported to determine the appropriateness of the spectrum of patients or the avoidance of review bias. Both the criteria used to select patients and the reference standard used to confirm diagnosis were inadequately described. This study reported a sensitivity of 89.2% and a specificity of 97.6%, suggesting that this combination may be a good

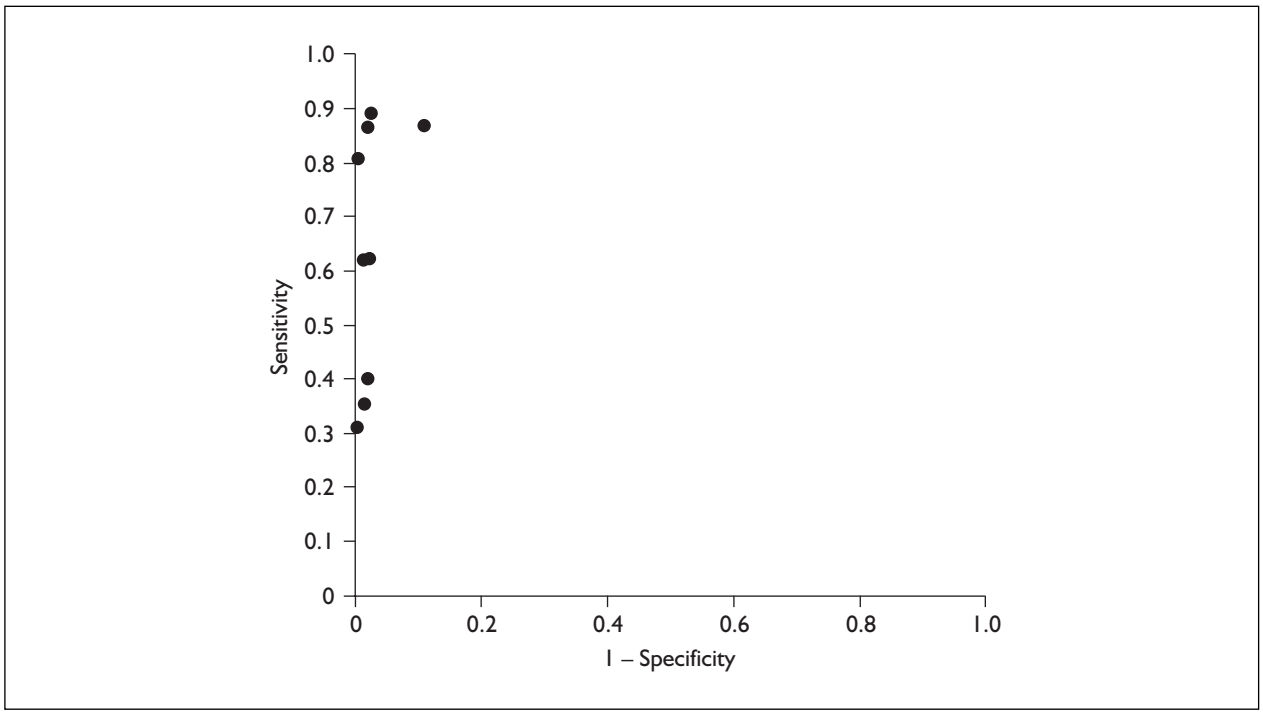

FIGURE 10 *LE and nitrite: study sensitivity and 1 – specificity plotted in ROC space*

test for the diagnosis of UTI. However, it is difficult to draw firm conclusions from only one study.

Combinations of three dipstick tests

Five studies including a total of ten data sets examined the diagnostic accuracy of various combinations of three dipstick tests.[45,75,112,232,233] Four studies each evaluated one combination of tests (nitrite, blood or protein positive;[45] nitrite, blood or LE positive;[49] nitrite, blood and LE positive;[52] nitrite, LE or protein positive[112]) and one combination was evaluated by two studies (nitrite, LE and protein positive).[75,112] Only one study included an appropriate spectrum of patients, and three studies did not adequately describe the criteria used to select patients. The reference standard used to confirm diagnosis was poorly described in two studies. Three studies reported inadequate information to assess the avoidance of review bias, and partial verification bias was a problem in one study. With little information available for each test combination it is not possible to draw firm conclusions regarding the utility of any of these testing strategies. However, one test combination, that of nitrite, LE and protein positive, did appear to be potentially informative for the diagnosis of UTI. This combination was evaluated by two studies.[75,112] One study reported excellent test performance, with a sensitivity of 96% and a specificity of 99%. The second study reported less accurate results, with a sensitivity of 89% and specificity of 72%. Further investigation of this test combination is needed.

Summary

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, blood or combinations of three different tests.

Figure 11 shows the estimates of sensitivity and 1 – specificity plotted in ROC space for all studies that examined the diagnostic accuracy of tests for glucose, and dipstick tests for nitrite and LE, alone and in combination. *Table 10* shows the pooled likelihood ratios for the different combinations of dipstick tests for LE and nitrite. The graph and the pooled estimates for the likelihood ratios suggest that glucose is considerably better than the other tests, both for ruling in disease and for ruling out disease. However, the confidence intervals around the pooled estimates are very large, especially for the negative likelihood ratios (ruling out disease), suggesting considerable uncertainty in these estimates. It should also be noted that very few

TABLE 10 *Pooled likelihood ratios for dipstick tests*

Dipstick	Pooled LR+ (95% CI)	Pooled LR– (95% CI)
Nitrite	15.9 (10.7 to 23.7)	0.51 (0.43 to 0.60)
LE	5.5 (4.1 to 7.3)	0.26 (0.18 to 0.36)
Nitrite or LE positive	6.1 (4.3 to 8.6)	0.20 (0.16 to 0.26)
Nitrite and LE positive	28.2 (17.3 to 46.0)	0.37 (0.26 to 0.52)
Glucose	66.3 (20.0 to 219.6)	0.07 (0.01 to 0.83)

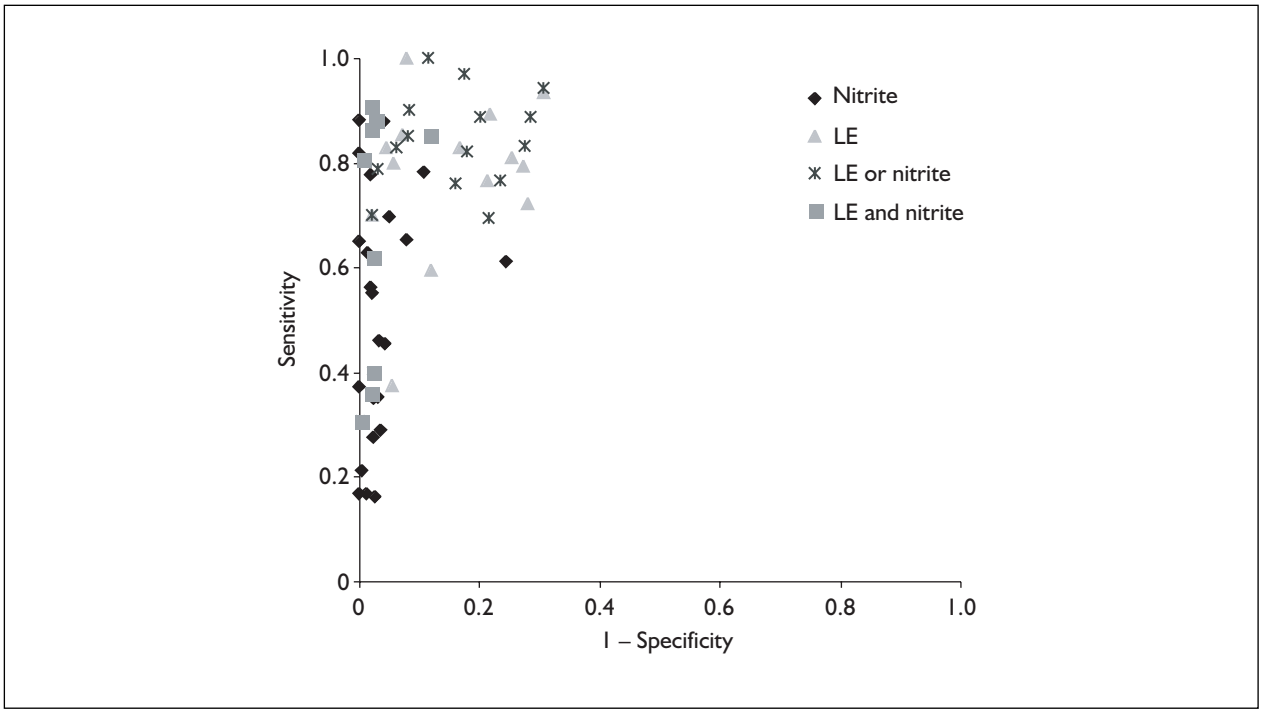

FIGURE 11 *Estimates of sensitivity and specificity for different dipstick tests plotted in ROC space*

studies of glucose tests were available and that they were all conducted over 30 years ago.

When looking at the dipstick tests for nitrite and LE, although there is considerable heterogeneity between studies, the data suggest that nitrite combined with LE (both positive) has the highest specificity and so may be useful for ruling in disease. The pooled positive likelihood ratios are highest for nitrite and LE combined, supporting the suggestion that these may be useful tests for ruling in disease. The negative likelihood ratio is lowest for the combination of nitrite or LE positive (i.e. both dipstick tests negative), suggesting that this combination may be useful for ruling out disease. A result of either test positive is not very good either for ruling in or ruling out disease.

What do these results mean?

The results of studies of dipstick tests are summarised in *Table 11*. As shown above, a dipstick test positive for both LE and nitrite is best for ruling in UTI and a test negative for both LE and nitrite is best for ruling out UTI. An indeterminate dipstick test result (positive for either LE or nitrite, negative for the other) provides very little useful diagnostic information. The glucose test was also found to be very good for both ruling in and ruling out UTI, but owing to the limitations of the studies of this test these results should be interpreted with caution. However, what these results mean in practical terms may not be immediately obvious. To understand what these results mean, the easiest measures to look at are the predictive values.

If one takes an estimate for the prevalence of UTI in children presenting to their GP with symptoms of possible UTI (the pretest probability of disease), that is, children in whom tests for the diagnosis of UTI are likely to be used, then the likelihood ratios can be used to calculate the post-test

TABLE 11 *Results of studies of dipstick tests*

Study; group	Index test manufacturer	Reference standard positive result	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Nitrite vs culture											
Armengol, 2001[230]	Ames/BM	$\geq 10^4$ cfu ml^{-1}	5	0	25	230	16.7	100.0	99.4	82.0	0.83
Boreland, 1986[45]	N-Labstix (Ames)	$\geq 10^5$ cfu ml^{-1}	37	10	30	459	55.2	97.9	53.8	25.9	0.46
	N-Labstix (Ames); read by Clinitek 200		8	2	7	147	53.3	98.7	66.9	31.9	0.48
Bulloch, 2000[47]	Clinitek (Ames)	Catheter $>10^4$, CVU $\geq 10^5$ cfu ml^{-1}	8	3	21	127	27.6	97.7	14.4	12.0	0.74
Dayan, 2002[53]	Automated urine analysis; Super UA	SPA $\geq 10^3$, Catheter $\geq 10^4$ cfu ml^{-1}	7	4	13	169	35.0	97.7	20.9	15.1	0.67
Demi, 1993[55]	BM	NS	4	3	20	220	16.7	98.7	13.8	12.4	0.84
Fennell, 1977[59]	Bac-U-Dip (Warner-Chilcutt)	$\geq 10^5$ cfu ml^{-1}	7	1	2	55	77.8	98.2	111.0	43.6	0.23
Fernandez, 2000[60]	Multistix (Bayer)	>50,000 cfu/ml	40	3	47	85	46.0	96.6	20.8	13.5	0.56
Giraldez, 1998[62]	NR	$\geq 10^5$ cfu ml^{-1}	22	1	3	24	88.0	96.0	105.0	22.0	0.13
Holland, 1968[70]	Stat-Test (Mallinckrodt)	$\geq 10^5$ cfu ml^{-1}	13	0	7	106	65.0	100.0	383.4	137.6	0.35
Kunin, 1977[72]	Microstix-Nitrite (Ames)	$\geq 10^5$ cfu ml^{-1}	18	0	4	55	81.8	100.0	456.3	90.1	0.18
Labbe, 1982[73]	N-Uristix (Ames); home based (mothers)	$\geq 10^5$ cfu ml^{-1}	37	0	5	89	88.1	100.0	1220.5	157.0	0.12
Liptak, 1993[78]	Chemstrip 9 (BM)	$\geq 10^5$ cfu ml^{-1}	68	18	36	207	65.4	92.0	21.0	8.2	0.38
Lohr, 1993[81]	Ames; read by Clinitek	SPA $>10^3$, catheter $>10^4$; CVU $>10^5$ cfu ml^{-1}	38	0	64	587	37.3	100.0	701.4	439.6	0.63
Marsik, 1986[84]	Chemstrip (biodynamics)	SPA any bacteria; catheter $>10^4$; CVU $>10^5$ cfu ml^{-1}	37	28	16	520	69.8	94.9	41.5	13.7	0.32
Ordonez, 1994[89]	Ames (Multistix)	$\geq 10^5$ cfu ml^{-1}	11	20	7	62	61.1	75.6	4.7	2.5	0.51
Parmington, 1989[91]	NR	NS	43	27	12	223	78.2	89.2	28.3	7.2	0.24
Sharief, 1998[101]											
Total population	Multistix (Bayer); automated reader	$\geq 10^5$ cfu ml^{-1}	6	10	11	298	35.3	96.8	16.1	10.9	0.67
Infants <1 year			1	2	7	114	12.5	98.3	9.2	7.8	0.85
Tahirovic, 1988[105]	Urocomb (BM)	$\geq 10^5$ cfu ml^{-1}	16	1	59	230	21.3	99.6	42.6	49.3	0.79
	Urocomb (BM); negative samples incubated for 4 hours at 37°C		55	20	20	211	73.3	91.3	27.9	8.3	0.30
	Urocomb (BM); negative samples incubated with $NaNO_3$ for 4 hours at 37°C		70	5	27	204	72.2	97.6	95.3	27.5	0.29
Wammanda, 2000[110]	Multistix (Bayer)	$\geq 10^5$ cfu ml^{-1}	13	5	32	135	28.9	96.4	10.2	8.1	0.74
Weinberg, 1991[111]	Multistix (Ames)	$\geq 10^5$ cfu ml^{-1}	23	19	18	959	56.1	98.1	62.5	28.9	0.45
Wiggelinkhuizen, 1988[112]	Combur 9 (BM)	Not clear	96	9	57	800	62.7	98.9	141.4	46.1	0.38
	Multistix (Ames)		96	11	57	798	62.7	98.6	116.5	44.1	0.38

continued

TABLE 11 *Results of studies of dipstick tests (cont'd)*

Study; group	Index test manufacturer	Reference standard positive result	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Nitrite vs culture and microscopy											
Lagos, 1994[74]	Combur-9 (BM)	$\geq 10^5$ cfu ml^{-1} and >10 cells mm^{-3}	158	27	190	615	45.4	95.8	18.6	10.8	0.57
Lejeune, 1991[75]	Multistick (Ames); read by Clinitek	Various WBC ml^{-1} and $>10^5$ cfu ml^{-1}	6	5	31	201	16.2	97.6	7.6	6.7	0.86
LE vs culture											
Armengol, 2001[230]	Ames/BM	$\geq 10^4$ cfu ml^{-1}	21	5	9	225	70.0	97.8	92.8	32.2	0.31
Bulloch, 2000[47]	Clinitek (Ames)	Catheter $>10^4$, CVU $\geq 10^5$ cfu ml^{-1}	24	6	5	124	82.8	95.4	85.3	17.9	0.18
Dayan, 2002[53]	Automated urine analysis; Super UA	SPA $\geq 10^3$, catheter $\geq 10^4$ cfu ml^{-1}	16	10	4	163	80.0	94.2	57.1	13.8	0.21
Demi, 1993[55]	BM	NR	9	12	15	211	37.5	94.6	10.4	7.0	0.66
Giraldez, 1998[62]	NR	$\geq 10^5$ cfu ml^{-1}	25	2	0	23	100.0	92.0	479.4	12.5	0.02
Liptak, 1993[78]	Chemstrip 9 (BM)	$\geq 10^5$ cfu ml^{-1}	62	27	42	198	59.6	88.0	10.6	5.0	0.46
Lohr, 1993[81]	Ames; read by Clinitek	SPA $>10^3$, catheter	81	160	21	427	79.4	72.7	10.1	2.9	0.28
		$>10^4$, CVU $>10^5$ cfu ml^{-1}	43	140	10	408	81.1	74.5	12.0	3.2	0.25
Marsik, 1986[84]	Chemstrip (Biodynamics)	SPA any bacteria; catheter $>10^4$, CVU $>10^5$ cfu ml^{-1}									
Ordonez, 1994[89]	Ames (Multistix)	$\geq 10^5$ cfu ml^{-1}	13	23	5	59	72.2	72.0	6.2	2.6	0.39
Sharief, 1998[101]											
Total population	Multistix (Bayer); automated reader	$\geq 10^5$ cfu ml^{-1}	13	66	4	242	76.5	78.6	10.9	3.6	0.30
Infants <1 year			6	30	2	86	75.0	74.1	7.4	2.8	0.38
Weinberg, 1991[111]	Multistix (Ames)	$\geq 10^5$ cfu ml^{-1}	35	71	6	907	85.4	92.7	69.3	11.8	0.16
Wiggelinkhuizen, 1988[112]	Multistix (Ames)	Not clear	143	248	10	561	93.5	69.3	30.9	3.0	0.09
	Combur-9 (BM)		144	252	9	557	94.1	68.9	33.6	3.0	0.09
LE vs culture and microscopy											
Lagos, 1994[74]	Combur-9 (BM)	$\geq 10^5$ cfu ml^{-1} and >10 cells mm^{-3}	289	107	59	535	83.0	83.3	24.2	5.0	0.20
Lejeune, 1991[75]	Multistick (Ames); read by Clinitek	Various WBC ml^{-1} and $>10^5$ cfu ml^{-1}	33	45	4	161	89.2	78.2	26.4	4.1	0.14
Protein vs culture											
Boreland, 1986[45]	N-Labstix (Ames)	$\geq 10^5$ cfu ml^{-1}	27	109	40	360	40.3	76.8	2.2	1.7	0.78
	N-Labstix (Ames); read by Clinitek 200		8	24	7	125	53.3	83.9	5.8	3.3	0.56

continued

TABLE 11 *Results of studies of dipstick tests (cont'd)*

Study; group	Index test manufacturer	Reference standard positive result	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Protein vs culture and microscopy											
Lejeune, 1991[75]	Multistick (Ames); read by Clinitek	Various WBC ml^{-1} and $>10^5$ cfu ml^{-1}	3	10	34	196	8.1	95.1	1.9	1.8	0.96
Glucose vs culture											
Dosa, 1973[56]	Uriglox	$\geq 10^5$ cfu ml^{-1}	2	1	63	356	3.1	99.7	9.4	9.0	0.97
Group A											
Group B			7	0	4	25	63.6	100.0	85.0	32.5	0.36
Kohler, 1970[71]	Uriglox	$\geq 10^5$ cfu ml^{-1}	22	12	0	1982	100.0	99.4	7137.0	166.2	0.02
Schersten, 1968[99]	NR	$\geq 10^5$ cfu ml^{-1}	6	19	0	510	100.0	96.4	340.3	27.8	0.07
Todd, 1974[106]	Uriglox	Catheter $\geq 10^3$, CVU $\geq 10^5$ cfu ml^{-1}	55	5	1	574	98.2	99.1	3864.8	113.7	0.02
Blood vs culture											
Boreland, 1986[45]	N-Labstix (Ames)	$\geq 10^5$ cfu ml^{-1}	17	52	50	417	25.4	88.9	2.8	2.3	0.84
	N-Labstix (Ames); read by Clinitek 200		8	23	7	126	53.3	84.6	6.1	3.4	0.56
Nitrite or LE positive vs culture											
Anad, 2001[39]	Nephu-Test + Leuco (BM)	Culture; SPA any; other: >10 cfu ml^{-1}	34	95	15	346	69.4	78.5	8.1	3.2	0.39
Armengol, 2001[230]	Ames/BM	$\geq 10^4$ cfu ml^{-1}	21	5	9	225	70.0	97.8	92.8	32.2	0.31
Bulloch, 2000[47]	Clinitek (Ames)	Catheter $>10^4$, CVU $\geq 10^5$ cfu ml^{-1}	24	8	5	122	82.8	93.8	64.2	13.4	0.18
Dayan, 2002[53]	Automated urine analysis; Super UA	SPA $\geq 10^3$, catheter: $\geq 10^4$ cfu ml^{-1}	17	14	3	159	85.0	91.9	55.0	10.5	0.16
Liptak, 1993[78]	Chemstrip 9 (BM)	$\geq 10^5$ cfu ml^{-1}	79	36	25	189	76.0	84.0	16.2	4.7	0.29
Lohr, 1993[81]	Ames; read by Clinitek	SPA $>10^3$, catheter $>10^4$, CVU $>10^5$ cfu ml^{-1}	85	162	17	425	83.3	72.4	12.8	3.0	0.23
Marsik, 1986[84]	Chemstrip (Biodynamics)	SPA any bacteria; catheter $>10^4$, CVU $>10^5$ cfu ml^{-1}	47	156	6	392	88.7	71.5	18.3	3.1	0.16
Sharief, 1998[101]											
Infants <1 year	Multistix (Bayer); automated reader	$\geq 10^5$ cfu ml^{-1}	6	31	2	85	75.0	73.3	7.1	2.7	0.38
Total population			13	72	4	236	76.5	76.6	9.8	3.3	0.31
Shaw, 1991[103]											
<2 years	Multistix; at least trace LE	Catheter $\geq 10^3$, CVU $\geq 10^5$ cfu ml^{-1}	10	10	4	121	71.4	92.4	27.0	8.8	0.33
All children	Multistix; at least trace LE		37	80	8	366	82.2	82.1	20.1	4.6	0.22
	Multistix; at least small LE		36	58	9	388	80.0	87.0	25.5	6.1	0.24

continued

TABLE 11 *Results of studies of dipstick tests (cont'd)*

Study; group	Index test manufacturer	Reference standard positive result	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Shaw, 1998[102]	Multistix (Bayer); at least moderate LE	$\geq 10^4$ cfu ml^{-1}	69	33	26	3266	72.6	99.0	255.7	71.3	0.28
	Multistix (Bayer); at least trace LE		75	99	20	3200	78.9	97.0	118.5	26.3	0.22
Waisman, 1999[109]	Multistix (Bayer)	SPA, catheter >103, CVU/bag $>10^5$ cfu ml^{-1}	34	15	1	71	97.1	82.6	106.1	5.6	0.03
Weinberg, 1991[111]	Multistix (Ames)	$\geq 10^5$ cfu ml^{-1}	37	82	4	896	90.2	91.6	90.6	10.6	0.11
Wiggelinkhuizen, 1988[112]	Multistix (Ames)	Culture; not clear	144	249	9	560	94.1	69.2	34.2	3.1	0.08
	Combur-9 (BM)		146	253	7	556	95.4	68.7	42.9	3.0	0.07
Nitrite or LE positive vs culture and microscopy											
Lagos, 1994[74]	Combur-9 (BM)	$\geq 10^5$ cfu ml^{-1} and >10 cells mm^{-3}	309	129	39	513	88.8	79.9	31.1	4.4	0.14
Woodward, 1993[113]	Multistix (Bayer)	$\geq 10^5$ cfu ml^{-1} and $\geq$ 20 WBC mm^{-3}	12	14	0	107	100.0	88.4	185.3	8.6	0.04
Nitrite and LE positive vs culture											
Dayan, 2002[53]	Automated urine analysis; Super UA	SPA $\geq 10^3$, catheter $\geq 10^4$ cfu ml^{-1}	6	0	14	173	30.0	100.0	155.6	107.7	0.69
Giraldez, 1998[62]	NR	$\geq 10^5$ cfu ml^{-1}	22	0	3	25	88.0	100.0	327.9	45.0	0.14
Marret, 1995[83]	Combur-9 (BM)	Not clear	20	5	3	41	87.0	89.1	44.2	7.3	0.17
Marsik, 1986[84]	Chemstrip (Biodynamics)	SPA any bacteria; catheter $>10^4$, CVU $>10^5$ cfu ml^{-1}	33	12	20	536	62.3	97.8	70.1	27.2	0.39
Sharief, 1998[101]											
Infants <1 year	Multistix (Bayer); automated reader	$\geq 10^5$ cfu ml^{-1}	1	1	7	115	12.5	99.1	15.4	13.0	0.84
Total population			6	4	11	304	36.1	98.5	38.3	24.8	0.65
Shaw, 1991[103]											
<2 years	Multistix	Catheter $\geq 10^3$, CVU $\geq 10^5$ cfu ml^{-1}	2	3	12	128	14.3	97.7	7.3	6.3	0.86
All children			18	9	27	437	40.0	98.0	31.0	18.9	0.61
Wiggelinkhuizen, 1988[112]	Multistix (Ames)	Not clear	95	10	58	799	62.0	98.7	124.3	47.8	0.38
	Combur-9 (BM)		94	8	59	801	61.4	99.0	149.8	58.5	0.39
Nitrite and LE positive vs culture and microscopy											
Lejeune, 1991[75]	Multistick (Ames); read by Clinitek	Various WBC ml^{-1} and $>10^5$ cfu ml^{-1}	33	5	4	201	89.2	97.6	272.7	33.2	0.12
Woodward, 1993[113]	Multistix (Bayer)	$\geq 10^5$ cfu ml^{-1} and $\geq$ 20 WBC mm^{-3}	10	0	2	121	83.3	100.0	1020.6	197.1	0.19

continued

TABLE 11 *Results of studies of dipstick tests (cont'd)*

Study; group	Index test manufacturer	Reference standard positive result	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
LE and protein positive vs culture and microscopy											
Lejeune, 1991[75]	Multistick (Ames); read by Clinitek	Various WBC ml^{-1} and $>10^5$ cfu ml^{-1}	33	10	4	196	89.2	95.1	139.3	17.4	0.12
Nitrite, blood or protein positive vs culture											
Boreland, 1986[45]	N-Labstix (Ames); read by Clinitek 200	Culture; $\geq 10^5$ cfu ml^{-1}	14	38	1	111	93.3	74.5	28.0	3.5	0.13
	N-Labstix (Ames)		54	138	13	331	80.6	70.6	9.7	2.7	0.28
Nitrite, blood or LE positive vs culture											
Cervilla, 2001[49]	NR	SPA $\geq 10^2$, catheter $\geq 10^4$, CVU $\geq 10^5$ cfu ml^{-1}	36	39	7	20	83.7	33.9	2.5	1.3	0.50
Nitrite, blood and LE positive vs culture											
Craver, 1997[52]	Multistix (Ames): at least trace blood	Catheter $>10^3$, CVU/bag >50,000 cfu ml^{-1}	22	80	2	132	91.7	62.3	14.8	2.4	0.16
	Multistix (Ames): at least blood 1+		21	52	3	160	87.5	75.5	18.8	3.5	0.19
Nitrite, LE and protein positive vs culture											
Wiggelinkhuizen, 1988[112]	Multistix (Ames)	Not clear	80	5	3	391	96.4	98.7	1637.2	69.2	0.04
	Combur-9 (BM)		68	4	3	437	95.8	99.1	1902.8	93.4	0.05
Nitrite, LE and protein positive vs culture and microscopy											
Lejeune, 1991[75]	Multistick (Ames); read by Clinitek	Various WBC ml^{-1} and $>10^5$ cfu ml^{-1}	33	58	4	148	89.2	71.8	18.9	3.1	0.17
Nitrite, LE or protein positive vs culture											
Wiggelinkhuizen, 1988[112]	Multistix (Ames)	Not clear	150	418	3	391	98.0	48.3	40.2	1.9	0.05
	Combur-9 (BM)		150	372	3	437	98.0	54.0	50.5	2.1	0.04

NR, not reported.

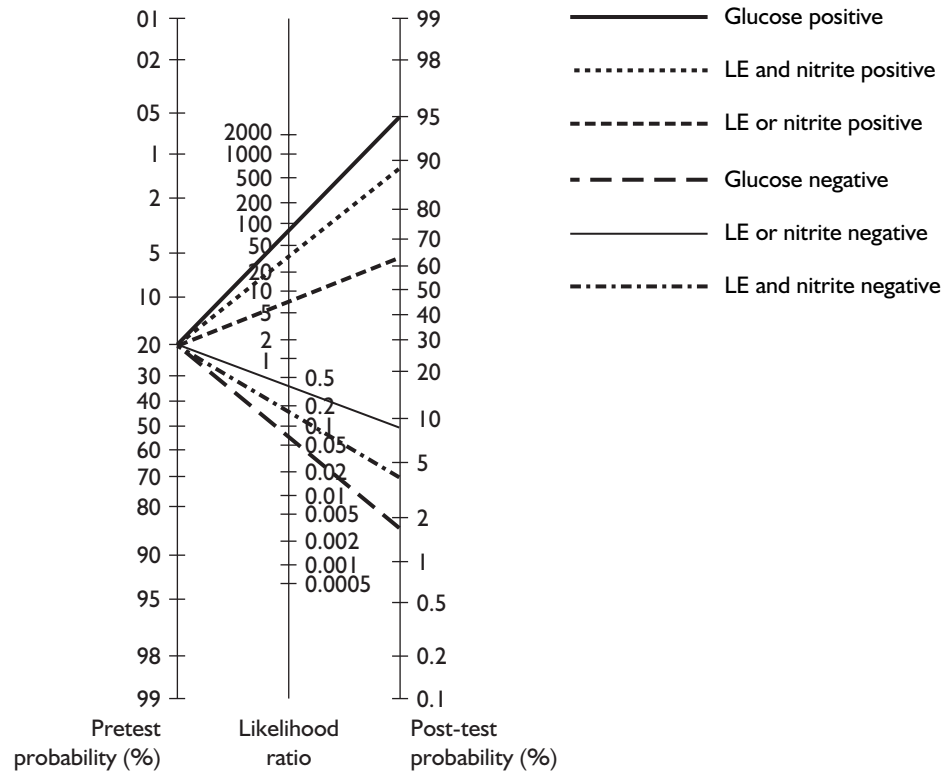

FIGURE 12 *Likelihood ratio monogram for dipstick tests*

probability of UTI. The reviewers were unable to find reliable estimates of the pretest probability of UTI in the literature; therefore, the results from the included studies were used to provide an estimate of the pretest probability of UTI. Only studies that included an appropriate patient spectrum were included in this analysis. UTI prevalence varied greatly between studies, ranging from 3 to 73%. As the distribution was highly skewed, the median, rather than the mean, UTI prevalence was used, which was 20%. *Figure 12* shows how the probability of UTI changes after the test has been given to give a post-test probability of disease in those with a positive and those with a negative test result.

Figure 13 uses an example to illustrate this. It takes a hypothetical 1000 children (numbers in italic) presenting to their GP with symptoms of UTI receiving a dipstick test for LE and nitrite. Based on the data from studies included in the review, on average, 110 would test positive for both LE and nitrite, 110 would test positive for LE or nitrite and negative for the other, and 780 would test negative for both LE and nitrite. Of the children testing positive for both LE and nitrite, 96 would have a UTI and 14 would not. Of the children testing negative for both LE and nitrite, 741 would not have a UTI and 39 would. Of those with an indeterminate test result (positive for either LE or nitrite), 65 would have a UTI and 45 would not. These numbers are calculated assuming a pretest probability of disease of 20%, a positive likelihood ratio of 28.2 and a negative likelihood ratio of 0.2.

Microscopy

A total of 39 studies reporting 101 data sets examined the accuracy of microscopy for diagnosing UTI. Microscopy was used for evidence of pyuria or bacteriuria, or combinations of the two. One study did not specify what microscopy parameter was used as the index test.[49] This study is not discussed further, but the results are recorded in *Table 17*. Some studies used standard microscopy, whereas others used automated microscopy. Culture was used as the reference standard in all studies, although different cut-off points were used to define a culture-positive result. One study used culture and microscopy as the reference standard,[108] where both tests were required to be positive to define positive UTI.

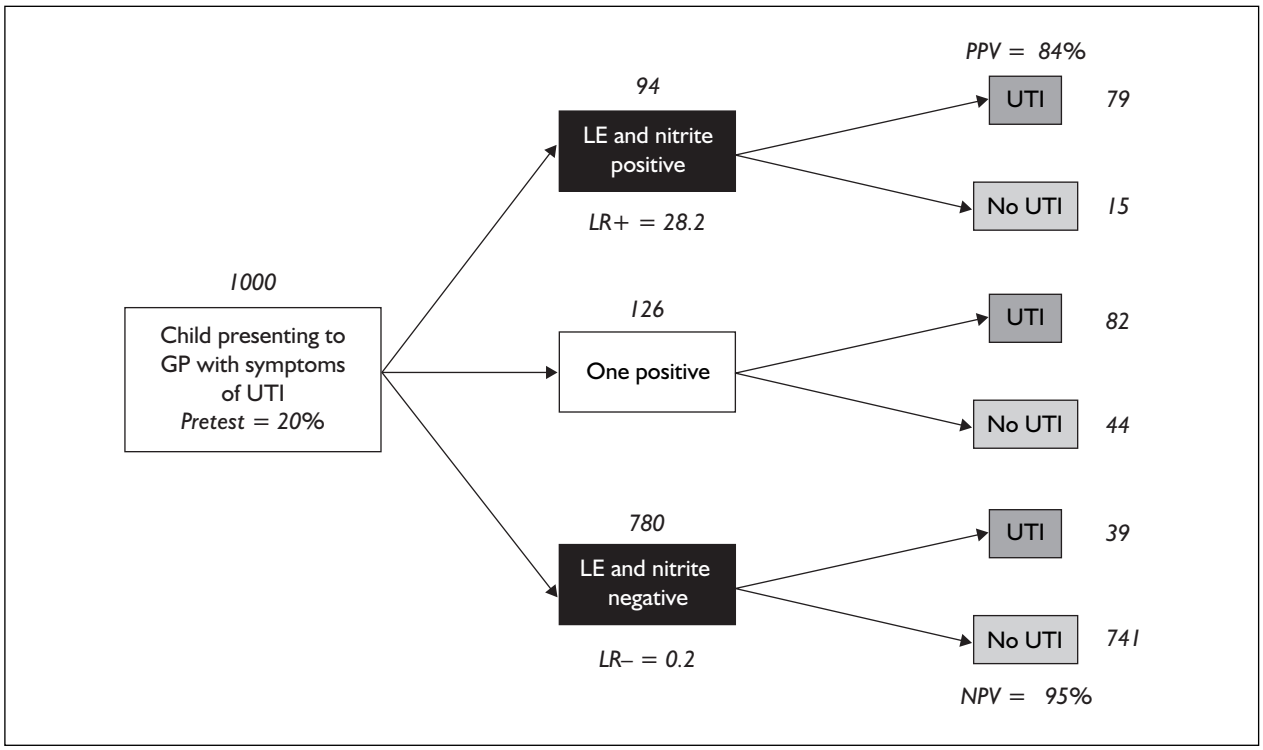

FIGURE 13 *Flowchart of an example of 1000 children being tested with dipstick for LE and nitrite. NPV, negative predictive value; PPV, positive predictive value.*

Pyuria

Twenty-seven studies reporting 49 data sets examined the microscopic detection of pyuria.[39,41,42,47,48,53,54,60,62,65–67,76,77,81,85,87,92,94,97,98,100,102,109–111,230] Only half of these studies included an appropriate spectrum of patients, and ten studies did not provide adequate description of the criteria used to select patients. The majority of the studies (25/28) did not report sufficient information to assess the avoidance of test review bias, and differential verification bias was a problem in one study. Approximately one-third of studies did not provide adequate descriptions of the index test and/or the reference standard. One study reported results stratified according to age,[41] a second reported results stratified according to age and antibiotic use,[100] and a third stratified results according to whether the urine sample was an early- or a late-stream sample.[54] As other studies did not make this distinction, the results from the subgroups were pooled to give an overall result for each test comparison from these studies. Several studies reported results for different cut-off points.[41,48,53,54,65,76,77,85,92,94,111] To prevent counting the same data twice, only one estimate from each of these studies was included in the analysis. Where possible, evaluations that reported a cut-off point of 10 WBC per high-power field (hpf) or per mm^3, the cut-off most commonly used by studies in this section, were retained. If studies did not report a result for this cut-off point then the evaluation with the most commonly reported alternative cut-off point was retained. One study reported results for three different methods of urine collection (all children had urine collected by all three methods).[65] The results were very similar for all three sampling methods. The evaluation retained in the analysis was that conducted on the SPA urine sample, as this is the accepted reference standard for urine sampling.

Sensitivity ranged from 36.6% (specificity 93%) to 96% (specificity 96%). Specificity ranged from 31.5% (sensitivity 89.4%) to 100% (sensitivity 50%).

Likelihood ratios showed considerable heterogeneity ($p < 0.001$). Positive likelihood ratios ranged from 1.3 (LR– 0.33) to 27.7 (LR– 0.09). Negative likelihood ratios ranged from 0.04 (LR+ 24.0) to 0.68 (LR+ 5.3). The pooled positive likelihood ratio was 5.9 (95% CI 4.1 to 8.5) and the pooled negative likelihood ratio was 0.27 (95% CI 0.20 to 0.37). *Figure 14* shows the estimates of sensitivity and 1 – specificity plotted in ROC space. This plot suggests that the considerable heterogeneity between studies is not just the result of different cut-off values, but is likely to be caused by other factors. Possible

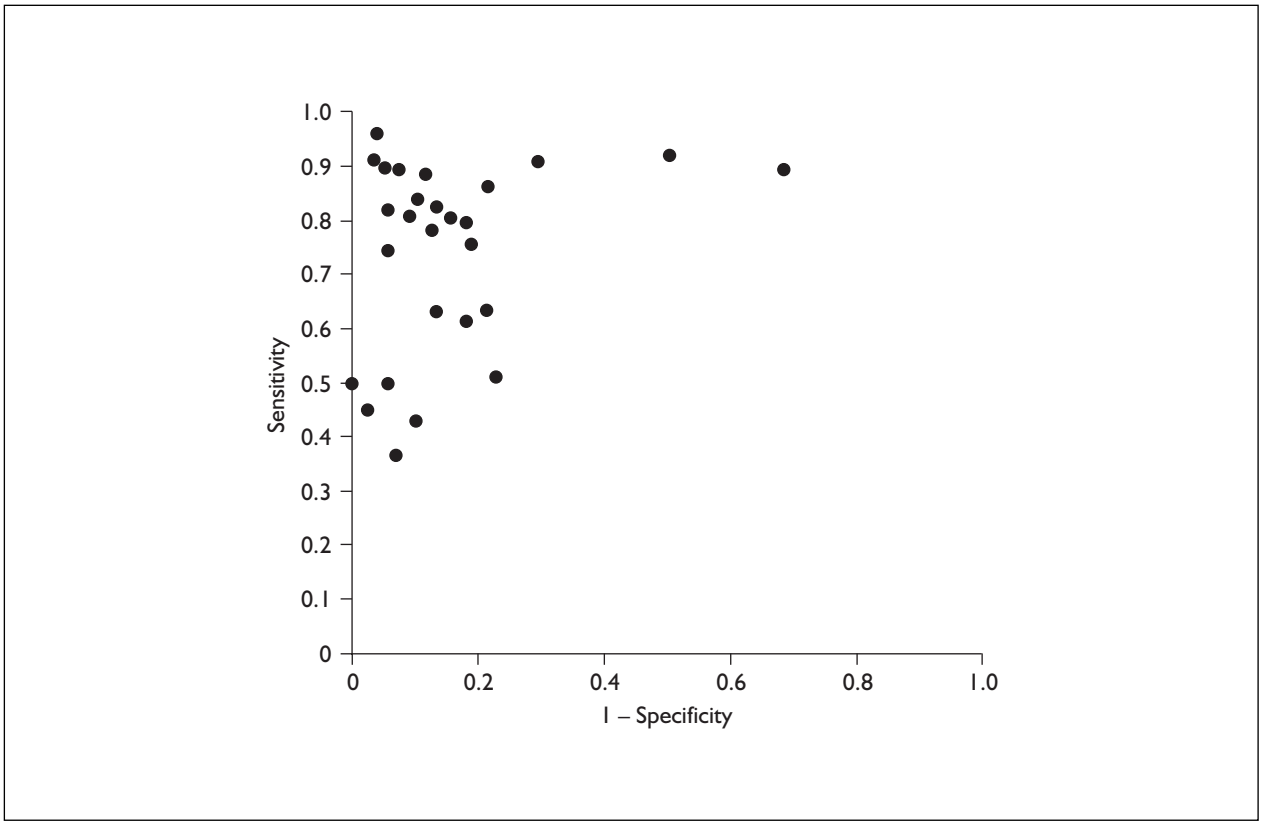

FIGURE 14 *Pyuria: study sensitivity and 1 – specificity plotted in ROC space*

explanations for the observed heterogeneity are discussed below. The median positive likelihood ratio was 6.2 (IQR 4.0–12.4) and the median negative likelihood ratio was 0.24 (IQR 0.17–0.47).

A regression analysis was carried out to investigate possible explanations for the observed heterogeneity. The regression model $D = \alpha + \beta S$ was extended to include variables for quality, age, region and whether samples were centrifuged. The results of the univariate regression analyses are shown in *Table 12*. The following items showed a significant association with D: centrifugation of the sample, description of selection criteria, test bias, review bias, description of study withdrawals and age. A stepwise multivariate regression analysis was then conducted (*Table 13*). Only two items remained significant in this analysis: centrifugation of the sample and the reporting of selection criteria. The DOR was 0.2 times less in centrifuged samples compared to non-centrifuged samples and was 3.3 times greater in studies that provided an adequate description of selection criteria.

Bacteriuria

Twenty-two studies reporting 34 data sets evaluated microscopic detection of bacteriuria.[40,42,47,53,54,60,66,69,79–82,85,87,93,94,97,102,107,108,111,230] Approximately half of theses studies did not include an appropriate spectrum of patients, and eight studies did not provide an adequate description of the criteria used to select participants. Only four studies reported whether or not investigators interpreting test results were blinded to the results of other tests (avoidance of test review bias), and differential verification bias and incorporation bias were problems in one study each. Approximately one-third of studies did not provide adequate descriptions of the index test and/or the reference standard. One study stratified results according to whether the urine was an early- or a late-stream catheter sample.[54] As other studies did not make this distinction, these data were pooled to give an overall result for each test comparison. This resulted in two data sets, one for microscopy of centrifuged unstained urine, and one for microscopy of uncentrifuged Gram-stained urine. The latter was used in all further analyses, as this method was most frequently used in other studies. One further study reported results for both centrifuged and uncentrifuged Gram-stained urine;[87] again, the data for uncentrifuged Gram-stained urine were used in further analyses. Four studies reported results using different cut-off points.[94,102,107,111]

TABLE 12 *Results of the regression analysis for microscopy for pyuria*

Variable		β	RDOR	*p*-Value	Adjusted r^2
Centrifuged		–1.4	0.2	0.005	0.46
Spectrum composition		–0.2	0.8	0.59	0.26
Selection criteria		0.9	2.5	0.057	0.36
Reference standard					
Time					
Partial verification					
Differential verification		–0.3	0.7	0.895	0.26
Incorporation					
Test details		0.7	2.0	0.371	0.28
Reference standard details		1.3	3.7	0.089	0.34
Test bias		1.6	5.0	0.008	0.44
Review bias		1.7	5.5	0.005	0.46
Clinical review bias		–1	0.4	0.153	0.31
Uninterpretable results		1.1	3.0	0.062	0.35
Withdrawals		1	2.7	0.059	0.36
Age:	< 2 years		0.56		
	< 5 years	–2.3	0.1	0.004	
	<12 years	–2.6	0.1	0.001	
	<18 years	–1.7	0.2	0.001	
Region:	North America		0.30		
	Europe	–1.1	0.3	0.19	
	Asia	–0.6	0.5	0.646	
	Other	–1.7	0.2	0.09	

RDOR, relative diagnostic odds ratio.

TABLE 13 *Results of the multivariate regression analysis for microscopy for pyuria*

Variable	Coefficient	RDOR	*p*-Value	Adjusted r^2
S	–0.2	0.8	0.047	0.64
Centrifuged	–1.7	0.2	<0.01	
Selection	1.2	3.3	0.001	
Constant	2.8	16.4	<0.01	

Where possible, evaluations that reported a cut-off point of 'any bacteria' (the cut-off most commonly used by studies in this section) were retained. If studies did not report a result for this cut-off point then the evaluation with the most commonly reported alternative cut-off point was retained. One study reported results for acridine orange stain using both diluted and undiluted urine. The results using undiluted urine were included in further analyses.[82]

Sensitivity ranged from 52.4% (specificity 98.7%) to 100% (specificity 98.1%). Specificity ranged from 40% (sensitivity 93.1%) to 99.7% (sensitivity 95.8%). Likelihood ratios showed considerably heterogeneity ($p < 0.001$). Positive likelihood ratios ranged from 1.6 (LR– 0.17) to 304.8 (LR– 0.04). Negative likelihood ratios ranged from 0.01 (LR+ 3.4) to 0.48 (LR+ 39.5). The pooled positive likelihood ratio was 14.7 (95% CI 8.7 to 24.9) and the pooled negative likelihood ratio was 0.19 (95% CI 0.14 to 0.24). *Figure 15* shows the estimates of sensitivity and 1 – specificity plotted in ROC space. This plot suggests that although different cut-off points may account for some of the heterogeneity in estimates of sensitivity and specificity, it is likely that other factors may be contributing to differences in estimates of test performance. The median positive likelihood ratio was 22.9 (IQR 5.5–46.3) and the median negative likelihood ratio was 0.19 (IQR 0.06–0.24).

A regression analysis was carried out to investigate possible explanations for the observed heterogeneity. *S* was not significant in the regression model $D = \alpha + \beta S$. Therefore, standard metaregression analysis using *D* as the dependent variable was performed. The following variables

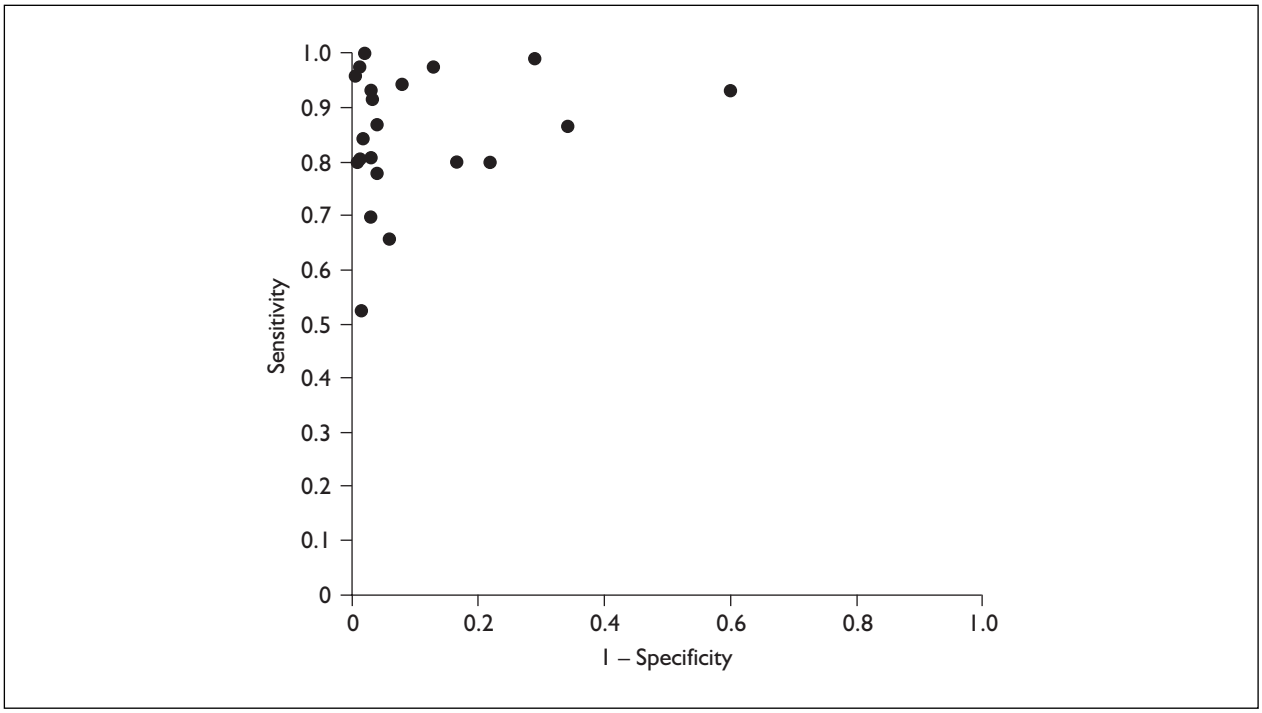

FIGURE 15 *Bacteriuria: study sensitivity and 1 – specificity plotted in ROC space*

were investigated: quality, age, region, whether samples were centrifuged and whether samples were Gram stained. The results of the univariate regression analyses are shown in *Table 14*. Gram stain and incorporation bias showed a significant association with D in the univariate analyses. Both items remained significant in the multivariate model (*Table 15*). The DOR was 5.5 times greater in samples that were Gram stained. In studies in which incorporation bias was not present, the DOR was 100 times greater than in studies in which it was a possibility.

Combinations of two microscopy tests

Presence of pyuria or bacteriuria

Eight studies reporting a total of ten data sets examined the combination of pyuria and bacteriuria, where a positive result from either test was taken as a positive result for UTI.[47,60,67,78,81,102,103,230] More than half of these studies (5/8) did not include an appropriate spectrum of patients, and two further studies did not provide an adequate description of the criteria used to select participants. The majority of studies (6/8) did not report sufficient information to assess the avoidance of test review bias, and differential verification bias was a problem in one study. One study reported results for all children and for children aged less than 2 years.[103] This study also reported results for two separate definitions of a positive index test: at least 5 WBC hpf^{-1} or few bacteria, and at least 10 WBC hpf^{-1} or moderate bacteria. Data for all children, using the cut-off point of at least 10 WBC hpf^{-1} or moderate bacteria, were retained for further analysis. Although thresholds of 10 and 5 WBC hpf^{-1} were equally used by studies in this group, 10 WBC hpf^{-1} was the most common cut-off for studies of pyuria in general.

Sensitivity ranged from 75% (specificity 92.9%) to 100% (specificity 32.3%). Specificity ranged from 32.3% (sensitivity 100%) to 92.9% (sensitivity 75%). Likelihood ratios showed considerable heterogeneity ($p<0.001$). Positive likelihood ratios ranged from 1.5 (LR– 0.05) to 12.9 (LR– 0.05). Negative likelihood ratios ranged from 0.02 (LR+ 2.8) to 0.27 (LR+ 4.1 and 10.5). The pooled positive likelihood ratio was 4.2 (95% CI 2.3 to 7.6) and the pooled negative likelihood ratio was 0.11 (95% CI 0.05 to 0.23). *Figure 16* shows the estimates of sensitivity and 1 – specificity plotted in ROC space. This plot suggests that the considerable heterogeneity between studies is not just the result of different cut-off values but is likely to be caused by other factors. No outliers were obvious from the ROC plot. The median positive likelihood ratio was 3.7 (IQR 2.7–7.1) and the median negative likelihood ratio was 0.08 (IQR 0.05–0.19).

TABLE 14 *Results of the regression analysis for microscopy for bacteriuria*

Variable		β	RDOR	*p*-Value	Adjusted r^2
Centrifuged		–1	0.4	0.179	0.04
Gram stain		1.3	3.7	0.014	0.23
Spectrum		0.6	1.8	0.247	0.02
Selection		0.8	2.2	0.14	0.06
Reference standard				Dropped	
Time		–2.6	0.1	0.188	0.04
Partial verification bias				Dropped	
Differential verification bias		–0.4	0.7	0.857	–0.05
Incorporation bias		–3.1	0.0	0.049	0.14
Test details		1.3	3.7	0.125	0.07
Reference details		1.4	4.1	0.106	0.08
Test bias		0.4	1.5	0.559	–0.03
Review bias		1.2	3.3	0.08	0.1
Clinical review bias		–0.2	0.8	0.829	–0.05
Uninterpretable results		0.7	2.0	0.213	0.03
Withdrawals		0.7	2.0	0.226	0.03
Region:	North America			Reference	–0.03
	Europe	–0.7	0.5	0.278	
	Asia	0.4	1.5	0.715	
	Other	0.4	1.5	0.555	
Age:	<2 years			Reference	–0.03
	< 5 years	–1.5	0.2	0.154	
	<12 years	–0.4	0.7	0.494	
	<18 years	–0.5	0.6	0.453	

TABLE 15 *Results of the multivariate regression analysis for microscopy for bacteriuria*

Variable	Coefficient	RDOR	*p*-Value	Adjusted r^2
Gram stain	1.7	5.5	<0.01	0.55
Incorporation bias	–4.4	0.01	0.001	
Constant	8.1	3294.5	<0.01	

There were insufficient studies to investigate heterogeneity using regression analysis. The studies were reasonably similar in terms of quality. Five of the eight studies did not include an appropriate patient spectrum, but this did not appear to have consistent effect on test performance.[47,67,78,81,230] Differential verification bias may have been a problem in one of the studies.[60]

Presence of pyuria and bacteriuria

Eight studies reporting a total of ten data sets evaluated the combination of pyuria and bacteriuria, where a positive result for both tests was taken as a positive result for UTI.[52,60,67–69,85,102,103] The majority (6/8) of these studies included an appropriate spectrum of patients. However, half of the studies did not adequately report the criteria used to select participants. The majority of studies (6/8) did not report sufficient information to assess the avoidance of test review bias, and differential verification bias was a problem in one study. One study reported results for all children and for children aged less than 2 years.[103] In this case only data for all children were retained and used for further analyses. Another study reported separate data for centrifuged and uncentrifuged Gram-stained urine samples; cut-off values used were at least 5 WBC hpf^{-1} and any bacteria, and at least 10 WBC mm^{-3}, respectively.[68] Data for uncentrifuged Gram-stained urine, using the cut-off at least 10 WBC mm^{-3}, were retained as these data were most similar to other studies in the group.

Sensitivity ranged from 46.7% (specificity 96%) to 93.1% (specificity 97.7%). Specificity ranged from 73.6% (sensitivity 70.8%) to 99.7% (sensitivity 84.4%). Likelihood ratios showed considerable heterogeneity ($p<0.001$). Positive likelihood ratios ranged from 2.7 (LR– 0.04) to 281 (LR– 0.16).

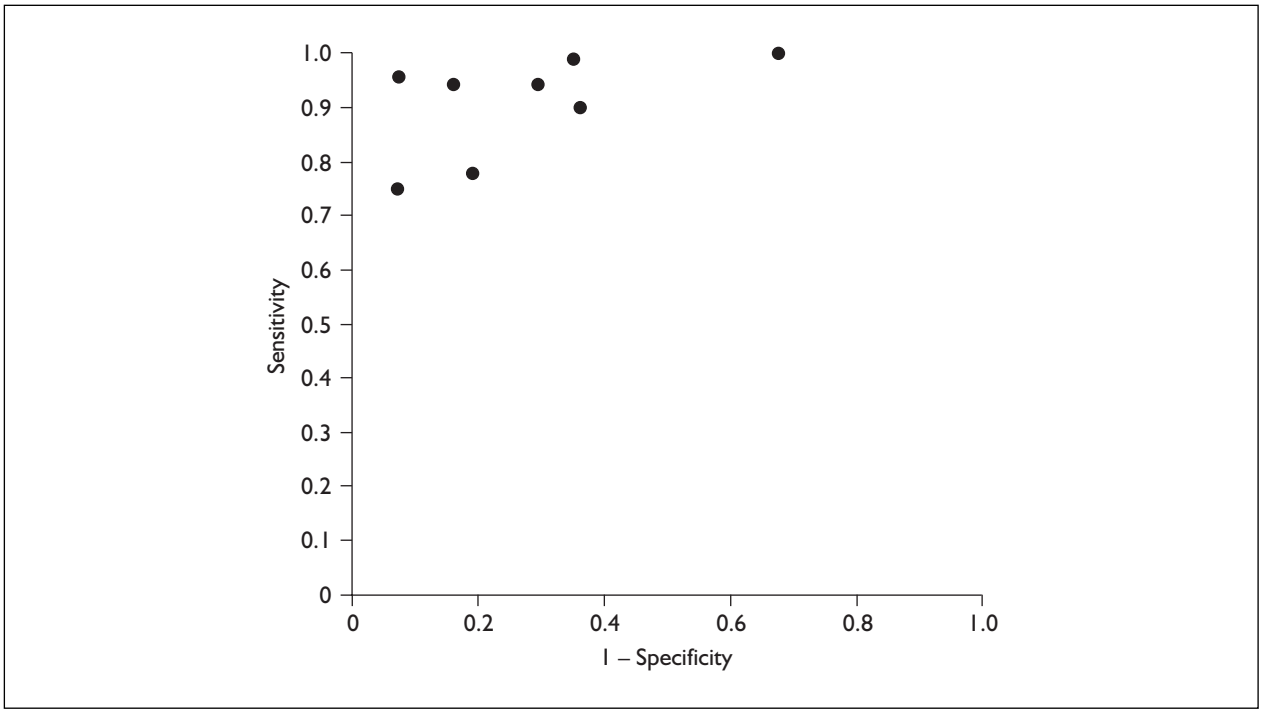

FIGURE 16 *Pyuria or bacteriuria: study sensitivity and 1 – specificity plotted in ROC space*

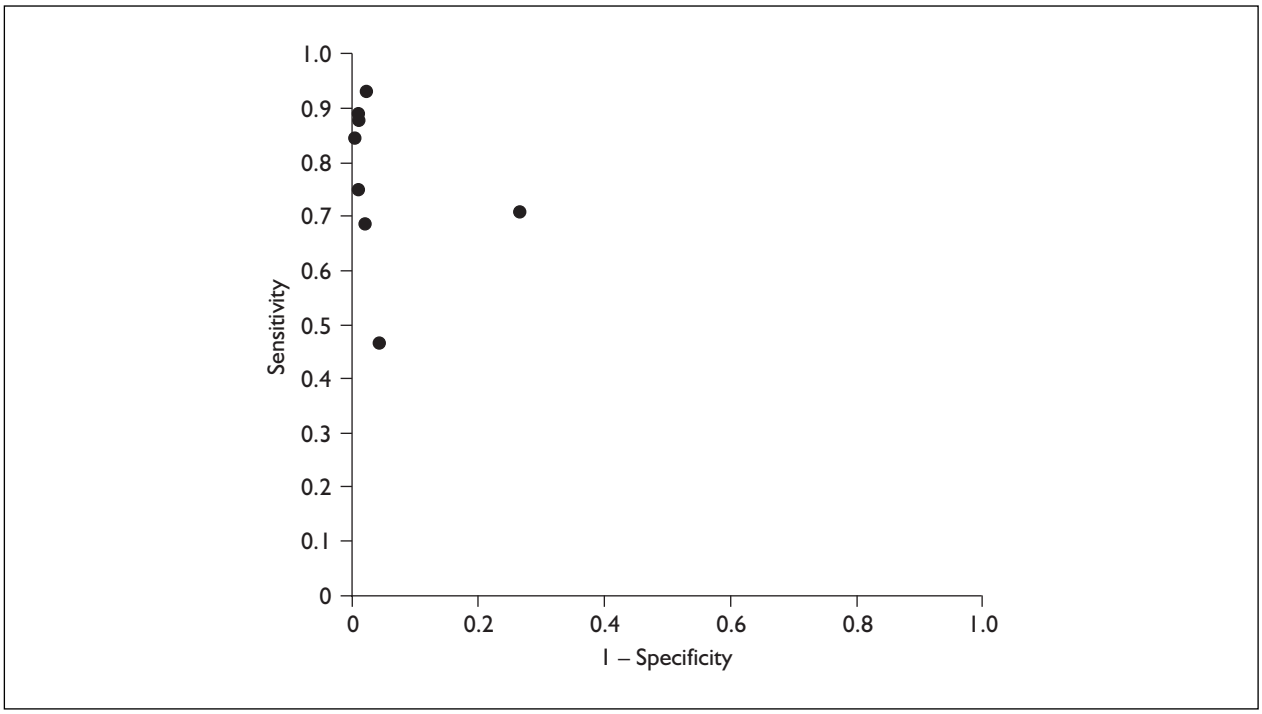

FIGURE 17 *Pyuria and bacteriuria: study sensitivity and 1 – specificity plotted in ROC space*

Negative likelihood ratios ranged from 0.07 (LR+ 41) to 0.56 (LR+ 11.6). The pooled positive likelihood ratio was 37.0 (95% CI 10.9 to 125.9) and the pooled negative likelihood ratio was 0.21 (95% CI 0.13 to 0.36). *Figure 17* shows the estimates of sensitivity and 1 – specificity plotted in ROC space. This plot suggests that the considerable heterogeneity between studies is not just the result of different cut-off values, but is likely to be caused by other factors. The plot indicated two apparent outliers.[52,103] These studies did not appear to differ from the other studies in

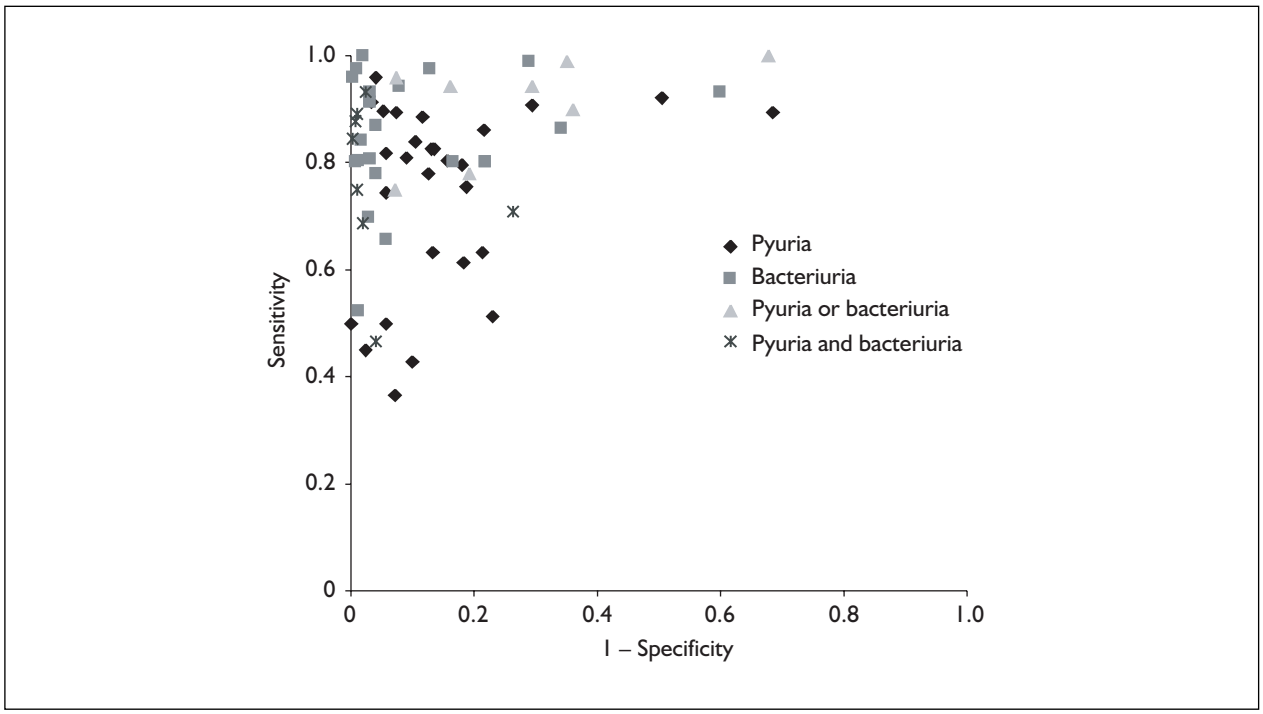

FIGURE 18 *Sensitivity and specificity plotted in ROC space for the different microscopy evaluations*

terms of quality or other factors.[52,103] The median positive likelihood ratio was 57.4 (IQR 28.6–94.2) and the median negative likelihood ratio was 0.21 (IQR 0.12–0.34).

There were insufficient studies to investigate heterogeneity using regression analysis. Six of the eight studies included an appropriate spectrum of patients.[52,60,69,85,102,103] This did not appear to be related to test performance. Differential verification bias may have been a problem in one study.[60] With this exception, study quality was reasonable in all studies.

Summary

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.

Figure 18 shows the estimates of sensitivity and 1 – specificity plotted in ROC space for all studies that evaluated the diagnostic accuracy of microscopic examinations for pyuria and bacteriuria, alone and in combination. This graph suggests that bacteriuria is considerably better than pyuria for ruling out disease and ruling in disease. The diagnostic performance of bacteriuria may be improved when combined with pyuria (both tests had to be positive to be defined as having a positive result). However, the confidence intervals around the pooled estimates are large, suggesting considerable uncertainty in these estimates. This is supported by the data in *Table 16*, that shows the pooled likelihood ratios for the different combinations of microscopy tests for pyuria and bacteriuria.

The pooled positive likelihood ratios are highest for pyuria and bacteriuria combined, supporting the suggestion that the combination of a positive result for both of these tests may be useful for ruling in disease. Pyuria alone, however, has a relatively poor positive likelihood ratio compared with bacteriuria alone or the two tests in combination, suggesting that it alone is not a useful test for ruling in disease. Both pyuria and bacteriuria, as well as the combination of both tests positive, have relatively poor negative likelihood ratios, suggesting that negative test results from these options are not useful for ruling out disease. The lowest likelihood ratio resulted from the combination of pyuria and bacteriuria, where a negative result was defined as both tests negative, and this may be useful for ruling out disease.

What do these results mean?

The results of studies of microscopy are summarised in *Table 17*. The results above indicate that a positive test for both pyuria and bacteriuria

TABLE 16 *Pooled likelihood ratios for microscopy*

Dipstick	Pooled LR+ (95% CI)	Pooled LR– (95% CI)
Pyuria	5.9 (4.1 to 8.5)	0.27 (0.20 to 0.37)
Bacteriuria	14.7 (8.6 to 24.9)	0.19 (0.14 to 0.24)
Pyuria or bacteriuria positive	4.2 (2.3 to 7.6)	0.11 (0.05 to 0.23)
Pyuria and bacteriuria positive	37.0 (11.0 to 125.9)	0.21 (0.13 to 0.36)

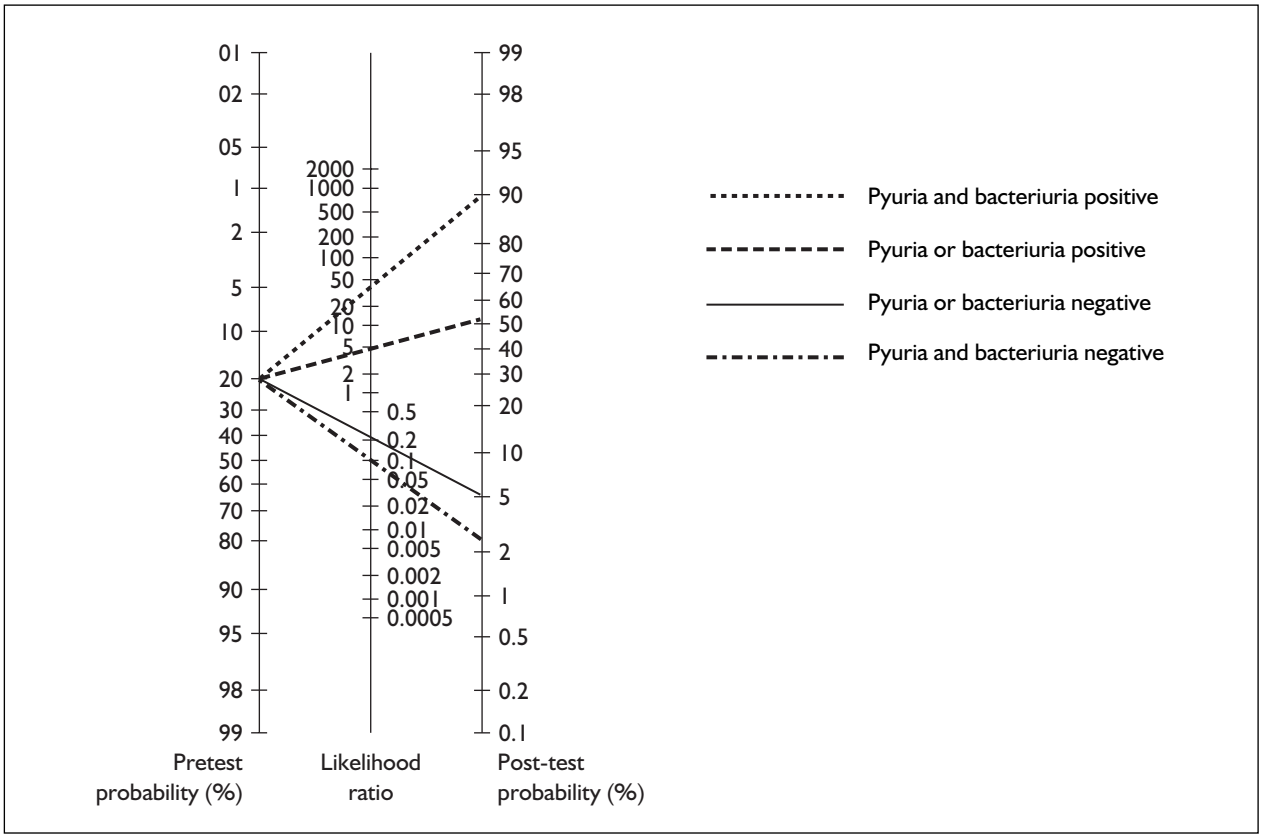

FIGURE 19 *Likelihood ratio monogram for microscopy*

is the best option for ruling in UTI using microscopy, while a negative result for both pyuria and bacteriuria is best for ruling out UTI. A result positive for either pyuria or bacteriuria, and negative for the other, may be seen as an indeterminate test result. To help to understand what these results mean, an estimate of the pretest probability of UTI and the likelihood ratios were used to calculate the post-test probability of UTI. The same estimate of the pretest probability of UTI was used as for the dipstick tests: 20%. *Figure 19* shows how the probability of UTI changes after the test has been given to give a post-test probability of disease in those with positive and negative test results.

Figure 20 uses an example to illustrate this. It takes a hypothetical group of 1000 children (numbers in italic) presenting to their GP with symptoms of UTI, receiving microscopy for pyuria and bacteriuria. Based on the data from studies included in the review, on average, 70 would test positive for both pyuria and bacteriuria, 140 would test positive for pyuria or bacteriuria and negative for the other, and 790 would test negative for both pyuria and bacteriuria. Of the children testing positive for both pyuria and bacteriuria, 63 would have a UTI and seven would not. Of the children testing negative for both pyuria and bacteriuria, 774 would not have a UTI and 16 would. Of those with an indeterminate test result (positive for pyuria or bacteriuria), 121 would have a UTI and 19 would not. Thus, quite a high proportion of these children would have a UTI and so in practice a positive result for either pyuria or bacteriuria could be treated similarly to

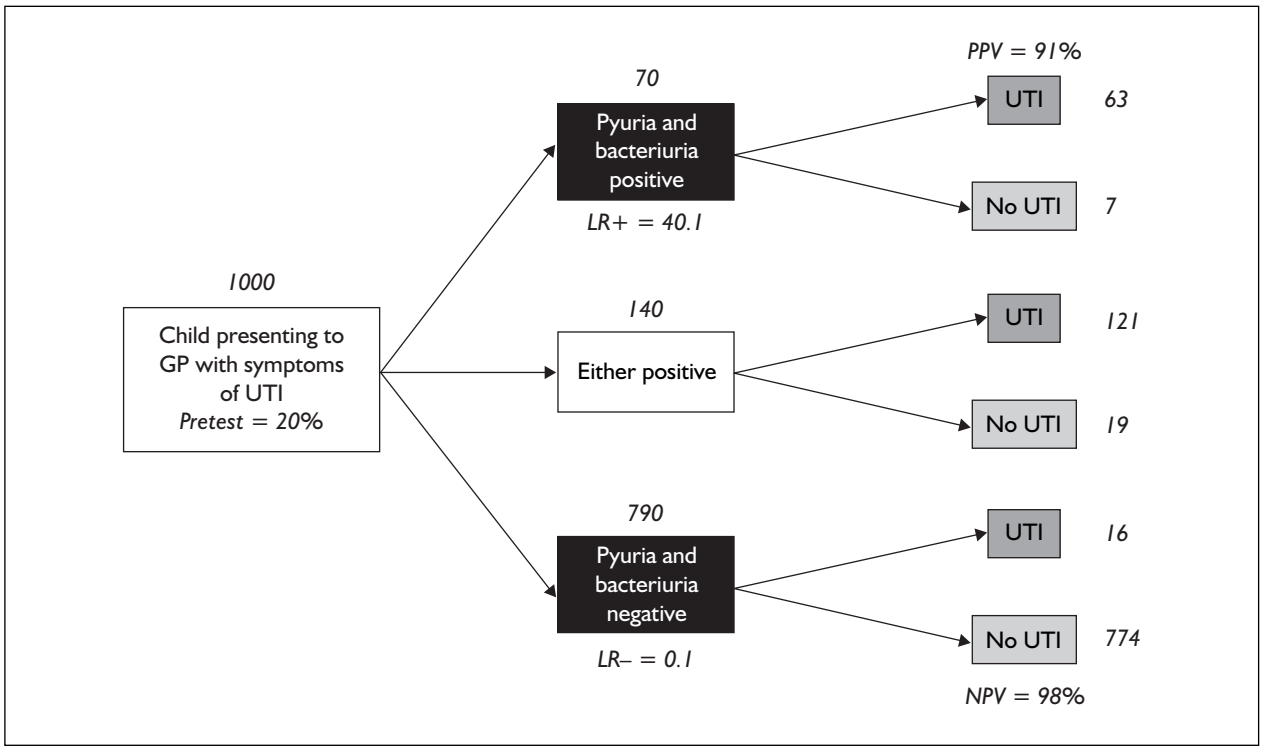

FIGURE 20 *Flow chart of an example of 100 children being tested with dipstick for pyuria and bacteriuria*

a positive result for both pyuria and bacteriuria. These numbers are calculated assuming a pretest probability of disease of 20%, a positive likelihood ratio of 40.1 and a negative likelihood ratio of 0.1.

Culture

Nine studies evaluated the accuracy of culture for the diagnosis of UTI[44,50,59,63,86,88,89,96,108] (*Table 18*). Eight studies examined dipslide cultures and one compared standard culture with a reference standard of microscopy and culture combined. This study found that culture was 100% sensitive and 93% specific.[108]

The eight studies that evaluated dipslide culture included a total of 11 data sets. Studies of this type were generally of poor quality and were poorly reported. Six of these studies did not use an appropriate spectrum of patients and six did not report the criteria used to select participants. Six studies did not provide an adequate description of the index test and/or the reference standard. The majority of studies (6/8) did not report sufficient information to assess avoidance of test review bias. Differential verification bias was a problem in three studies, and disease progression bias was also a problem in one of these. One study evaluated two different dipslide tests,[86] one examined results at 4 and 9 hours post-inoculation,[88] and one presented two analyses of the same results. In the latter study unsatisfactory samples were excluded in one analysis and in the other they were classed as negative.[96] The study that evaluated culture results after 4 and 9 hours reported a very low sensitivity of 18.6% after 4 hours. Therefore, only the results after 9 hours were included in the analysis. The reference standard in all studies was laboratory-based culture, and the cut-off point for a positive test result was 10^5 cfu ml^{-1}, based on the seven studies that provided this information.

Sensitivity ranged from 56.3% (specificity 96.5%) to 100% (specificity 91.8%). Specificity ranged from 70.7% (sensitivity 77.8%) to 100% (sensitivity 83.3%). The positive likelihood ratios ranged from 2.7 (LR– 0.31) to 135.4 (LR– 0.17). Negative likelihood ratios ranged from 0.02 (LR+ 12.18) to 0.46 (LR+ 7.8). There was considerable statistical heterogeneity in both positive and negative likelihood ratios ($p < 0.0001$). The pooled positive likelihood ratio was 14.6 (95% CI 6.7 to 31.8) and the pooled negative likelihood ratio was 0.23 (95% CI 0.14 to 0.39). However, these estimates should be interpreted with some degree of caution owing to the significant heterogeneity present. *Figure 21* shows the results from these studies plotted in ROC space. This shows the considerable heterogeneity across all studies with no clear outliers. Six studies are clustered towards the

TABLE 17 *Results of studies of microscopy*

Study	Sample	Index test positive result	Reference standard test positive result	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Pyuria vs culture												
Anad, 2001[39]	Uncentrifuged	Any WBC hpf^{-1}	SPA any, other >10 $cfu\ ml^{-1}$	37	83	12	358	75.5	81.2	12.9	4.0	0.30
Aronson, 1973[41]												
3–12 years	SPA	>10 WBC mm^{-3}	SPA ≥ 10^2 $cfu\ ml^{-1}$	13	6	1	14	92.9	70.0	20.1	2.9	0.14
	CVU	>250 WBC mm^{-3}		13	9	1	11	92.9	55.0	10.9	2.0	0.18
<18 mths	CVU	>250 WBC mm^{-3}		20	0	4	62	83.3	100.0	569.4	103.3	0.18
	SPA	>10 WBC mm^{-3}		21	0	3	62	87.5	100.0	767.9	108.4	0.14
Arslan, 2002[42]	Centrifuged	> 5 WBC	≥ 10^5 $cfu\ ml^{-1}$	30	3	40	27	42.9	90.0	5.9	4.3	0.63
Bulloch, 2000[47]	Centrifuged	≥ 5 WBC hpf^{-1}	Catheter >10^4 , CVU ≥ 10^5 $cfu\ ml^{-1}$	25	28	4	102	86.2	78.5	20.4	4.0	0.18
Caballero, 2001[48]	Centrifuged	Any WBC	≥ 10^5 $cfu\ ml^{-1}$	80	193	7	190	92.0	49.6	10.6	1.8	0.16
		>18 WBC		72	62	15	321	82.8	83.8	24.1	5.1	0.21
Dayan, 2000[54]	Early sample;	≥ 5 WBC/hpf	Catheter late-stream sample	4	20	2	59	66.7	74.7	5.2	2.5	0.48
	centrifuged	≥ 10 WBC/hpf	≥ 50,000 $cfu\ ml^{-1}$	3	5	3	74	50.0	93.7	13.5	7.3	0.54
	Late sample;	≥ 10 WBC/hpf		3	4	3	75	50.0	94.9	16.8	8.9	0.53
	centrifuged	≥ 5 WBC/hpf		5	10	1	69	83.3	87.3	24.3	6.0	0.25
Fernandez, 2000[60]	NS	NS	>50,000 $cfu\ ml^{-1}$	79	26	8	62	90.8	70.5	22.1	3.1	0.13
Giraldez, 1998[62]	Centrifuged	≥ 10 WBC hpf^{-1}	≥ 10^5 $cfu\ ml^{-1}$	24	1	1	24	96.0	96.0	266.8	24.0	0.04
Hardy, 1976[65]	CVU	>10 WBC mm^{-3}	SPA ≥ 10^4 $cfu\ ml^{-1}$	2	1	2	25	50.0	96.2	17.0	9.0	0.53
	Bag			1	1	3	25	25.0	96.2	7.3	5.4	0.74
	SPA			2	0	2	26	50.0	100.0	53.0	27.0	0.50
Hiraoka, 1995[66]	NS	≥ 10 WBC mm^{-3}	Catheter >10^3, CVU >10^5 $cfu\ ml^{-1}$	19	9	4	57	82.6	86.4	26.2	6.1	0.20
Hoberman, 1996[67]	Uncentrifuged	≥ 10 WBC mm^{-3}	≥ 50,000 $cfu\ ml^{-1}$	190	206	22	3835	89.6	94.9	157.3	17.6	0.11
		> 10 WBC mm^{-3}	≥ 50,000 $cfu\ ml^{-1}$	93	67	9	1969	91.2	96.7	287.2	27.7	0.09
Lin, 2000[76]	Uncentrifuged	≥ 10 WBC μl^{-1}	≥ 10^2 $cfu\ ml^{-1}$	18	8	4	132	81.8	94.3	64.1	14.3	0.19
	Centrifuged	≥ 5 WBC hpf^{-1}		13	10	9	130	59.1	92.9	17.7	7.9	0.45
Lin, 2000[77]	Uncentrifuged	≥ 10 WBC μl^{-1}	≥ 10^3 $cfu\ ml^{-1}$	31	20	6	173	83.8	89.6	41.0	8.1	0.18
	Centrifuged	≥ 5 WBC hpf^{-1}		24	23	13	170	64.9	88.1	13.2	5.3	0.40
Lohr, 1993[81]	Centrifuged	>5 WBC hpf^{-1}	SPA >10^3, catheter >10^4, CVU ≥ 10^5 $cfu\ ml^{-1}$	82	92	20	495	80.4	84.3	21.6	5.1	0.23
Matthai, 1995[85]	Centrifuged	>5 WBC hpf^{-1}	≥ 10^5 $cfu\ ml^{-1}$	190	50	36	100	84.1	66.7	10.4	2.5	0.24
		>20 WBC hpf^{-1}		49	9	77	141	38.9	94.0	9.5	6.2	0.65
		>10 WBC hpf^{-1}		180	27	46	123	79.6	82.0	17.4	4.4	0.25

continued

TABLE 17 *Results of studies of microscopy (cont'd)*

Study	Sample	Index test positive result	Reference standard test positive result	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Morton, 1982[87]	Uncentrifuged	>10 WBC mm^{-3}	SPA any; CVU >105 cfu ml^{-1}	43	32	25	210	63.2	86.8	11.0	4.7	0.42
Pryles, 1965[92]	Uncentrifuged	≥ 100 WBC mm^{-3}	≥ 10^5 cfu ml^{-1}	10	2	34	229	22.7	99.1	27.9	21.7	0.78
		≥ 10 WBC mm^{-3}		27	42	17	189	61.4	81.8	7.0	3.4	0.47
	Centrifuged	≥ 5 WBC mm^{-3}		19	6	25	235	43.2	97.5	27.7	16.1	0.58
Pylkkanen, 1979[94]	CVU	≥ 11 WBC mm^{-3}	SPA	178	50	21	23	89.4	31.5	3.9	1.3	0.33
		≥ 200 WBC mm^{-3}		117	16	82	57	58.8	78.1	5.0	2.6	0.53
Santos, 1982[97]	Centrifuged; stained	NS	NS	233	95	403	1269	36.6	93.0	7.7	5.3	0.68
Saxena, 1975[98]	Uncentrifuged	≥ 10 WBC mm^{-3}	NS	21	4	5	40	80.8	90.9	35.2	8.9	0.21
Schreiter, 1971[100]												
<1 year antibiotics	NS	>20 WBC μl^{-1}	≥ 10^5 cfu ml^{-1}	5	14	4	67	55.6	82.7	5.7	3.1	0.55
<1 year no antibiotics				51	7	14	237	78.5	97.1	112.5	25.5	0.23
1–14 years antibiotics				2	22	5	69	28.6	75.8	1.4	1.3	0.91
1–14 years no antibiotics				53	6	15	438	77.9	98.6	232.9	53.1	0.23
Shaw, 1998[102]	Uncentrifuged	≥ 10 WBC mm^{-3}	≥ 10^4 cfu ml^{-1}	47	278	10	1858	82.5	87.0	30.2	6.3	0.20
Wammanda, 2000[110]	Centrifuged	≥ 10 WBC hpf^{-1}	≥ 10^5 cfu ml^{-1}	23	32	22	108	51.1	77.1	3.5	2.2	0.63
Weinberg, 1991[111]	Centrifuged	≥ 5 WBC hpf^{-1}	≥ 10^5 cfu ml^{-1}	36	223	5	755	87.8	77.2	22.4	3.8	0.17
		≥ 10 WBC hpf^{-1}		32	122	9	856	78.0	87.5	23.9	6.3	0.25
Pyuria vs culture (automated microscopy)												
Armengol, 2001[230]	Centrifuged	≥ 5 WBC hpf^{-1}	≥ 10^4 cfu ml^{-1}	19	49	11	181	63.3	78.7	6.2	3.0	0.47
Dayan, 2002[53]	Uncentrifuged	≥ 5 WBC hpf^{-1}	SPA ≥ 10^3, catheter ≥ 10^4 cfu ml^{-1}	13	13	7	160	65.0	92.5	21.4	8.3	0.39
		≥ 10 WBC hpf^{-1}		9	4	11	169	45.0	97.7	31.1	19.4	0.56
Waisman, 1999[109]	Uncentrifuged (positives centrifuged and re-examined with standard microscopy)	>10 WBC hpf^{-1}	SPA ≥ 10^3, catheter ≥ 10^3, CVU/bag 10^5 cfu ml^{-1}	31	10	4	76	88.6	88.4	51.0	7.6	0.13

continued

TABLE 17 *Results of studies of microscopy (cont'd)*

Study	Sample	Index test positive result	Reference standard test positive result	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Bacteriuria vs culture												
Armengol, 2000[40]	NS	Any bacteria	$\geq 10^5$ 10^5 cfu ml^{-1}	24	50	6	180	80.0	78.3	13.5	3.7	0.26
Arslan, 2002[42]	Centrifuged; Gram stain	Any bacteria	$\geq 10^5$ 10^5 cfu ml^{-1}	56	5	14	25	80.0	83.3	18.1	4.8	0.24
Bulloch, 2000[47]	Centrifuged	Any bacteria	Catheter $>10^4$, CVU $\geq 10^5$ cfu ml^{-1}	27	78	2	52	93.1	40.0	7.4	1.6	0.17
Dayan, 2000[54]	Late sample; uncentrifuged; Gram stain	Any bacteria	Catheter late-stream sample $\geq$ 50,000 cfu ml^{-1}	6	1	0	78	100.0	98.7	680.3	49.5	0.07
	Late sample; centrifuged			6	10	0	69	100.0	87.3	86.0	7.1	0.08
	Early sample; centrifuged			4	16	2	63	66.7	79.7	6.9	3.1	0.45
	Early sample; uncentrifuged; Gram stain			6	2	0	77	100.0	97.5	403.0	29.7	0.07
Dayan, 2002[53]	Uncentrifuged; Gram stain	Any bacteria	SPA $\geq 10^3$, Catheter $\geq 10^4$ cfu ml^{-1}	16	1	4	172	80.0	99.4	421.7	138.4	0.20
Fernandez, 2000[60]	Uncentrifuged; Gram stain	NS	$>$50,000 cfu ml^{-1}	70	1	17	87	80.5	98.9	235.0	70.8	0.20
Hiraoka, 1995[66]	NS	NS	Catheter $>10^3$, CVU $>10^5$ cfu ml^{-1}	21	2	2	64	91.3	97.0	221.9	30.1	0.09
Hoberman, 1994[69]	Uncentrifuged; Gram stain	Any bacteria	$\geq$ 50,000 cfu ml^{-1}	95	61	7	1975	93.1	97.0	409.0	31.1	0.07
Littlewood, 1977[79]	Centrifuged	$\geq$ 10 bacteria hpf^{-1}	$\geq 10^5$ cfu ml^{-1}	33	6	5	145	86.8	96.0	136.3	21.9	0.14
Lockhart, 1995[80]	Uncentrifuged; Gram stain	Any bacteria	$\geq 10^3$ cfu ml^{-1}	17	15	1	174	94.4	92.1	131.3	11.3	0.09
Lohr, 1993[81]	Centrifuged; Gram stain	Any bacteria	SPA $>10^3$, catheter $>10^4$, CVU $>10^5$ cfu ml^{-1}	101	170	1	417	99.0	71.0	165.7	3.4	0.01
Matthai, 1995[85]	Centrifuged; unstained and Gram stain	NS	$\geq 10^5$ cfu ml^{-1}	176	6	50	144	77.9	96.0	77.7	19.5	0.23

continued

TABLE 17 *Results of studies of microscopy (cont'd)*

Study	Sample	Index test positive result	Reference standard test positive result	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Morton, 1982[87]	Uncentrifuged	>15 bacteria hpf^{-1}	SPA any, CVU >10^5 cfu ml^{-1}	30	22	38	220	44.1	90.9	7.8	4.8	0.61
	Centrifuged; Gram stain			25	10	24	200	51.0	95.2	19.9	10.2	0.52
	Uncentrifuged; Gram stain			11	2	10	149	52.4	98.7	65.5	39.5	0.48
Purwar, 1972[93]	Uncentrifuged; Gram stain	Any bacteria	≥ 10^4 cfu ml^{-1}	39	2	1	190	97.5	99.0	2006.6	93.6	0.03
Pylkkanen, 1979[94]	CVU	≥ 30 bacteria hpf^{-1}	SPA	83	5	116	68	41.7	93.2	8.9	5.6	0.63
		>1 bacteria hpf^{-1}		172	25	27	48	86.4	65.8	11.9	2.5	0.21
Santos, 1982[97]	Gram stain	NS	NS	278	29	52	1641	84.2	98.3	295.2	48.5	0.16
Shaw, 1998[102]	Uncentrifuged;	Single organism	≥ 10^4 cfu ml^{-1}	49	45	13	2198	79.0	98.0	177.2	38.8	0.22
	Gram stain	Any bacteria		50	67	12	2176	80.6	97.0	130.3	27.0	0.20
Vangone, 1985[107]	Uncentrifuged	>5 bacteria ml^{-1}	≥ 10^5 cfu ml^{-1}	50	28	26	457	65.8	94.2	30.6	11.4	0.36
		>10 bacteria ml^{-1}		47	20	29	465	61.8	95.9	36.6	14.6	0.40
Weinberg, 1991[111]	Uncentrifuged;	≥ 5 bacteria oif^{-1}	≥ 10^5 cfu ml^{-1}	37	22	4	956	90.2	97.8	354.3	38.8	0.11
	Gram stain	≥ 2 bacteria oif^{-1}		40	57	1	921	97.6	94.2	432.7	16.4	0.04
		≥ 1 bacteria oif^{-1}		40	86	1	892	97.6	91.2	278.6	10.9	0.04
		Any bacteria		40	124	1	851	97.6	87.3	184.7	7.7	0.03
Bacteriuria vs culture and microscopy												
Vickers, 1991[108]	NS	≥ 10^7 bacteria ml^{-1}	Both positive	23	1	1	317	95.8	99.7	3316.1	304.8	0.04
Bacteriuria vs culture (automated microscopy)												
Armengol, 2001[230]	Centrifuged	Any bacteria	≥ 10^4 cfu ml^{-1}	24	50	6	180	80.0	78.3	13.5	3.7	0.26
Manson, 1985[82]	Diluted urine; borate buffer + acridine orange	≥ 10^5 bacteria ml^{-1}	≥ 10^5 cfu ml^{-1}	43	152	3	1482	93.5	90.7	120.8	9.9	0.08
	Borate buffer + acridine orange			37	54	16	1786	69.8	97.1	74.5	23.8	0.31
Pyuria or bacteriuria vs culture												
Bulloch, 2000[47]	Centrifuged	≥ 5 WBC hpf^{-1} or any bacteria	Catheter >104, CVU ≥ 10^5 cfu ml^{-1}	29	88	0	42	100.0	32.3	28.3	1.5	0.05

continued

TABLE 17 *Results of studies of microscopy (cont'd)*

Study	Sample	Index test positive result	Reference standard test positive result	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Fernandez, 2000[60]	Uncentrifuged; Gram stain	Both positive	>50,000 cfu ml^{-1}	81	2	6	86	93.1	97.7	433.8	41.0	0.07
Hoberman, 1996[67]	Uncentrifuged; Gram stain	≥ 10 WBC mm^{-3} or any bacteria	>50,000 cfu ml^{-1}	203	300	9	3741	95.8	92.6	266.7	12.9	0.05
Liptak, 1993[78]	NS	≥ 10 WBC hpf^{-1} or any bacteria	≥ 10^5 cfu ml^{-1}	78	16	26	209	75.0	92.9	37.6	10.5	0.27
Lohr, 1993[81]	Centrifuged; Gram stain	≥ 5 WBC hpf^{-1} or any bacteria	SPA >10^3, catheter >10^4, CVU >10^5 cfu ml^{-1}	101	207	1	382	99.0	64.9	124.7	2.8	0.02
Shaw, 1998[102]	Uncentrifuged; Gram stain	≥ 10 WBC mm^{-3} or positive Gram stain	≥ 10^4 cfu ml^{-1}	49	314	3	1650	94.2	84.0	74.2	5.9	0.07
		≥ 10 WBC mm^{-3} and positive Gram stain	≥ 10^4 cfu ml^{-1}	39	20	13	1944	75.0	99.0	277.5	73.7	0.25
Shaw, 1991[103] All children	Centrifuged	≥ 5 WBC hpf^{-1} or few bacteria	Catheter ≥ 10^3, CVU ≥ 10^5 cfu ml^{-1}	43	246	2	200	95.6	44.8	14.2	1.7	0.12
		≥ 10 WBC hpf^{-1} and moderate bacteria	Catheter ≥ 10^3, CVU ≥ 10^5 cfu ml^{-1}	21	18	24	428	46.7	96.0	20.3	11.6	0.56
		≥ 10 WBC hpf^{-1} or moderate bacteria	Catheter ≥ 10^3, CVU ≥ 10^5 cfu ml^{-1}	35	85	10	361	77.8	80.9	14.3	4.1	0.27
<2 years		≥ 5 WBC hpf^{-1} or few bacteria	Catheter ≥ 10^3, CVU ≥ 10^5 cfu ml^{-1}	12	62	2	69	85.7	52.7	5.6	1.8	0.32
		≥ 10 WBC hpf^{-1} and moderate bacteria	Catheter ≥ 10^3, CVU ≥ 10^5 cfu ml^{-1}	5	3	9	128	35.7	97.7	21.3	13.8	0.65
Pyuria or bacteriuria vs culture (automated microscopy)												
Armengol, 2001[230]	Centrifuged	≥ 5 WBC hpf^{-1} or any bacteria	≥ 10^4 cfu ml^{-1}	27	83	3	147	90.0	63.9	13.9	2.5	0.16
Pyuria and bacteriuria vs culture												
Craver, 1997[52]	Centrifuged	>10 WBC and any bacteria	Catheter >10^3, CVU/bag >50,000 cfu ml^{-1}	17	56	7	156	70.8	73.6	6.5	2.7	0.40
Fernandez, 2000[60]	Uncentrifuged; Gram stain	Both positive	>50,000 cfu ml^{-1}	81	2	6	86	93.1	97.7	433.8	41.0	0.07
		Either positive		82	26	5	62	94.3	70.5	35.4	3.2	0.08

continued

TABLE 17 *Results of studies of microscopy (cont'd)*

Study	Sample	Index test positive result	Reference standard test positive result	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Hoberman, 1996[67]	Uncentrifuged; Gram stain	≥ 10 WBC mm^{-3} or any bacteria	>50,000 cfu ml^{-1}	203	300	9	3741	95.8	92.6	266.7	12.9	0.05
		≥ 10 WBC mm^{-3} and any bacteria		186	34	26	4007	87.7	99.2	817.5	104.3	0.12
Hoberman, 1993[68]	Uncentrifuged; Gram stain	≥ 10 WBC mm^{-3} and any bacteria	>50,000 cfu ml^{-1}	27	2	5	664	84.4	99.7	1329.0	281.0	0.16
	Centrifuged	≥ 5 WBC hpf^{-1} and any bacteria		21	5	11	661	65.6	99.2	224.9	79.0	0.35
Hoberman, 1994[69]	Uncentrifuged; Gram stain	≥ 10 WBC mm^{-3} and any bacteria	>50,000 cfu ml^{-1}	91	20	11	2016	89.2	99.0	782.7	90.8	0.11
Matthai, 1995[85]	Centrifuged; Gram stain	Both positive: >10 WBC hpf^{-1} for pyuria; not clear for others	≥ 10^5 cfu ml^{-1}	155	3	71	147	68.6	98.0	91.7	34.3	0.32
Shaw, 1998[102]	Uncentrifuged; Gram stain	≥ 10 WBC mm^{-3} or positive Gram stain	≥ 10^4 cfu ml^{-1}	49	314	3	1650	94.2	84.0	74.2	5.9	0.07
		≥ 10 WBC mm^{-3} and positive Gram stain	≥ 10^4 cfu ml^{-1}	39	20	13	1944	75.0	99.0	277.5	73.7	0.25
Shaw, 1991[103] All children	Centrifuged	≥ 5 WBC hpf^{-1} or few bacteria	Catheter ≥ 10^3, CVU ≥ 10^5 cfu ml^{-1}	43	246	2	200	95.6	44.8	14.2	1.7	0.12
		≥ 10 WBC hpf^{-1} and moderate bacteria	Catheter ≥ 10^3, CVU ≥ 10^5 cfu ml^{-1}	21	18	24	428	46.7	96.0	20.3	11.6	0.56
		≥ 10 WBC hpf^{-1} or moderate bacteria	Catheter ≥ 10^3, CVU ≥ 105 cfu ml^{-1}	35	85	10	361	77.8	80.9	14.3	4.1	0.27
<2 years		≥ 5 WBC hpf^{-1} or few bacteria	Catheter ≥ 10^3, CVU ≥ 10^5 cfu ml^{-1}	12	62	2	69	85.7	52.7	5.6	1.8	0.32
		≥ 10 WBC hpf^{-1} and moderate bacteria	Catheter ≥ 10^3, CVU ≥ 10^5 cfu ml^{-1}	5	3	9	128	35.7	97.7	21.3	13.8	0.65
Not stated vs culture												
Cervilla, 2001[49]	NS	NS	SPA ≥ 10^2, catheter ≥ 10^4, CVU ≥ 10^5 cfu ml^{-1}	36	39	7	20	83.7	33.9	2.5	1.3	0.50

hpf, high-power field; oif, oil immersion field.

TABLE 18 *Results of studies of culture*

Study	Test	Definition of positive result	Reference standard; details; definition of positive result	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Baum, 1972[44]	Dipslide culture	Not clear	Culture; NS	9	2	7	55	56.3	96.5	28.1	16.0	0.45
Cid, 1992[50]	Dipslide culture; Uricult-plus	Not clear	Culture; $\geq 10^5$ cfu ml^{-1}	17	11	3	69	85.0	86.3	30.2	6.2	0.17
Fennell, 1977[59]	Dipslide culture; Bacturcult	Not clear	Culture; $\geq 10^5$ cfu ml^{-1}	32	5	5	136	86.5	96.5	146.7	24.4	0.14
Godard, 1979[63]	Dipslide culture; home	$\geq 10^5$ cfu ml^{-1}	Culture; $\geq 10^5$ cfu ml^{-1}	29	22	0	246	100.0	91.8	646.4	12.2	0.02
Mongeau, 1972[86]	Dipslide culture; Uricult	$\geq 10^5$ cfu ml^{-1}	Culture; $\geq 10^5$ cfu ml^{-1}	17	1	1	81	94.4	98.8	633.9	77.4	0.06
	Dipslide culture; Testuria	>25 colonies		15	0	3	82	83.3	100.0	730.7	135.4	0.17
Navarrete, 1966[88]	Dipslide culture; Uroscreen; results after 4 hours	Not clear	Culture; $\geq 10^5$ cfu ml^{-1}	8	1	35	96	18.6	99.0	15.4	12.6	0.82
	Dipslide culture; Uroscreen; results after 9 hours			20	7	15	89	57.1	92.7	15.8	7.8	0.46
Ordonez, 1994[89]	Dipslide culture; Uricult plus	Not clear	Culture; $\geq 10^5$ cfu ml^{-1}	14	24	4	58	77.8	70.7	7.7	2.7	0.31
Rich, 1976[96]	Dipslide culture; home: unsatisfactory samples classed as negative	$\geq 10^5$ cfu ml^{-1}	Culture dipslide; $\geq 10^5$ cfu ml^{-1}	23	23	8	1226	74.2	98.2	144.3	40.3	0.26
	Dipslide culture; home; unsatisfactory samples excluded			23	23	3	884	88.5	97.5	252.7	34.9	0.12
Vickers, 1991[108]	Culture	$\geq 10^5$ cfu ml^{-1}	Microscopy and culture; both positive	24	21	0	297	100.0	93.4	678.0	14.5	0.02

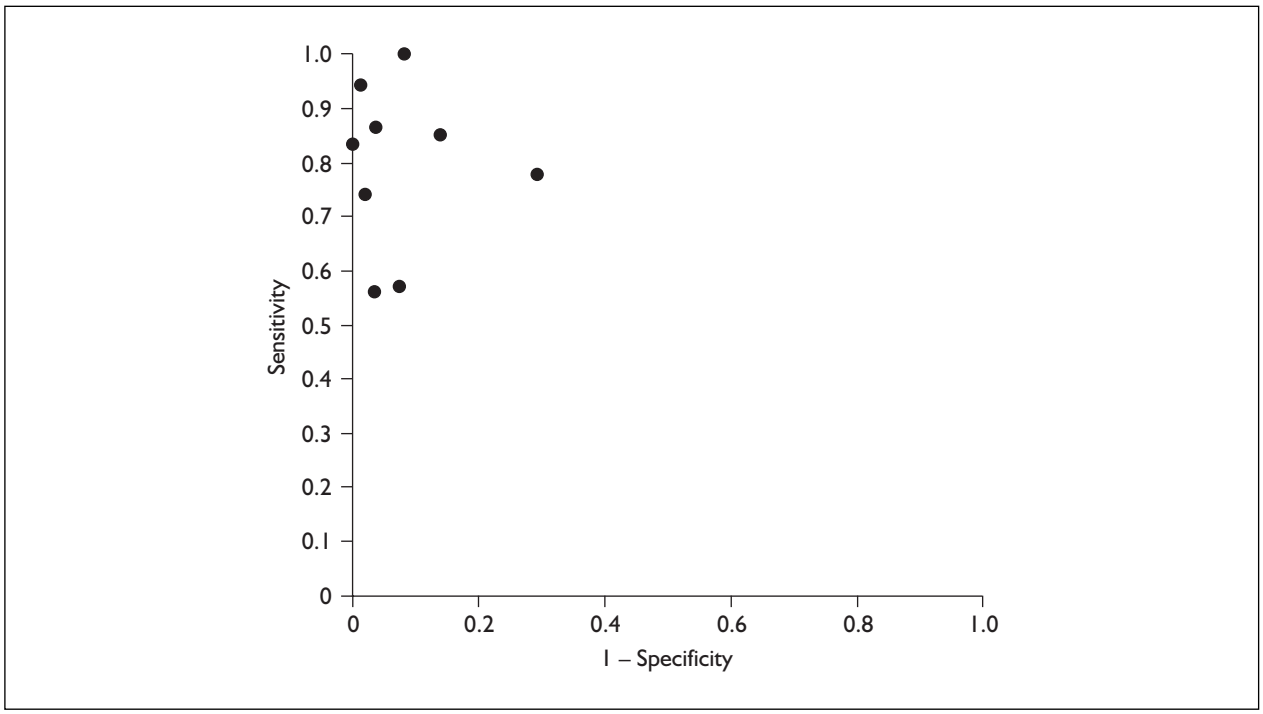

FIGURE 21 *Culture: study sensitivity and 1 – specificity plotted in ROC space*

upper left-hand corner of the graph suggesting a good diagnostic performance. The other three studies suggested poorer diagnostic performance.[44,88,89]

There were not enough studies to conduct a regression analysis to investigate possible explanations for the observed heterogeneity. The median positive likelihood ratio was 20.2 (IQR 8.9 –38.9). The median negative likelihood ratio was 0.2 (IQR 0.12–0.30).

In general, the studies were of relatively poor quality. Only two studies included an appropriate patient spectrum,[50,89] and in one the selection criteria were not clearly described.[89] Of the studies that did not include an appropriate spectrum of patients, three involved population screening,[59,63,96] three did not select patients on the basis of suspected UTI[86,88,108] and one provided no details.[44] In two studies it was unclear whether the reference standard used was appropriate.[59,63] Differential verification bias may have been a problem in both these studies and in one additional study.[96] Disease progression bias may also have been a problem in these three studies. There were no apparent differences between the studies that reported a good diagnostic performance of dipslide culture and those in which performance was poorer.

Other tests

Six studies examined other tests for the diagnosis of UTI[70,76,83,85,90,109] (*Table 19*). Studies of this type were generally of poor quality and were poorly reported. Only one study used an appropriate spectrum of patients, and reported the criteria used to select study participants. Two studies provided adequate descriptions of both the index test and the reference standard. No study in this category reported sufficient information to assess the avoidance of test review bias. A study published in 1968 examined the triphenyl-tetrazolium chloride (TTC) reduction test and the Greiss nitrate reduction test.[70] One study evaluated three laboratory-based blood tests for the diagnosis of UTI.[76] This study examined peripheral WBC levels, ESR and CRP levels. All were found to be fairly poor tests for the diagnosis of UTI. Other tests investigated included FiltraCheck-UTI™ for bacteriuria,[83] quantitative estimation of proteinuria[85] and Uriscreen™ test for catalase.[90,109] FiltraCheck-UTI and quantitative estimation of proteinuria both reported sensitivities and specificities of around 80%. The two studies of Uriscreen reported contrasting results. While one study[109] reported a sensitivity of 100% (specificity 69%), the other reported a sensitivity of only 65% (specificity 86%).[90] The cut-off point was the same in these two studies; thus, this cannot account for the observed differences.

TABLE 19 *Results of studies of other tests*

Study details	Test	Definition of positive result	Reference standard; positive result	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Holland, 1968[70]	TTC test	Formation of red precipitate	Culture; $\geq 10^5$ cfu ml^{-1}	20	1	6	345	76.9	99.7	726.4	175.6	0.24
	Greiss nitrate reduction test	Pink or red colour change	Culture; $\geq 10^5$ cfu ml^{-1}	26	2	0	42	100.0	95.5	901.0	17.7	0.02
Lin, 2000[76]	Peripheral WBC levels	>15000 µl^{-1}	Culture; $\geq 10^2$ cfu ml^{-1}	8	28	14	112	36.4	80.0	2.3	1.8	0.79
	ESR	>30 mm h^{-1}	Culture; $\geq 10^2$ cfu ml^{-1}	16	31	6	109	72.7	77.9	8.8	3.2	0.36
	CRP levels	>20 mg l^{-1}	Culture; $\geq 10^2$ cfu ml^{-1}	13	14	9	126	59.1	90.0	12.4	5.7	0.46
Marret, 1995[83]	FiltraCheck-UTI for bacteriuria	NS NS	Culture; NS	19	5	4	41	82.6	89.1	32.7	6.9	0.21
Matthai, 1995[85]	Quantitative estimation of proteinuria	NS	Culture; $\geq 10^5$ cfu ml^{-1}	177	30	49	120	78.3	80.0	14.2	3.9	0.27
Palmer, 1997[90]	Catalase test; Uriscreen	Formation of foam ring	Culture; $\geq$ 50,000 cfu ml^{-1}	30	22	16	132	65.2	85.7	10.9	4.5	0.41
Waisman, 1999[109]	Catalase test; Uriscreen	Formation of foam ring	Culture dipslide; SPA $>10^3$, catheter $>10^3$, CVU/bag $>10^5$ cfu ml^{-1}	35	27	0	59	100.0	68.6	153.6	3.1	0.02

Owing to the very small number of studies that examined these tests there was insufficient information to judge how useful these may be in the diagnosis of UTI.

Combinations of tests from different categories

Ten studies including a total of 20 data sets examined the accuracy of different combinations of tests for the diagnosis of UTI.[39,43,47,52,60,74,80,81,85,102] Half of these studies did not include an appropriate spectrum of patients, and a further three did not report the criteria used to select participants. The majority of the studies (8/10) did not report sufficient information to assess the avoidance of test review bias, and one of these also did not report sufficient information to assess avoidance of disease progression and partial verification bias; differential verification bias was a problem in one study. Given the results of individual tests, the test combinations that appear to be potentially the most interesting are dipstick for LE, and nitrite and microscopy for pyuria and bacteriuria. Five studies investigated different permutations of these tests.[42,47,80,81,102] Three studies evaluated the accuracy of a positive result in one of these four tests (i.e. dipstick positive for LE or nitrite or microscopy positive for pyuria or bacteriuria).[42,47,102] The results varied considerably between studies with sensitivity ranging from 67 to 100% and specificity from 3 to 87%. It is therefore not possible to draw overall conclusions from these studies. One study examined the combination of a positive result for all four tests.[102] This study reported an excellent specificity of 98%; that is, the combination was found to be very good for ruling in disease, but sensitivity was less good at 73%. This result might be expected, given the results from the studies that examined combinations of dipstick tests or combinations of microscopy tests.

The other test combinations evaluated by these studies differed widely, and none was repeated between studies. Test combinations investigated included combinations of microscopy (for pyuria and bacteriuria), dipstick tests (for LE, nitrite and blood), visual examination and quantitative estimation of proteinuria. As each test was only evaluated by one study it was not possible to draw conclusions regarding the diagnostic accuracy of these test combinations. The results of these studies are presented in *Table 20*.

Comparison of different tests

The results suggest that the best tests for ruling in disease were dipstick positive for nitrite and LE or microscopy positive for bacteriuria and pyuria. Conversely, the best tests for ruling out disease were dipstick negative for nitrite and LE or microscopy negative for bacteriuria and pyuria. Comparison of the pooled likelihood ratios of these tests suggests that the microscopy combinations were more accurate than the dipstick combinations. Further analysis was conducted to investigate the statistical significance of this difference.

Only one study evaluated both dipstick positive for nitrite and LE and microscopy positive for bacteriuria and pyuria.[103] This study found that the dipstick combination was best for ruling in disease: it found a higher positive likelihood ratio for dipstick positive for LE and nitrite than for microscopy positive for bacteriuria and pyuria (18.9 versus 11.6). Five studies examined dipstick negative for nitrite and LE and microscopy negative for both pyuria and bacteriuria.[47,78,81,102,103] All but one of these studies found that microscopy negative for pyuria and bacteriuria was better for ruling out disease than dipstick negative for nitrite and LE. The one study that found that the dipstick combination was better for ruling out disease reported very similar negative likelihood ratios for both test combinations.[103] The statistical significance of these differences was not formally assessed.

Accuracy of tests used for the further investigation of UTI

A total of 105 studies that evaluated the diagnostic accuracy of tests for the further investigation of UTI met the inclusion criteria. These included 215 test evaluations, investigating tests for the localisation of UTI, detection of reflux, prediction and detection of renal scarring, and detection of anatomical abnormalities.

Localisation of UTI

Thirty-seven studies reporting 82 data sets evaluated the diagnostic accuracy of tests for the localisation of UTI. Of the studies included in this section, 22 used acute ^{99m}Tc-DMSA renal scintigraphy as the reference standard.

Clinical

Five studies assessed the utility of various clinical features for the localisation of UTI.[125,129,140,143,168] Two of these studies did not include an appropriate spectrum of patients. Three studies did not report sufficient information to assess the avoidance of disease progression bias, and three

TABLE 20 *Results of studies of combinations of tests*

Study	Index test positive result	Reference standard positive result	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Dipstick and microscopy vs culture											
Nitrite and pyuria											
Fernandez, 2000[60]	Either positive	>50,000 cfu ml^{-1}	81	28	6	60	93.1	68.2	26.6	2.9	0.11
	Both positive		75	2	12	86	86.2	97.7	209.0	30.5	0.15
LE, nitrite, and pyuria											
Anad, 2001[39]	LE or nitrite and pyuria	SPA; any; other >10 cfu ml^{-1}	28	50	6	313	82.4	86.2	27.2	5.9	0.22
Bachur, 2001[43]	Any one positive	SPA ≥ 10^3, catheter ≥ 10^4, CVU ≥ 10^5 cfu ml^{-1}	578	640	127	7355	82.0	92.0	52.1	10.2	0.20
Lohr, 1993[81]	LE and nitrite positive, or pyuria	SPA >103, catheter >10^4, CVU >10^5 cfu ml^{-1}	90	166	12	421	88.2	71.7	18.3	3.1	0.17
LE, nitrite and bacteriuria											
Lohr, 1993[81]	LE and nitrite positive, or any bacteria	SPA >10^3, catheter >10^4, CVU >10^5 cfu ml^{-1}	102	249	0	338	100.0	57.6	278.1	2.3	0.01
LE, nitrite, blood, and pyuria											
Craver, 1997[52]	All positive: LE ≥ trace, nitrite positive blood ≥ 1+, and WBC>10	Catheter >10^3, CVU/bag >50,000 cfu ml^{-1}	21	85	3	127	87.5	59.9	9.2	2.1	0.23
	All positive: LE ≥ trace, nitrite positive blood ≥ trace, and WBC>10		22	109	2	103	91.7	48.6	8.5	1.8	0.21
LE, nitrite, pyuria and bacteriuria											
Bulloch, 2000[47]	≥ 5 WBC hpf^{-1} or any bacteria or ≥ small LE or nitrite	Catheter >10^4 cfu ml^{-1}, CVU ≥ 10^5 cfu ml^{-1}	29	87	0	43	100.0	33.1	29.3	1.5	0.05
Lohr, 1993[81]	Not clear	SPA >10^3, catheter >10^4, CVU >10^5 cfu ml^{-1}	102	234	0	353	100.0	60.1	309.0	2.5	0.01
Shaw, 1998[102]	Any one positive: ≥ 5 WBC hpf^{-1} or any bacteria	≥ 10^4 cfu ml^{-1}	79	429	16	2870	83.2	87.0	32.2	6.4	0.20
	Dipstick positive and ≥ 5 WBC hpf^{-1} and any bacteria		69	66	26	3233	72.6	98.0	127.5	35.9	0.28
Lockhart, 1995[80]	Any positive: ≥ trace LE, ≥ 5 WBC hpf^{-1}, ≥ 'slight' bacteria	≥ 10^3 cfu ml^{-1}	12	40	6	149	66.7	78.8	7.1	3.1	0.43
Arslan, 2002[42]	Any one positive	≥ 10^5 cfu ml^{-1}	52	29	18	1	74.3	3.3	0.1	0.8	5.38

continued

TABLE 20 *Results of studies of combinations of tests (cont'd)*

Study		Index test positive result	Reference standard positive result	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Dipstick and visual examination vs culture and microscopy												
Lagos, 1994[74]	LE, nitrite and not limpid	Any one positive	$\geq 10^5$ cfu ml^{-1} and >10 cells mm^{-3}	325	220	23	422	93.4	65.7	26.5	2.7	0.10
	LE and not limpid	Either positive		325	201	23	441	93.4	68.7	30.3	3.0	0.10
Microscopy and quantitative estimation vs culture												
Matthai, 1995[85]	Proteinuria and bacteriuria	Both positive	$\geq 10^5$ cfu ml^{-1}	146	2	80	148	64.6	98.7	108.1	39.0	0.36
	Pyuria and proteinuria	Both positive: >10 WBC hpf^{-1} for pyuria		159	17	67	133	70.4	88.7	18.0	6.1	0.34
	Pyuria, bacteriuria and proteinuria	Any two positive: >10 WBC hpf^{-1} for pyuria not clear for others		184	20	42	130	81.4	86.7	27.6	6.0	0.22
		Any one positive: >10 WBC hpf^{-1} for pyuria not clear for others		207	45	19	105	91.6	70.0	24.7	3.0	0.12
		All three positive		138	1	88	149	61.1	99.3	156.0	61.4	0.39

did not report sufficient information to assess avoidance of review bias.

Two studies examined the accuracy of body temperature for the diagnosis of acute pylonephritis (APN). Both used ^{99m}Tc-DMSA renal scintigraphy as the reference standard.[125,143] Test performance was poor in both studies, with one reporting a sensitivity of 64% and a specificity of 40%, and the other reporting a sensitivity of 87% and a specificity of 64% for cut-off points of 39.1°C and 38°C, respectively.

Two studies evaluated the diagnostic accuracy of symptoms of APN; both used ^{99m}Tc-DMSA renal scintigraphy as the reference standard.[129,140] These found relatively poor sensitivities of 57%[140] and 71%,[129] which corresponded to 100% specificity in both studies. Both were of reasonable quality, although one did not include an appropriate spectrum of patients.[140] One study assessed the presence of physical symptoms or positive laboratory findings for the diagnosis of APN.[168] This study also used ^{99m}Tc-DMSA renal scintigraphy as the reference standard. Sensitivity was higher in this study at 98%, but specificity was only 33%. In general, the clinical features used and the methods of determination were too diverse and poorly described to allow any conclusions to be drawn regarding the value of clinical examination in making a diagnosis of APN.

Laboratory-based tests

Approximately one-third of all studies in this category did not include an appropriate spectrum of patients, and a similar number did not report the criteria used to select participants. One-third of studies did not use an appropriate reference standard to confirm diagnosis, or did not report sufficient detail to judge this. Half of all studies in this section did not report sufficient information to assess avoidance of disease progression bias, and half of studies did not report sufficient information to assess avoidance of test review bias. Less than half of the studies (7/16) adequately described both the index test and the reference standard. The accuracy of circulatory CRP for diagnosing APN, at various concentrations, was assessed in seven studies.[125,129,143,145,146,204,208] All of these studies used acute ^{99m}Tc-DMSA renal scintigraphy as the reference standard. Three studies used a concentration of 20 mg ml^{-1} to define a positive result;[146,204,208] all reported high sensitivity (above 85%), but poor specificity (between 19 and 60%). The remaining studies of CRP used widely varying definitions of a positive result (20 μg l^{-1} to 880 mg l^{-1}) and generally reported poor diagnostic performance. For higher concentrations, sensitivity ranged from 65 to 70% and specificity from 55 to 68%. One study[129] using a very low concentration of 20 μg l^{-1} to define a positive result reported a specificity of 100%; however, sensitivity was very poor at 14%.

Other laboratory analytes evaluated were β_2M, *N*-acetyl-β-glucosaminidase NAG, NAG/creatinine ratio, PCT, polymorphonuclearelastase-α_1-antitrypsin complex and urinary α_1-microglobulin/creatinine ratio. Given the small number of studies using each analyte, along with the diverse methodologies and cut-off points, it is not possible to draw any conclusions regarding the utility of these tests in making the diagnosis of APN.

ESR, immunofluorescent detection of bacteria and various microscopic evaluations were also assessed for the diagnosis of APN. Again, the small number of studies using each test, and variety of test methods and reference standard tests used, make it impossible to reach conclusions on the potential utility of these tests.

Ultrasound

The primary imaging technique evaluated for the diagnosis of APN was ultrasound. The diagnostic accuracy of ultrasound was assessed in 20 studies.[116,121,125–127,134,146,148,150,155–157,160,166,169,177,183,201,205,211] Renal scintigraphy was the reference standard in 18 of these studies,[116,121,125–127,146,148,150,155–157,160,166,169,177,183,201,205] with acute ^{99m}Tc-DMSA used in 14 of these. Of the 18 ultrasound studies that used an appropriate reference standard, ten did not use an appropriate spectrum of patients,[116,121,127,148,150,155–157,177,183] and four did not describe the criteria used to select patients.[116,126,148,201] Only six studies provided an adequate description of both the index test and the reference standard.[121,150,155,160,169,183] Eleven studies did not report sufficient information to assess avoidance of test review bias,[116,127,146,148,150,156,157,166,169,201,205] and six did not report sufficient information to assess avoidance of disease progression bias.[116,125,146,155,177,201] The remaining two evaluations of ultrasound techniques used inappropriate reference standard tests of clinical and laboratory diagnosis of APN[211] and computed tomography (CT).[134] These studies are not considered in the analysis below.

One study[121] reported data separately for different age groups and for first or multiple UTI. To prevent the same population being included more

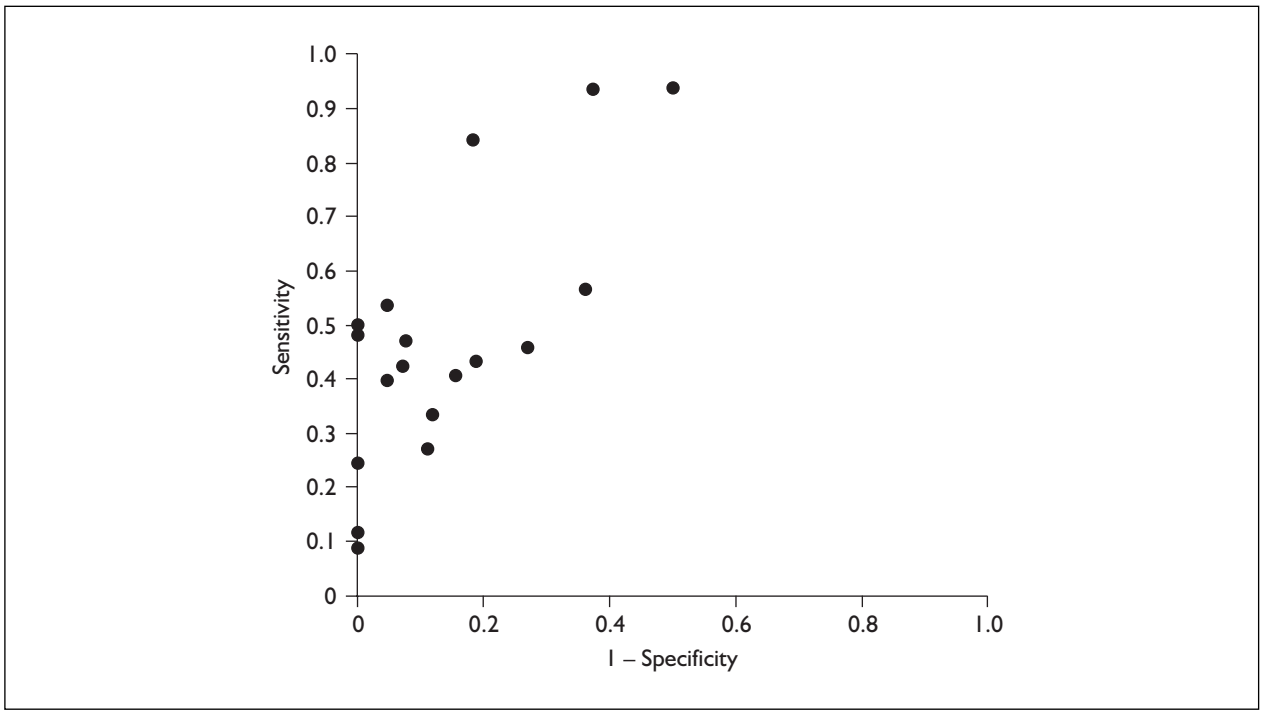

FIGURE 22 *Ultrasound for the localisation of UTI: study sensitivity and 1 – specificity plotted in ROC space*

than once, the combined results for the whole population (all age groups, first or multiple UTI) were used for further analyses. A second study[155] reported data by patient and renal unit; the patient data were used as these were considered more relevant to the issue of localisation of UTI and were used by the majority of the other studies in this group. A further study[160] reported data using both standard and Doppler ultrasound techniques. The data for standard ultrasound were used for further analyses as this was the method used by the majority of other studies in this group.

In the 18 studies using ultrasound techniques with renal scintigraphy as the reference standard, sensitivity was generally found to be poor, with relatively high specificity. Sensitivity ranged from 9.2% (specificity 100%) to 93.6% (specificity 50%). However, all but three of the studies reported sensitivities below 60%. Specificity ranged from 50% (sensitivity 93.6%) to 100% (sensitivity 9.2–50%); all but four estimates of specificity were above 80%. Likelihood ratios showed considerable heterogeneity ($p < 0.0001$). Positive likelihood ratios ranged from 1.6 (LR– 0.68) to 55.0 (LR– 0.50). Negative likelihood ratios ranged from 0.10 (LR+ 2.5) to 0.91 (LR+ 12.7). The pooled positive likelihood ratio was 3.11 (95% CI 2.3 to 4.3) and the pooled negative likelihood ratio was 0.62 (95% CI 0.53 to 0.73). These estimates should be interpreted with extreme caution owing to the significant heterogeneity present. *Figure 22* shows estimates of sensitivity and 1 – specificity plotted in ROC space. The median positive likelihood ratio was 5.3 (IQR 2.4–9.3) and the median negative likelihood ratio was 0.66 (IQR 0.51–0.76). This ROC plot shows considerable heterogeneity between studies but overall suggests that ultrasound is a poor test for the localisation of UTI.

A regression analysis was carried out to investigate possible explanations for the observed heterogeneity. The regression model $D = \alpha + \beta S$ was extended to include variables for quality items, age, region and study design. The results of the univariate regression analysis are shown in *Table 21*. None of the items investigated showed a significant association with D in this analysis.

MCUG

Seven studies evaluated the performance of MCUG for the diagnosis of APN.[143,146,156,157,169,183,208] All of these studies used ^{99m}Tc-DMSA renal scintigraphy as the reference standard; six used acute DMSA and one used the presence or absence of renal scarring as determined by follow-up DMSA. The studies described in this section were generally poorly reported, and were of average methodological quality. Only four of the seven studies included an appropriate spectrum of patients,[143,146,169,208] although all provided an adequate description of selection criteria. All but one of the studies reported that disease

TABLE 21 *Results of the regression analysis for ultrasound for the localisation of UTI*

Variable		β	RDOR	*p*-Value	Adjusted r^2
Spectrum		–0.5	0.6	0.435	0.22
Selection		–0.4	0.7	0.507	0.20
Reference standard		–0.6	0.5	0.322	0.23
Time		–0.6	0.5	0.281	0.24
Partial verification bias				Dropped	
Differential verification bias				Dropped	
Incorporation bias				Dropped	
Test details		–0.6	0.5	0.25	0.25
Reference details		–0.2	0.8	0.642	0.19
Test bias		–0.4	0.7	0.429	0.21
Review bias		–0.3	0.7	0.504	0.2
Clinical review bias		–0.2	0.8	0.886	0.18
Uninterpretable results		–0.7	0.5	0.19	0.27
Withdrawals		0.5	1.6	0.35	0.22
Age:	<2 years			Reference	0.09
	<5 years	–0.5	0.6	0.769	
	<12 years	–0.7	0.5	0.498	
	<18 years	–0.3	0.7	0.74	
Region:	North America			Reference	0.19
	Europe	–0.3	0.7	0.661	
Study design		–0.2	0.8	0.734	0.18

progression bias had been avoided.[146] Only two studies provided sufficient details on how MCUG was performed to permit replication,[157,169,183] although all but one provided details on how the reference standard was performed.[157] Two studies reported that the MCUG results were interpreted without knowledge of the reference standard results, and that the reference standard results were interpreted without knowledge of the MCUG results.[183,208] A further study reported that the reference standard results were interpreted without knowledge of the MCUG results.[143] The remaining studies failed to report on blinding. None of the studies provided details regarding what clinical information was available when the MCUG results were interpreted.

Sensitivity was generally found to be poor, with higher specificity. Sensitivity ranged from 21.6% (specificity 96.2%) to 47.1% (specificity 60%). Specificity ranged from 50% (sensitivity 29%) to 96.2% (sensitivity 21.6%); all but two estimates of specificity were above 80%. Positive likelihood ratios showed considerable heterogeneity ($p < 0.001$); however, negative likelihood ratios were statistically homogeneous ($p = 0.575$). Positive likelihood ratios ranged from 0.6 (LR– 1.42) to 5.8 (LR– 0.81). Negative likelihood ratios ranged from 0.72 (LR+ 2.8) to 1.42 (LR+ 0.6). The pooled positive likelihood ratio was 1.9 (95% CI 1.2 to 3.1). The pooled negative likelihood ratio was 0.80 (95% CI: 0.74 to 0.87). The pooled positive likelihood ratio should be interpreted with caution owing to the significant heterogeneity present. However, pooled likelihood ratios suggest that MCUG is a very poor test for the localisation of UTI. *Figure 23* shows estimates of sensitivity and 1 – specificity plotted in ROC space. All points indicate that MCUG is a poor test for the localisation of UTI.

The median positive likelihood ratio was 2.3 (IQR 1.6–2.8) and the median negative likelihood ratio was 0.81 (IQR 0.80 to 0.85).

Other imaging studies

One study assessed the accuracy of gadolinium-enhanced magnetic resonance imaging (MRI)[173] and reported high sensitivity (92%), but poor specificity (44%). This study was of generally good quality, but the index test and reference standard were not reported in sufficient detail to allow replication. A second study assessed the accuracy of CT[169] for the diagnosis of APN. Conversely, this study reported high specificity (100%) and poor sensitivity (56%). This study was also of good quality, but did not report sufficient information to determine whether or not investigators were blinded to other test results or clinical data when interpreting the index test and reference standard images. Both studies used ^{99m}Tc-DMSA renal scintigraphy as the reference standard. As only one study evaluating each of these imaging modalities was identified, it is not possible to reach any firm conclusions about their possible contribution to the localisation of UTI.

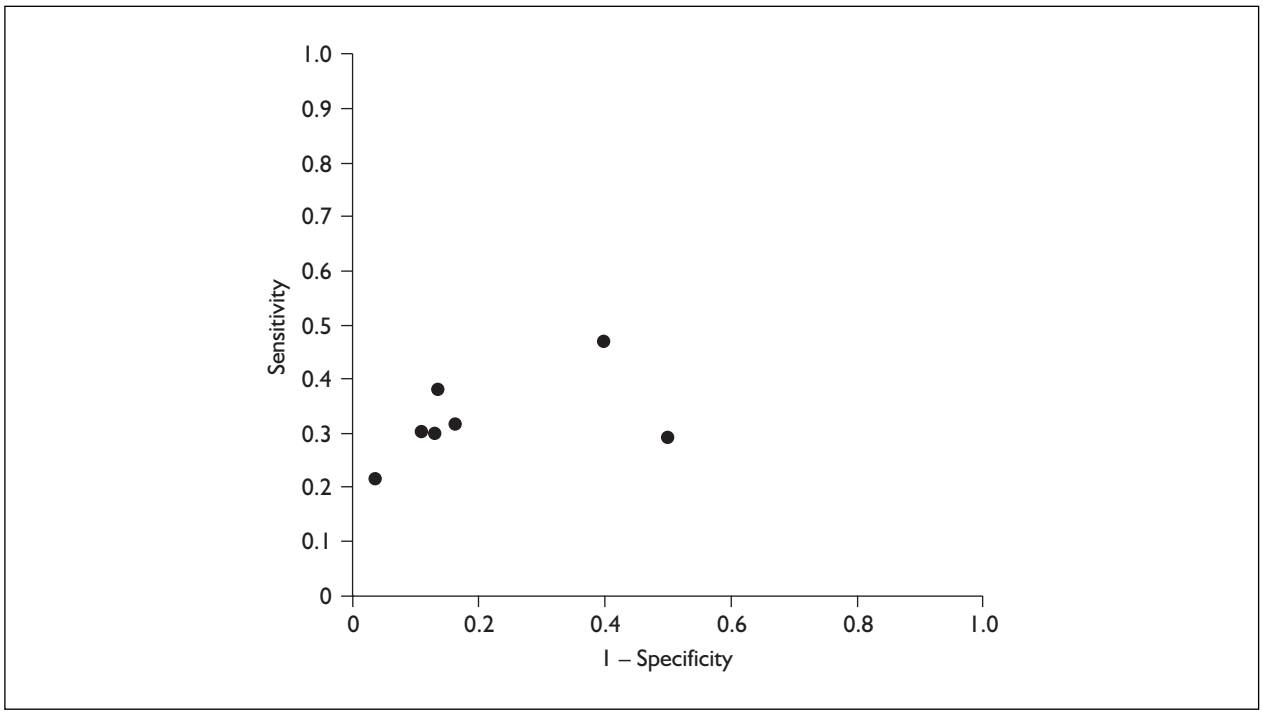

FIGURE 23 *MCUG for the localisation of UTI: study sensitivity and 1 – specificity plotted in ROC space*

Three studies evaluated the diagnostic performance of IVP using ^{99m}Tc-DMSA renal scintigraphy as the reference standard.[126,157,208] The details of the index test were poorly reported in these studies, and one study did not include an appropriate spectrum of patients. A fourth study compared IVP with the inappropriate reference standard of clinical and laboratory diagnosis. Therefore, no conclusions can be drawn from this study.[211] These studies reported relatively high specificity (75–100%), but poor sensitivity (9–44%) for the diagnosis of APN. However, given the small number of studies involved, no firm conclusions can be drawn regarding the utility of this test.

One study evaluated cystography using ^{99m}Tc-DMSA renal scintigraphy and the reference standard.[116] This study reported poor diagnostic performance (sensitivity 41% and specificity 68%). The quality of this study was poor: an appropriate spectrum of patients was not included and no selection criteria were reported, and details of the index test and reference standard were inadequate, and reporting was generally poor, with insufficient detail provided to assess the avoidance of review bias or disease progression bias. A second study used clinical and laboratory diagnosis as the reference standard.[211] These studies do not provide sufficient information to draw any conclusions about the usefulness of cystography for localising UTI.

Three studies assessed the accuracy of various scintigraphic techniques for the diagnosis of APN.[200,211,215] The techniques investigated were ^{99m}Tc-DMSA,[215] ^{99m}-Tc-glucoheptonate, and ^{131}I-orthoiodohippurate,[211] and ^{123}I-hippuran.[200] All three studies used inappropriate reference standards, including clinical diagnosis,[215] clinical and laboratory diagnosis,[211] and IVP.[200] All reported good diagnostic accuracy, with sensitivity ranging from 80 to 98% and specificity from 98 to 100%. However, as the reference standards used have been shown to be poor tests for the localisation of UTI, these results should be interpreted with some degree of caution.

Summary

A wide variety of tests has been investigated for the localisation of UTI. There was insufficient information to draw overall conclusions regarding the accuracy of clinical examination or laboratory-based tests for the localisation of UTI. The clinical and laboratory-tests investigated showed fairly poor accuracy, but only a limited number of tests was investigated.

Ultrasound was investigated in a large number of studies and was found to be a poor test for the localisation of UTI. The pooled negative likelihood ratio was 0.62 (95% CI 0.53 to 0.73), suggesting that this is a poor test for ruling out APN; that is, a negative ultrasound cannot be used

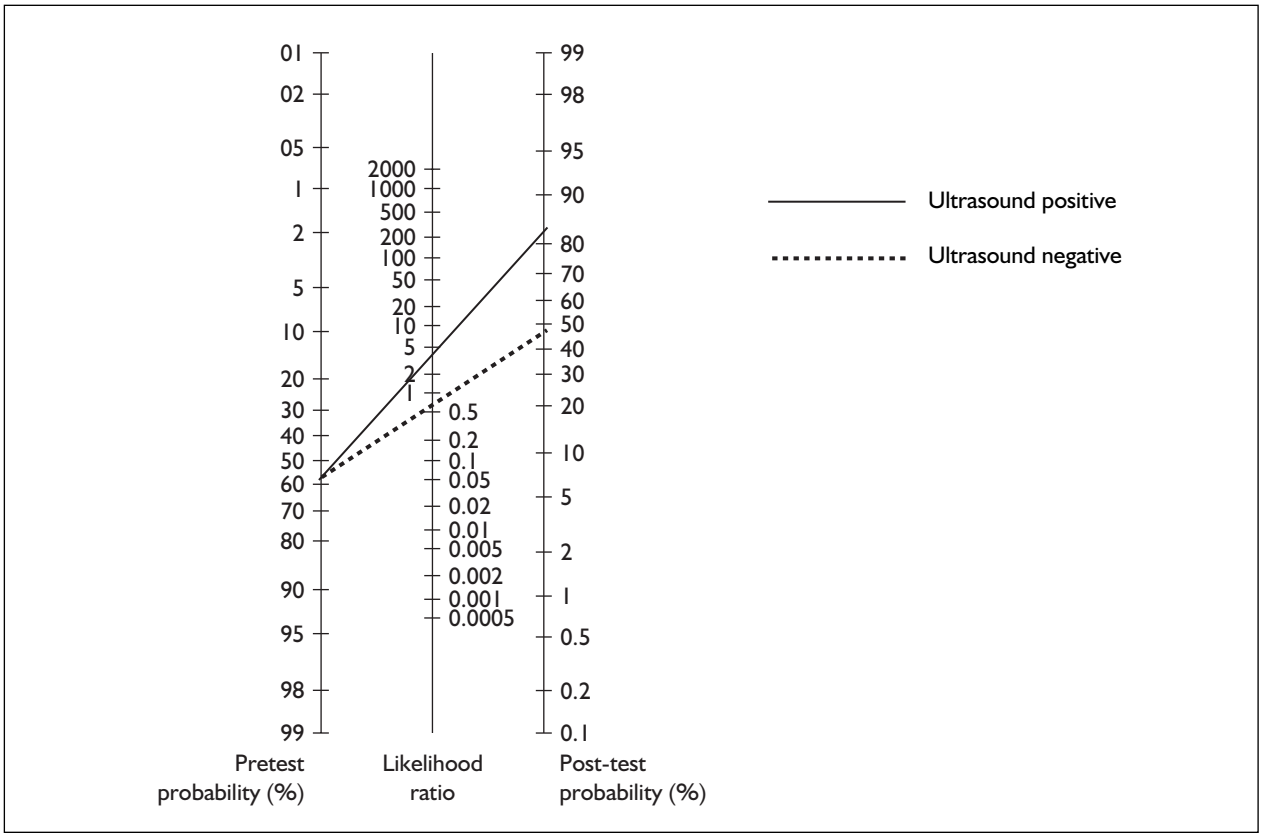

FIGURE 24 *Likelihood ratio monogram for ultrasound for the localisation of UTI*

as evidence to identify if UTI is lower rather than upper. The pooled positive likelihood ratio was low at 3.11 (95% CI 2.3 to 3.4) suggesting that ultrasound is also poor at ruling in APN; that is, a positive ultrasound scan cannot be used as evidence of APN. MCUG, which was investigated in seven studies, was found to be an even poorer test for the localisation of UTI. The pooled positive likelihood ratio was 1.9 (95% CI 1.2 to 3.1) and the pooled negative likelihood ratio was 0.80 (95% CI 0.74 to 0.87). All pooled estimates should be interpreted with extreme caution owing to significant heterogeneity between studies.

Other imaging studies that investigated localisation of UTI included MRI, CT, IVP, cystography and various different scintigraphic techniques. The only techniques found to have good accuracy for the localisation of UTI were the scintigraphic techniques. However, all the studies that investigated these tests used inappropriate reference standards.

What do these results mean?

The results of studies of the localisation of UTI are summarised in *Table 22*. The results indicate that scintigraphy is the only test that can accurately localise UTI; this test is generally used as a reference standard for UTI localisation. The limited value of ultrasound in the localisation of UTI is illustrated in the next section. The likelihood ratios and an estimate of the pretest probability of upper UTI were used to calculate the post-test probability of upper UTI. The reviewers were unable to find reliable estimates of the proportion of children with UTI who have an upper UTI (the pretest probability of disease) in the literature. Therefore, they used the data from the studies included in this section of the review to provide an estimate of the pretest probability of upper UTI. As before, only studies that included an appropriate patient spectrum were included in this analysis. The median prevalence of upper UTI in these studies was 60%, and was used as an estimate of the pretest probability of upper UTI. *Figure 24* shows how a positive or negative ultrasound scan for APN changes the probability of upper UTI.

Detection of reflux

A total of 34 studies reporting 57 data sets investigated tests for the detection of reflux. All but one of these studies evaluated imaging

TABLE 22 *Results of studies of the localisation of UTI*

Study	Test details; time	Definition of positive result	Reference standard; definition of positive result; time	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Clinical features of APN vs scintigraphy													
Biggi, 2001[125]	Temperature; acute	≥ 39.1°C	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	45	18	25	12	64.3	40.0	1.2	1.1	0.89
Buyan, 1993[129]	Flank pain, chills, nausea, vomiting, fever, tenderness of the costovertebral angle; acute	Presence of any symptoms	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	12	0	9	3	57.1	100.0	9.2	4.5	0.49
Everaert, 1998[140]	NS; acute	Symptoms of APN	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	31	0	13	18	70.5	100.0	86.3	26.6	0.31
Fretzayas, 2000[143]	Temperature; acute	≥ 38°C	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	26	19	4	34	86.7	64.2	10.4	2.4	0.23
Landau, 1994[168]	Physical examination; blood WBC; band forms; urinalysis; WBC in stool (when diarrhoea present); acute	Presence of any symptoms	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	48	53	1	26	98.0	32.9	16.0	1.5	0.09
CRP concentration vs scintigraphy													
Biggi, 2001[125]	CRP; acute	>880 mg l^{-1}	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	45	10	25	21	64.3	67.7	3.7	2.0	0.53
Buyan, 1993[129]	CRP; acute	>20 μg/l	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	3	0	18	3	14.3	100.0	1.3	1.3	0.96

continued

TABLE 22 *Results of studies of the localisation of UTI (cont'd)*

Study	Test details; time	Definition of positive result	Reference standard; definition of positive result; time	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Fretzayas, 2000[143]	CRP; acute	>200 mg l^{-1}	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	21	23	9	30	70.0	56.6	2.9	1.6	0.54
Gervaix, 2001[145]	CRP; acute	≥ 400 mg l^{-1}	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	23	9	11	11	67.6	55.0	2.5	1.5	0.60
Girona, 1995[146]	CRP; acute	>20 mg l^{-1}	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; NS	Patients	18	14	3	21	85.7	60.0	7.8	2.1	0.27
Smolkin, 2002[204]	CRP; acute	≥ 20 mg l^{-1}	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	18	34	0	8	100.0	19.0	9.1	1.2	0.13
Stokland, 1996[208]	CRP; acute	>20 mg l^{-1}	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	69	73	4	29	94.5	28.4	6.2	1.3	0.21
ESR vs scintigraphy													
Biggi, 2001[125]	ESR; acute	>68 mm h^{-1}	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	34	15	36	16	48.6	51.6	1.0	1.0	1.00
Buyan, 1993[129]	ESR; acute	>25 mm h^{-1}	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	7	0	14	3	33.3	100.0	3.6	2.7	0.75
Fretzayas, 2000[143]	ESR; acute	30 mm h^{-1}	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	27	22	3	31	90.0	58.5	11.0	2.1	0.19

continued

TABLE 22 *Results of studies of the localisation of UTI (cont'd)*

Study	Test details; time	Definition of positive result	Reference standard; definition of positive result; time	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Microscopy vs scintigraphy													
Biggi, 2001[125]	Microscopy; acute	>14,601 WBC mm^{-3}	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	39	13	31	18	55.7	58.1	1.7	1.3	0.77
		>52% granulocytes			36	11	34	20	51.4	64.5	1.9	1.4	0.76
Buyan, 1993[129]	Microscopy; acute	>15,000 WBC cm^{-2}	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	5	0	16	3	23.8	100.0	2.3	2.0	0.86
Landau, 1994[167]	Microscopy; acute	≥ 5 WBC hpf^{-1}	Scintigraphy (^{99m}Tc-DMSA); NS; acute	Patients	48	56	4	34	92.3	37.8	6.6	1.5	0.22
Landau, 1994[168]	Microscopy; acute	≥ 5 WBC hpf^{-1}	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	45	52	4	27	91.8	34.2	5.3	1.4	0.26
Immunofluorescence detection of ACB													
Barnett, 1978[118]	Immunofluorescence; acute	Any fluorescence	IVU and MCUG ; NS; NS	Patients	11	6	10	39	52.4	86.7	6.7	3.7	0.56
Buyan, 1993[129]	Immunofluorescence; acute	>2 ACB per 200 fields	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	14	1	7	2	66.7	66.7	3.2	1.8	0.55
Hellerstein, 1978[153]	Immunofluorescence; acute	>1% ACB	Bladder washout test; ≥ 1000 cfu ml^{-1} in 1/3 final urine specimens; NS	Patients	8	14	4	21	66.7	60.0	2.8	1.6	0.58
Montplaisir, 1981[182]	Immunofluorescence; acute	>3 ACB per 200 fields	Not clear (clinical and radiological criteria); NS; NS	Patients	75	97	19	91	79.8	48.4	3.6	1.5	0.42

continued

TABLE 22 *Results of studies of the localisation of UTI (cont'd)*

Study details	Test details; time	Definition of positive result	Reference standard; definition of positive result; time	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Pylkkanen, 1978[190]	Immunofluorescence; acute	≥ 2 ACB	Clinical (Jodal criteria); evidence of APN; acute	Patients	29	4	49	32	37.2	88.9	4.3	3.1	0.71
Other biochemical tests for the diagnosis of APN vs scintigraphy													
Capa Kaya, 2001[130]	NAG and NAG/creatinine ratio; acute	5 U l^{-1} for NAG, 7 U g^{-1} for NAG:CR	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	30	5	0	65	100.0	92.9	726.5	12.7	0.02
Everaert, 1998[140]	Urinary α_1-MG/ creatinine ratio; acute	>10 mg g^{-1}	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	43	0	1	18	97.7	100.0	1073.0	36.7	0.03
Fretzayas, 2000[143]	Polymorphonuclear elastase-α_1-antitrypsin complex; acute	95th percentile of reference range	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	29	27	1	26	96.7	49.1	19.0	1.9	0.10
Gervaix, 2001[145]	PCT; acute	≥ 0.5 ng ml^{-1}	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	25	3	9	17	73.5	85.0	13.4	4.4	0.33
Jantausch, 1994[158]	β_2M; acute	≥ 0.5 μg mg^{-1} CR	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; NS	Patients	10	2	4	1	71.4	33.3	1.4	1.1	0.80
	NAG/creatinine ratio and β_2M; acute	NS		Patients	6	2	8	1	42.9	33.3	0.5	0.7	1.51
Jantausch, 1994[158]	NAG/creatinine ratio; acute	≥ 40 μmol h^{-1} mg^{-1} CR	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; NS	Patients	13	4	6	1	68.4	20.0	0.7	0.9	1.30
Smolkin, 2002[204]	PCT; acute	≥ 0.5 μg l^{-1}	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	17	4	1	38	94.4	90.5	99.8	8.8	0.09

continued

TABLE 22 *Results of studies of the localisation of UTI (cont'd)*

Study	Test details; time	Definition of positive result	Reference standard; definition of positive result; time	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Ultrasound vs scintigraphy													
Andrich, 1992[116]	Standard; NS	NS	Scintigraphy (^{99m}Tc-DMSA); NS; NS	Patients	3	0	23	24	11.5	100.0	7.3	6.5	0.88
Benador, 1994[121]	Standard; acute	Renal changes indicative of APN	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	12	3	19	6	38.7	66.7	1.2	1.1	0.94
≥ 1 year, any UTI													
≥ 1 year, 1st UTI					5	2	13	1	27.8	33.3	0.2	0.5	1.89
≥ 1 year, multiple UTI					7	1	6	5	53.8	83.3	4.2	2.5	0.59
<1 year, any UTI					20	4	23	24	46.5	85.7	4.7	3.0	0.63
<1 year, 1st UTI					18	4	22	22	45.0	84.6	4.1	2.7	0.66
< 1year, multiple UTI					2	0	1	2	66.7	100.0	8.3	3.8	0.45
all ages, any UTI					32	7	42	30	43.2	81.1	3.1	2.3	0.70
all ages, 1st UTI					23	6	35	23	39.7	79.3	2.4	1.8	0.77
all ages, multiple UTI					9	1	7	7	56.3	87.5	6.3	3.4	0.53
Biggi, 2001[125]	Standard; NS	Renal changes indicative of APN	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Renal units	19	12	51	96	27.1	88.9	2.9	2.4	0.82
Bircan, 1995[126]	Standard; acute	Renal changes indicative of APN and presence of congenital abnormalities	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	11	0	34	18	24.4	100.0	12.3	9.5	0.76
Boudailliez, 1998[127]	Doppler; NS	NS	Scintigraphy (^{99m}Tc-DMSA); NS; acute	Renal units	10	8	20	60	33.3	88.2	3.6	2.8	0.76

continued

TABLE 22 *Results of studies of the localisation of UTI (cont'd)*

Study	Test details; time	Definition of positive result	Reference standard; definition of positive result; time	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Girona, 1995[146]	Standard; NS	Abnormal kidney size	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; NS	Renal units	17	36	20	97	45.9	72.9	2.3	1.7	0.74
Guermazi, 1993[148]	Standard; NS	Renal changes indicative of APN or scarring	Scintigraphy (^{99m}Tc-DMSA); presence of acute or chronic lesions; NS	Patients	14	5	19	64	42.4	92.8	8.7	5.9	0.62
Hajjar, 2002[150]	Doppler; acute	Renal changes indicative of APN	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	15	1	13	20	53.6	95.2	15.7	11.3	0.49
Hitzel, 2002[155]	Doppler; acute	Renal changes indicative of APN	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	44	5	3	5	93.6	50.0	12.7	1.9	0.13
				Renal units	43	11	11	48	79.6	81.4	16.0	4.1	0.26
Hitzel, 2000[177]	Colour Doppler; NS	NS	Scintigraphy (^{99m}Tc-DMSA); NS; acute	Renal units	43	11	8	49	84.3	81.7	22.0	4.6	0.19
Ilyas, 2002[156]	Standard; acute	Renal changes indicative of APN	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	9	0 (8)	89	65 (57)	9.2	100.0	13.9	12.7	0.91
Jakobsson, 1992[157]	Standard; acute	Renal changes indicative of APN	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Renal units	61	13	47	23	56.5	63.9	2.3	1.6	0.69
Jequier, 1998[160]	Standard; acute	Renal changes indicative of APN	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	71	18	104	97	40.6	84.3	3.6	2.6	0.70
	Doppler; acute				22	1	89	61	19.8	98.4	10.3	8.4	0.82

continued

TABLE 22 *Results of studies of the localisation of UTI (cont'd)*

Study	Test details; time	Definition of positive result	Reference standard; definition of positive result; time	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Krzemien, 2002[166]	Doppler; acute	Renal changes indicative of APN	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Renal units	15	2	17	24	46.9	92.3	8.7	6.1	0.58
Lavocat, 1997[169]	Standard; acute	Renal changes indicative of APN	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Renal units	28	0	28	54	50.0	100.0	109.0	55.0	0.50
Morin, 1999[183]	Standard; acute	Renal changes indicative of APN	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	58	3	4	5	93.5	62.5	20.4	2.5	0.10
Sfakianakis, 1989[201]	Standard; NS	NS	Scintigraphy (^{99m}Tc-glucoheptonate); NS; NS	Patients	12	0	13	23	48.0	100.0	43.5	23.1	0.52
Sreenarasimhaiah, 1995[205]	Standard; acute	NS	Scintigraphy (^{99m}Tc-glucoheptonate); renal changes indicative of APN; acute	Renal units	21	2 (7)	32	41 (36)	39.6	95.3	11.0	8.5	0.63
Ultrasound vs other													
Dacher, 1996[134]	Doppler; acute	Renal changes indicative of APN	CT; renal changes indicative of APN; acute	Patients	17	1	2	10	89.5	90.9	49.0	7.0	0.14
Traisman, 1986[211]	Standard; acute	NS	Clinical and laboratory diagnosis of APN; acute	Patients	8	0	11	5	42.1	100.0	8.1	5.1	0.63
MRI/CT vs scintigraphy													
Lavocat, 1997[169]	CT scan (sodium meglumine ioxitalamate); acute	Renal changes indicative of APN	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	14	0	11	11	56.0	100.0	29.0	13.4	0.46

continued

TABLE 22 *Results of studies of the localisation of UTI (cont'd)*

Study	Test details; time	Definition of positive result	Reference standard; definition of positive result; time	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Lonergan, 1998[173]	MRI (gadolinium enhanced); NS	Renal changes indicative of APN	Scintigraphy (^{99m}Tc-DMSA or ^{99m}Tc-glucoheptonate); renal changes indicative of APN; NS	Patients	23	5	2	4	92.0	44.4	7.7	1.6	0.21
				Renal units	26	11	4	25	86.7	69.4	13.1	2.8	0.21
Cystography for the diagnosis of APN													
Andrich, 1992[116]	Cystography; NS	NS	Scintigraphy (^{99m}Tc-DMSA); NS; NS	Patients	13	13	19	27	40.6	67.5	1.4	1.2	0.88
Traisman, 1986[211]	Cystography; NS	NS	Clinical and laboratory diagnosis of AP; acute	Patients	9	0	10	4	47.4	100.0	8.1	4.8	0.58
IVP for the diagnosis of APN													
Bircan, 1995[126]	IVP (sodium megluminediatrizoate); acute	Presence of anatomical pathologies	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	4	0	41	18	8.9	100.0	4.0	3.7	0.93
Jakobsson, 1992[157]	IVP; follow-up	Presence of renal scarring	Scintigraphy (^{99m}Tc-DMSA); presence of renal scarring; follow-up	Kidneys	24	17	33	50	42.1	74.6	2.1	1.6	0.78
Stokland, 1996[208]	IVP; NS	Presence of renal scarring	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Renal units	9	4	76	261	10.6	98.5	7.2	6.5	0.90
Traisman, 1986[211]	IVP (renographin 60); NS	NS	Clinical and laboratory findings diagnosis of A; acute	Patients	4	0	13	4	23.5	100.0	3.0	2.5	0.83
MCUG vs scintigraphy													
Fretzayas, 2000[143]	MCUG; acute	Presence of reflux	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	9	7	21	46	30.0	86.8	2.7	2.3	0.81

continued

TABLE 22 *Results of studies of the localisation of UTI (cont'd)*

Study	Test details; time	Definition of positive result	Reference standard; definition of positive result; time	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Girona, 1995[146]	MCUG; NS	Presence of reflux grade 2 or more	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Renal units	8	5	29	128	21.6	96.2	6.7	5.8	0.81
Ilyas, 2002[156]	MCUG; acute	Presence of reflux	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	65	28	73	42	47.1	60.0	1.3	1.2	0.88
Jakobsson, 1992[157]	MCUG; follow-up	Presence of reflux	Scintigraphy (^{99m}Tc-DMSA); presence of renal scarring; follow-up	Kidneys	18	11	39	56	31.6	83.6	2.3	1.9	0.82
Lavocat, 1997[169]	MCUG; NS	Presence of reflux	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Renal units	17	6	39	48	30.4	88.9	3.3	2.7	0.78
Morin, 1999[183]	MCUG; acute	Presence of reflux	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Patients	18	4	44	4	29.0	50.0	0.4	0.58	1.42
Stokland, 1996[208]	MCUG; NS	Presence of reflux	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Renal units	32	36	52	228	38.1	86.4	3.9	2.8	0.72
Scintigraphy for the diagnosis of APN													
La Cava, 1990[200]	^{123}I-hippuran; not clear	Presence of renal scarring (abnormal ERPF 270 ml minute^{-1})	IVP; presence of renal scarring; not clear	Renal units	83	5	2	231	97.6	97.9	1405.8	41.8	0.03
Traisman, 1986[211]	^{99m}Tc-glucoheptonate and ^{131}I-orthoiodohippurate; acute	Renal changes indicative of APN	Clinical and laboratory; diagnosis of APN; acute	Patients	24	0 (1)	4	4 (3)	85.7	100.0	49.0	8.4	0.17

continued

TABLE 22 *Results of studies of the localisation of UTI (cont'd)*

Study	Test details; time	Definition of positive result	Reference standard; definition of positive result; time	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Verboven, 1990[215]	^{99m}Tc-DMSA; acute	NS	Clinical (Pylkkanen criteria); diagnosis of AP; acute	Patients	12	0 (1)	3	9 (8)	80.0	100.0	67.9	15.6	0.23

ACB, antibody-coated bacteria; CR, creatinine; ERPF, effective renal plasma flow.

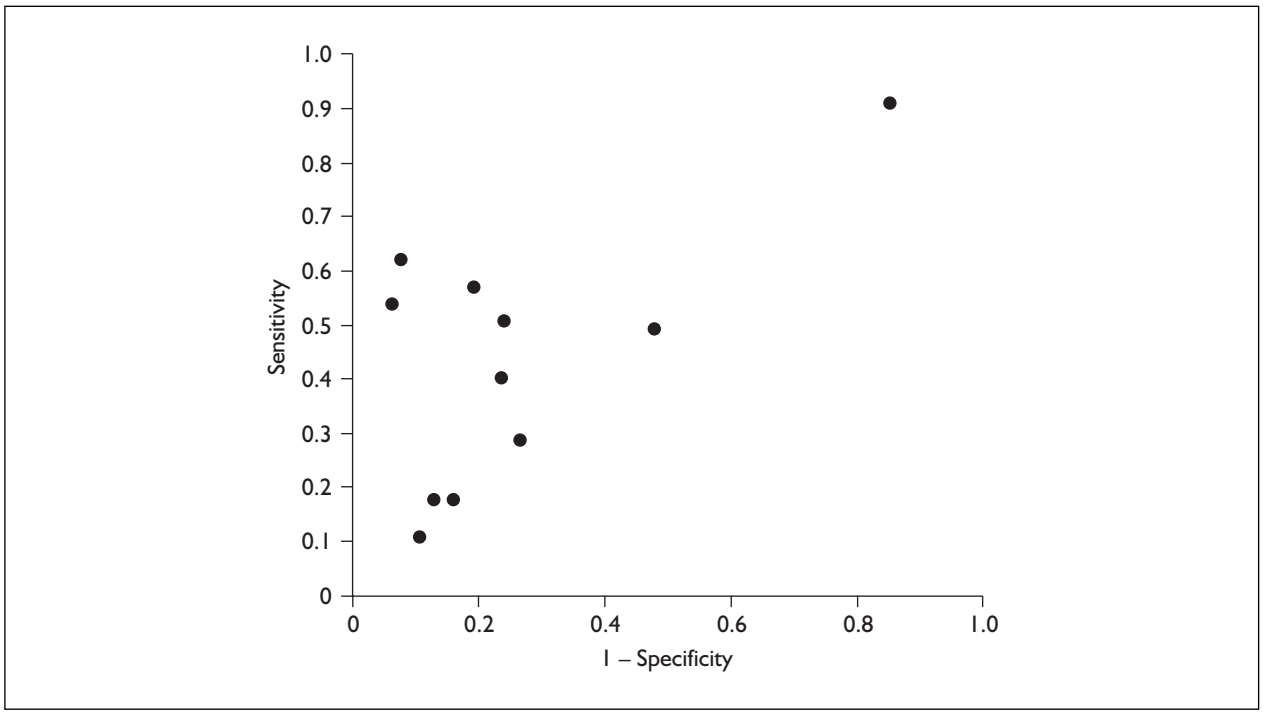

FIGURE 25 *Standard ultrasound for the detection of reflux: study sensitivity and 1 – specificity plotted in ROC space*

techniques. The majority of studies (30) used MCUG as the reference standard.

Ultrasound

The main imaging technique evaluated for the detection of reflux was ultrasound. The diagnostic accuracy of ultrasound was assessed in 26 of the 34 studies.

Conventional ultrasound

Eleven studies evaluated standard[119,139,175,176,183,185,186,210,212,214] or duplex[142] ultrasound compared with the reference standard of MCUG. One study[185] reported data for a variety of cut-off points. To prevent duplicate inclusion of study participants, only data examining the presence of reflux (dilatation) by ultrasound compared with presence of any reflux by MCUG were used in the analyses. Around half of the studies in this section did not include an appropriate spectrum of patients. Poor quality of reporting was also a particular problem: most studies (9/11) did not report sufficient details of the reference standard to permit replication, half of all studies did not report sufficient information to allow assessment of the avoidance of review bias, and half of all studies did not report sufficient information to judge whether the time elapsing between the index test and reference standard was appropriate (disease progression bias). Partial verification bias was a problem in two studies.

Estimates of sensitivity and specificity reported by these studies varied greatly. Sensitivity ranged from 10.5% (specificity 89.4%) to 90.9% (14.6%). Specificity ranged from 14.6% (sensitivity 90.9%) to 93.8% (sensitivity 53.7%). Likelihood ratios showed considerable heterogeneity ($p < 0.0001$). Positive likelihood ratios ranged from 1.0 (LR– ~1.0) to 8.7 (LR– 0.49). Negative likelihood ratios ranged from 0.41 (LR+ 8.2) to 0.98 (LR+ ~1.0). The pooled positive likelihood ratio was 1.9 (95% CI 1.2 to 2.9). The pooled negative likelihood ratio was 0.76 (95% CI 0.63 to 0.93). The pooled likelihood ratios suggest that standard ultrasound is a poor test for detecting reflux; however, owing to the presence of significant heterogeneity, these should be interpreted with caution. *Figure 25* shows estimates of sensitivity and 1 – specificity plotted in ROC space. The median positive likelihood ratio was 1.4 (IQR 1.1–2.5) and the median negative likelihood ratio was 0.79 (IQR 0.58–0.98).

Contrast-enhanced ultrasound

Fourteen studies assessed the diagnostic accuracy of cystosonography,[123,124,144,149,165,179,187,191,195,199,202,213] or other contrast-enhanced techniques,[115,197] using a variety of contrast agents. All but one of these studies used MCUG as the reference standard. None of these studies included an appropriate spectrum of patients. The quality of reporting was also an issue: nine studies did not

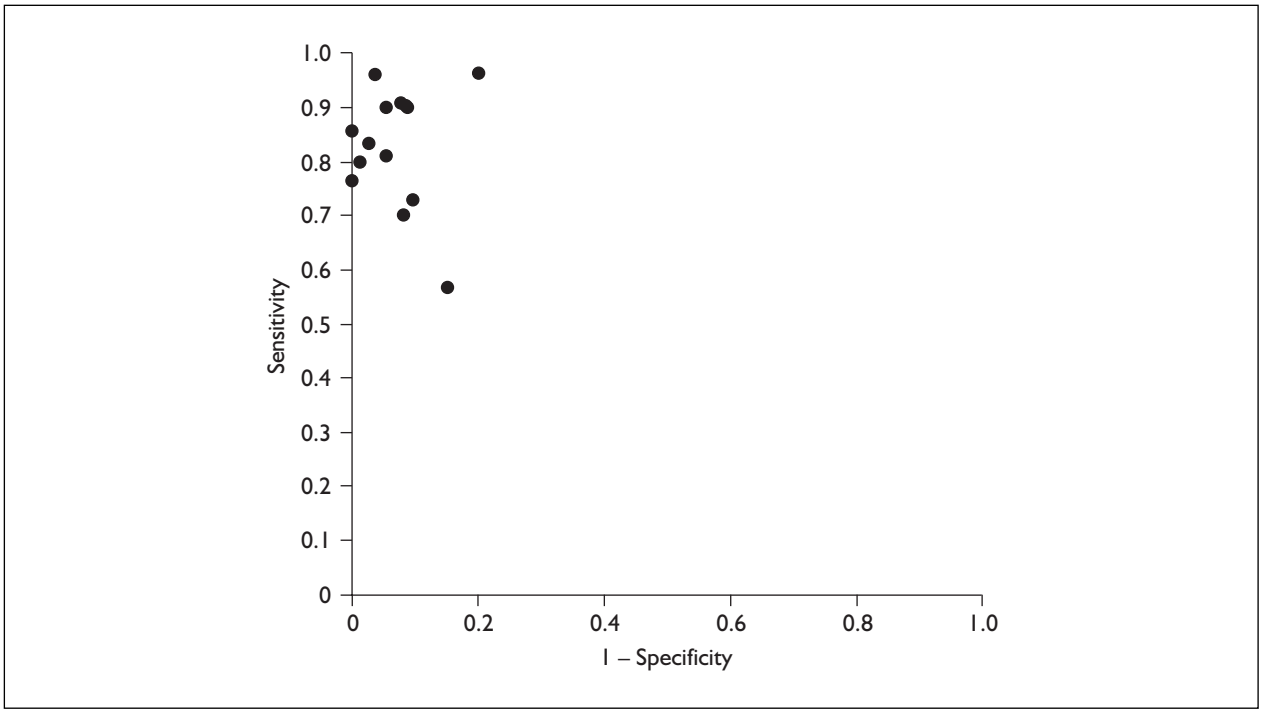

FIGURE 26 *Contrast-enhanced ultrasound for the detection of reflux: study sensitivity and 1 – specificity plotted in ROC space*

describe the reference standard in sufficient detail to permit replication;[115,123,149,165,187,191,195,199,202] six studies reported insufficient detail to allow assessment of the avoidance of test review bias,[115,123,187,195,199,202] and three studies reported insufficient detail to allow assessment of the avoidance of disease progression bias.[165,187,195] The remaining study used direct radionuclide voiding cystography as the reference standard.[164] This is not considered to be an appropriate reference standard and so the results are not included in the analysis below. Where studies reported more than one data set for the same population, only one data set was included in further analyses; this was to prevent duplicate inclusion of the same study participants. One study[123] reported data using cut-off thresholds of both reflux grade 2 or above, and reflux grade 3 or above; a second study[199] reported data using cut-off points of both reflux grade 2 or above, and presence of any reflux; data for the lower cut-off were used in both cases. One study[124] reported data by patient and by renal unit. In this case the data calculated by renal unit was used in the present analyses as they were deemed more pertinent to the question of detection of reflux and were reported in the majority of studies in this category. Two further studies[202,213] reported data for more than one cystosonographic technique (air and fluid cystosonography, and grey-scale and Doppler cystosonography, respectively). Data for air-contrast cystosonography and grey-scale cystosonography were used in the analysis.

In general, estimates of sensitivity and specificity reported were relatively high. Sensitivity ranged from 56.8% (specificity 84.8%) to 96.3% (80%); in all but three studies sensitivity was above 75%. Specificity ranged from 80% (sensitivity 96.3%) to 100% (sensitivity 76.5 and 85.7%). Likelihood ratios showed considerable heterogeneity ($p < 0.0001$). Positive likelihood ratios ranged from 3.8 (LR– 0.51) to 71.2 (LR– 0.20). Negative likelihood ratios ranged from 0.04 (LR+ 25.6) to 0.51 (LR+ 3.8). The pooled positive likelihood ratio was 12.3 (95% CI 8.2 to 18.3) and the pooled negative likelihood ratio was 0.17 (95% CI 0.11 to 0.27). The median positive likelihood ratio was 13.7 (IQR 9.1–30.8) and the median negative likelihood ratio was 0.16 (IQR 0.11–0.23). *Figure 26* shows estimates of sensitivity and 1 – specificity plotted in ROC space. The points on this graph are all scattered towards the upper left-hand corner of the chart, suggesting that contrast-enhanced ultrasound may be a good test for the detection of reflux. The graph suggests that one study may be an outlier.[187] This study does not appear to be significantly different from the other studies in terms of quality, spectrum composition, reference standard or ultrasound methods.

TABLE 23 *Results of the regression analysis for ultrasound for the detection of reflux*

Variable	β	RDOR	*p*-Value	Adjusted r^2
Ultrasound	3.2	24.5	<0.01	0.67

TABLE 24 *Results of the regression analysis for ultrasound for the detection of reflux, with ultrasound type forced into the model*

Variable		β	RDOR	*p*-Value	Adjusted r^2
Spectrum		0.9	2.5	0.3	0.68
Selection		–0.7	0.5	0.18	0.68
Reference standard		–2.0	0.1	0.05	0.71
Time		1.1	3.0	0.05	0.71
Partial verification bias		1.2	3.3	0.24	0.67
Differential verification bias		0.9	2.5	0.65	0.65
Incorporation bias				Dropped	
Test details		–0.8	0.4	0.29	0.67
Reference details		0.1	1.1	0.84	0.65
Test bias		0.6	1.8	0.17	0.68
Review bias		0.8	2.2	0.13	0.68
Clinical review bias		0.4	1.5	0.67	0.65
Uninterpretable results		0.2	1.2	0.73	0.65
Withdrawals		0.7	2.0	0.19	0.67
Age:	<2 years			Reference	0.66
	< 5 years	–0.3	0.7	0.84	
	<12 years	–1.1	0.3	0.5	
	<18 years	0	1.0	0.98	
Region:	North America			Reference	0.64
	Europe	0.6	1.8	0.46	
	Asia	–0.7	0.5	0.74	
Study design		0	1.0	0.98	0.65
Population (UTI only/mixed)		0.8	2.2	0.2	0.67

TABLE 25 *Results of the multivariate regression analysis for ultrasound for the detection of reflux*

Variable	β	RDOR	*p*-Value	Adjusted r^2
Ultrasound	3	20.1	<0.01	0.71
Time	1.1	3.0	0.05	

A regression analysis was carried out to investigate possible explanations for the observed heterogeneity. The regression model $D = \alpha + \beta S$ was extended to include a variable for type of ultrasound (contrast enhanced or standard). The results of this model are shown in *Table 23*. The type of ultrasound showed a highly significant association with D. The DOR was over 24 times greater in studies that used contrast-enhanced ultrasound than in those that used standard ultrasound. Given the significance of this association, and the clinical importance of distinguishing between ultrasound types, this variable was included in the model to investigate other possible sources of heterogeneity. The regression model $D = \alpha + \beta_1 S + \beta_1$(ultrasound) was extended to include variables for quality items, age, region, study design (prospective versus retrospective) and population (children with confirmed UTI only versus mixed population). The results of this regression analysis are shown in *Table 24*. Two items, the use of an appropriate reference standard and time delay, were significant in this analysis. These items were included in a multivariate analysis, but only time delay (the possibility of disease progression bias) and ultrasound type remained significant. The DOR was around three times higher in studies that avoided disease progression bias than in studies in which disease progression bias may have been a problem. The DOR remained over 20 times higher in studies that used contrast-enhanced ultrasound than in those that used standard ultrasound. The final model, shown in *Table 25*, provided a reasonable fit for the data, with an adjusted r^2 of 0.71.

IVP

Four studies assessed the accuracy of IVP for detecting reflux of any grade[131,137,192,214] and all used MCUG as the reference standard. These studies were poorly reported: only one study provided adequate detail of either the index test or the reference standard;[214] two studies did not report the criteria used to select participants,[192,214] three studies did not report sufficient detail to assess the avoidance of diagnostic review bias;[137,192,214] and two studies did not report sufficient detail to assess the avoidance of disease progression bias.[192,214] All studies showed that a positive IVP result was fairly specific for reflux (73–100%), but sensitivity was low (28–48%). One of these studies[131] also evaluated the accuracy of IVP for detecting reflux of grade 3 or above. Sensitivity was increased to 100% for this comparison, with little impact on specificity.

Other tests and combinations of tests

Eight studies investigated a variety of imaging and other techniques, including indirect voiding radionuclide cystography,[128,135,152] NAG/creatinine ratio,[162] scintigraphy,[181,212,214] and a risk scoring system.[186] A study of NAG/creatinine ratio found that this was a reasonable test for the detection of reflux in children with cystitis according to clinical criteria, but a poor test in children with APN.[162] The study of the risk scoring system used a combination of gender, family history, age, CRP level and ultrasound to produce an overall risk score.[186] This study reported results for several different cut-off points and for the diagnosis of any grade reflux and reflux grade 3 or above. The DOR ranged between 4 and 14, suggesting that this system was a poor test for the detection of reflux at any of the cut-off points reported and for both grades of reflux.

Two studies evaluated the diagnostic accuracy of standard ^{99m}Tc-DMSA scintigraphy for the diagnosis of reflux.[212,214] Neither of these studies was of good quality. Only one of them reported sufficient detail of the index test to permit replication,[214] and neither reported the criteria used to select participants. Neither study reported sufficient detail to assess the avoidance of review bias, and one did not report sufficient information for the assessment of disease progression bias;[214] partial verification bias was also a problem in this study. Both reported relatively poor test performance, with sensitivities of 67% and 77%, and specificities of 63% and 74%. An additional study looked at ^{131}I-*o*-hippurat and found this to be a reasonable test for the detection of reflux.[181] However, this study did not use an appropriate spectrum of patients or reference standard, and was very poorly reported (criteria used to select patients were not reported, details of the reference standard were inadequate, and neither avoidance of review bias nor avoidance of disease progression bias could be assessed). Three studies evaluated indirect radionuclide voiding cystography.[128,135,152] Two used ^{99m}Tc-diethylenetriamine pentaacetic acid (DTPA) and one used ^{99m}Tc-MAG3. These studies were generally of good quality. However, only one of them included an appropriate spectrum of patients,[152] and none reported the criteria used to select participants. These reported high specificity (95–100%), but low sensitivity (33–68%).

Summary

The main test investigated for the detection of reflux was ultrasound. The studies of ultrasound were divided into two main categories: standard and contrast enhanced. Standard ultrasound was found to have poor performance for the detection of reflux. The pooled positive likelihood ratio was 1.9 (95% CI 1.2 to 2.9) and the pooled negative likelihood ratio was 0.76 (95% CI 0.63 to 0.93). These figures should be interpreted with caution owing to the significant heterogeneity between studies; however, they suggest that ultrasound is a very poor test both for ruling in and for ruling out disease. Contrast-enhanced ultrasound was found to have much better performance for the detection of reflux. The pooled positive likelihood ratio was 12.3 (95% CI 8.2 to 18.3) and the pooled negative likelihood ratio was 0.17 (95% CI 0.11 to 0.26). There was also considerable heterogeneity between studies of contrast-enhanced ultrasound, so these figures should be interpreted with caution, but suggest that contrast-enhanced ultrasound is a good test both for ruling in and for ruling out reflux.

Other studies that investigated the detection of reflux include IVP, indirect voiding radionuclide cystography, NAG/creatinine ratio, scintigraphy and a clinical risk scoring system. None of these tests showed both high sensitivity and specificity for the detection of reflux. IVP and indirect radionuclide voiding cystography were found to have very good specificity but poor sensitivity for the detection of reflux.

What do these results mean?

The results of studies of detection of reflux are summarised in *Table 26*. To help to understand what these results mean in practice, on estimate of the pretest probability of reflux and the likelihood ratios were used to calculate the post-test probability of reflux. Published estimates of the

TABLE 26 *Results of studies of detection of reflux*

Study	Test details	Definition of positive result	Reference standard; definition of positive result	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Standard ultrasound vs MCUG													
Baronciani, 1986[119]	Standard	Presence of reflux: dilatation or hydronephrosis	MCUG; presence of reflux	Patients	13	4	8	49	61.9	92.5	17.5	8.2	0.41
Evans, 1999[139]	Standard	Presence of reflux (change in pelvic diameter)	MCUG; presence of reflux	Renal units	2	10	17	84	10.5	89.4	1.1	1.0	1.0
Foresman, 2001[142]	Duplex	Any abnormality	MCUG; presence of reflux	Patients	24	43	25	47	49.0	52.2	1.0	1.0	0.98
Mage, 1989[175]	Standard	NS	MCUG; presence of reflux	Patients	22	5	19	76	53.7	93.8	16.0	8.7	0.49
Mahant, 2002[176]	Standard	Presence of reflux (dilatation)	MCUG; presence of reflux	Patients	14	30	21	97	40.0	76.4	2.2	1.7	0.79
Morin, 1999[183]	Standard	Renal changes indicative of APN	MCUG; presence of reflux	Patients	20	41	2	7	90.9	14.6	1.5	1.1	0.62
Muensterer, 2002[185]	Standard	Abnormal kidney size or dilatation	MCUG; presence of reflux ≥ grade 3	Renal units	21	118	2	245	91.3	67.5	17.8	2.8	0.15
		Presence of reflux (dilatation)	MCUG; presence of reflux	Renal units	35	76	34	241	50.7	76.0	3.2	2.1	0.65
		Abnormal kidney size	MCUG; presence of reflux	Renal units	20	28	49	289	29.0	91.2	4.2	3.3	0.78
		Abnormal kidney size	MCUG; presence of reflux ≥ grade 3	Renal units	11	37	12	326	47.8	89.8	8.0	4.7	0.58
		Presence of reflux (dilatation)	MCUG; presence of reflux ≥ grade 3	Renal units	18	93	5	270	78.3	74.4	9.7	3.0	0.31
Oostenbrink, 2000[186]	Standard	Presence of reflux (at least mild dilatation)	MCUG; presence of reflux	Patients	21	20	16	83	56.8	80.6	5.3	2.9	0.54
Salih, 1994[197]	Colour Doppler	Presence of reflux (blue jet)	MCUG; presence of reflux	Renal units	26	3	1	12	96.3	80.0	63.1	4.8	0.05

continued

TABLE 26 *Results of studies of detection of reflux (cont'd)*

Study	Test details	Definition of positive result	Reference standard; definition of positive result	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Tan, 1988[210]	Standard	NS	MCUG; presence of reflux	Patients	3	6	14	32	17.6	84.2	1.2	1.1	0.98
Trave, 1997[212]	Standard	NS	MCUG; presence of reflux	Renal units	3	4	14	27	17.6	87.1	1.5	1.4	0.95
Verber, 1988[214]	Standard	Presence of reflux or scarring	MCUG (Hypaque); presence of reflux	Renal units	8	9	20	25	28.6	73.5	1.1	1.1	0.97
Contrast-enhanced ultrasound vs MCUG													
Alzen, 1994[115]	Air contrast	NS	MCUG; presence of reflux	Renal units	20	6	2	73	90.9	92.4	92.7	12.0	0.10
Bergius, 1990[123]	Cystosonography (Isopaque)	Presence of reflux ≥ grade 3 (air bubbles)	MCUG; presence of reflux ≥ grade 3	Renal units	19	1	2	226	90.5	99.6	1177.8	134.7	0.11
		Presence of reflux ≥ grade 2 or air bubbles	MCUG; presence of reflux ≥ grade 2	Renal units	56	2	14	176	80.0	98.9	275.1	71.2	0.20
Berrocal, 2001[124]	Cystosonography (SH U 508A)	Presence of reflux (microbubbles)	MCUG (Plenigraf); presence of reflux	Renal units	94	29	10	307	90.4	91.4	93.8	10.5	0.11
		Presence of reflux (microbubbles)	MCUG (Plenigraf); presence of reflux	Patients	67	16	9	124	88.2	88.6	53.6	7.5	0.14
Frutos, 2000[144]	Cystosonography (Levograf)	Presence of reflux (microbubbles)	MCUG; presence of reflux	Renal units	63	19	7	204	90.0	91.5	88.8	10.6	0.11
Haberlick, 1997[149]	Colour Doppler cystosonography	Presence of reflux (blue jet)	MCUG; presence of reflux	Renal units	21	10	9	114	70.0	91.9	24.7	8.7	0.33
Kessler, 1982[165]	Cystosonography (Cysto-Conray)	Presence of reflux (microbubbles and/or dilatation)	MCUG; presence of reflux ≥ grade 2	Renal units	13	0	4	38	76.5	100.0	231.0	58.5	0.24
Mentzel, 2002[179]	Cystosonography (Levovist)	Presence of reflux	MCUG; presence of reflux	Renal units	36	10	4	174	90.0	94.6	134.8	16.6	0.11
Piaggio, 2003[187]	Cystosonography (Levovist)	NS	MCUG; presence of reflux	Renal units	42	35	32	196	56.8	84.8	7.2	3.8	0.51

continued

TABLE 26 *Results of studies of detection of reflux (cont'd)*

Study	Test details	Definition of positive result	Reference standard; definition of positive result	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Radmayr, 2002[191]	Doppler cystosonography (galactose-based contrast agent)	Presence of reflux (microbubbles)	MCUG; presence of reflux	Renal units	71	5	3	129	95.9	96.3	481.0	25.7	0.04
Rohden, 1995[195]	Cystosonography (Echovist)	NS	MCUG; presence of reflux	Patients	6	0	1	19	85.7	100.0	169.0	32.5	0.14
Schneider, 1984[199]	Cystosonography (Conray FL/air)	Presence of reflux (increased separation in the central renal echo complex)	MCUG; presence of reflux ≥ grade 2	Renal units	34	18	5	162	87.2	90.0	55.1	8.4	0.15
		Presence of reflux	MCUG; presence of reflux	Renal units	46	15	17	141	73.0	90.4	24.3	7.6	0.30
Siamplis, 1996[202]	Cystosonography (air)	NS	MCUG; presence of reflux	Renal units	15	4	3	154	83.3	97.5	152.0	32.9	0.17
	Cystosonography (fluid)			Renal units	17	8	1	150	94.4	94.9	206.6	17.2	0.08
Valentini, 2001[213]	Grey-scale cystosonography (Levovist)	Presence of reflux (microbubbles)	MCUG; presence of reflux	Renal units	34	4	8	72	81.0	94.7	65.4	15.4	0.20
	Colour Doppler cystosonography (Levovist)	Presence of reflux (colour signals)		Renal units	42	5	0	71	100.0	93.4	1105.0	13.8	0.01
Ultrasound vs cystography													
Kenda, 2000[164]	Cystosonography (levovist)	Presence of reflux (microbubbles)	Direct radionuclide voiding cystography (^{99m}Tc-pertechnetate-labelled colloid and Levovist); presence of reflux	Renal units	50	11	13	124	79.4	91.9	40.5	9.3	0.23

continued

TABLE 26 *Results of studies of detection of reflux (cont'd)*

Study	Test details	Definition of positive result	Reference standard; definition of positive result	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
IVP vs MCUG													
Cavanagh, 1983[131]	IVP (sodium meglumine diatrizoate)	Presence of renal scarring or anatomical abnormality	MCUG; presence of reflux ≥ grade 3	Patients	11	3	0	48	100.0	94.1	318.7	14.2	0.04
			MCUG; presence of reflux		13	1	16	32	44.8	97.0	17.7	10.2	0.58
Drachman, 1984[137]	IVP	NS	MCUG; presence of reflux	Patients	27	0 (19)	70	94 (75)	27.8	100.0	73.7	53.3	0.72
Redman, 1984[192]	IVP (diatrizoate)	NS	MCUG; presence of reflux	Patients	8	0 (14)	31	161 (147)	20.5	100.0	87.2	68.9	0.79
Verber, 1988[214]	IVP (Hypaque or Niopam)	Presence of renal scarring	MCUG; presence of reflux	Renal units	23	16	25	44	47.9	73.3	2.5	1.8	0.71
Other													
Bower, 1985[128]	Indirect radionuclide voiding cystography (^{99m}Tc-DTPA)	Presence of reflux	Direct radionuclide voiding cystography (^{99m}Tc-DTPA renal scan and a delayed voiding cystogram); presence of reflux	Renal units	13	1 (5)	6	34 (30)	68.4	97.1	47.8	16.2	0.34
De Sadeleer, 1994[135]	Indirect radionuclide voiding cystography (^{99m}Tc-MAG3)	Presence of reflux	MCUG; presence of reflux	Renal units	14	0 (3)	29	37 (34)	32.6	100.0	36.9	25.0	0.68
Hedman, 1978[152]	Dynamic micturating scintigraphy (^{99m}Tc-DTPA)	NS	MCUG; presence of reflux	Renal units	13	4	8	77	61.9	95.1	27.4	11.2	0.41
Johnson, 1990[162] Children with cystitis (Jodal criteria)	NAG/creatinine ratio	>1 SD from the mean	NS; presence of reflux ≥ grade 2	Patients	5	15	1	60	83.3	80.0	14.3	3.9	0.27

continued

TABLE 26 *Results of studies of detection of reflux (cont'd)*

Study	Test details	Definition of positive result	Reference standard; definition of positive result	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Children with APN (Jodal criteria)				Patients	5	24	11	26	31.3	52.0	0.5	0.7	1.30
Misselwitz, 1971[181]	Scintigraphy (^{131}I-*o*-hippurat)	Positive for reflux; semi-quantitative assessment	IVP; presence of reflux	Renal units	75	48	2	161	97.4	77.0	100.6	4.2	0.04
		Positive for reflux; qualitative assessment		Renal units	69	38	8	245	89.6	86.6	52.1	6.6	0.13
Oostenbrink, 2000[186]	Combined risk score: gender, family history, age, CRP and ultrasound	≥ 1	MCUG; presence of reflux	Patients	37	96	0	17	100.0	15.0	13.6	1.2	0.09
		≥ 6			34	64	3	39	91.9	37.9	6.0	1.5	0.24
		≥ 11			30	49	7	54	81.1	52.4	4.5	1.7	0.38
		≥ 16			24	29	13	74	64.9	71.8	4.6	2.3	0.50
		>25			19	8	18	95	51.4	92.2	11.8	6.3	0.53
		≥ 11	MCUG; presence of reflux > grade 3		25	54	3	58	89.3	51.8	7.8	1.8	0.23
		≥ 16			20	33	8	79	71.4	70.5	5.7	2.4	0.42
		>25			16	11	12	101	57.1	90.2	11.7	5.6	0.48
Trave, 1997[212]	Scintigraphy (^{99m}Tc-DMSA)	Renal changes indicative of APN	MCUG; presence of reflux	Renal units	13	8	4	23	76.5	74.2	8.3	2.8	0.34
Verber, 1988[214]	Scintigraphy (^{99m}Tc-DMSA)	Presence of renal scarring	MCUG; presence of reflux	Renal units	40	26	20	44	66.7	62.9	3.3	1.8	0.54

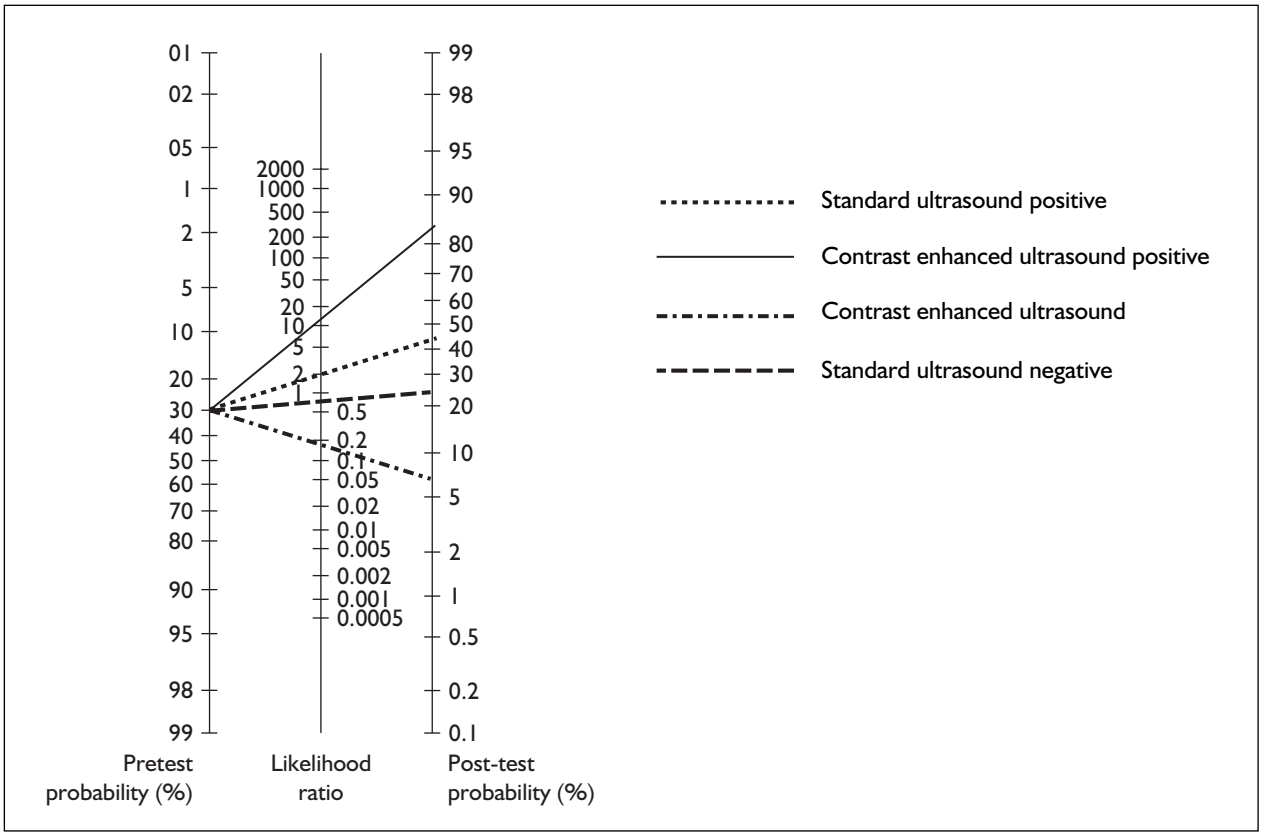

FIGURE 27 *Likelihood ratio monogram for ultrasound for the detection of reflux*

prevalence of reflux in children with UTI range from around 30 to 50%. The median prevalence of reflux in the studies included in this section of the review was 25% and ranged from 5 to 65%. The reviewers therefore decided to use 30% as an estimate of the pretest probability of reflux. *Figure 27* shows how the probability of reflux changes after the test has been given to give a post-test probability of disease in those with a positive and those with a negative ultrasound scan for standard and contrast-enhanced ultrasound.

Prediction of scarring

Four studies reporting nine data sets (*Table 27*) investigated the performance of tests for the prediction of renal scarring;[160,177,207,209] that is, the ability of a test for renal inflammation, performed during the acute phase, to predict the occurrence of renal scarring after the infection has resolved (2–24 months later in the included studies). All of these studies used follow-up ^{99m}Tc-DMSA renal scintigraphy as the reference standard. These studies were not generally of good quality. No study adequately reported details of both the index test and the reference standard. Disease progression bias was a problem in two of the studies[207,209] and insufficient information was reported to assess its avoidance in a third;[177] differential verification bias was a problem in the remaining study. In general, the diagnostic accuracy reported in these studies was poor. Two studies, one using acute Doppler ultrasound[160] and the other using acute IVP,[209] reported high specificities (92% and 99%, respectively); however, these were associated with very low sensitivities (27% and 12%). Conversely, a study of non-invasive indicators including fever and acute CRP[207] reported high sensitivity (92%) and low specificity (20%) for both. As each test was only evaluated by one or two studies, it is not possible to draw conclusions regarding their utility in the prediction of renal scarring.

Detection of scarring

Thirty studies reporting 50 data sets evaluated imaging tests for the detection of renal scarring. Of these studies, 17 used some form of renal scintigraphy as the reference standard and 13 used IVP.

IVP

Four studies evaluated the diagnostic accuracy of IVP for detecting renal scarring using a

TABLE 27 *Study results for the prediction of scarring*

Study details	Test details; time	Definition of positive result	Reference standard; definition of positive result; time	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
MCUG vs scintigraphy													
Stokland, 1998[209]	MCUG; NS	Presence of reflux	Scintigraphy (^{99m}Tc-DMSA); presence of renal scarring; follow-up	Renal units	26	38	39	209	40.0	84.6	3.7	2.6	0.71
Stokland, 1996[207]	MCUG; acute	Presence of reflux	Scintigraphy (^{99m}Tc-DMSA); presence of renal scarring; follow-up	Patients	28	17	31	80	47.5	82.5	4.2	2.7	0.64
Ultrasound vs scintigraphy													
Hitzel, 2000[177]	Colour Doppler; NS	NS	Scintigraphy (^{99m}Tc-DMSA); NS; follow-up	Renal units	15	21	8	31	65.2	59.6	2.7	1.6	0.60
Jequier, 1998[160]	Doppler; acute	Renal changes indicative of APN	Scintigraphy (^{99m}Tc-DMSA); presence of renal scarring; follow-up	Patients	18	3	49	34	26.9	91.9	3.7	3.0	0.80
	Standard; acute				43	23	58	46	42.6	66.7	1.5	1.3	0.86
Other													
Stokland, 1996[207]	Temperature; acute	≥ 38.5°C	Scintigraphy (^{99m}Tc-DMSA); presence of renal scarring; follow-up	Patients	54	78	5	20	91.5	20.4	2.6	1.1	0.44
	CRP; acute	>20 mg l^{-1}			54	78	5	20	91.5	20.4	2.6	1.1	0.44
Stokland, 1998[209]	IVP; acute	Presence of renal scarring	Scintigraphy (^{99m}Tc-DMSA); presence of renal scarring; follow-up	Renal units	8	2	57	247	12.3	99.2	14.6	12.9	0.88
	Scintigraphy (^{99m}Tc-DMSA); acute				36	44	29	205	55.4	82.3	5.7	3.1	0.54

scintigraphic technique as the reference standard.[178,180,188,209] Positive IVP investigations were found to be highly specific (above 98%) for renal scarring, with sensitivities ranging from 23% to 86%. Only one of these studies included an appropriate patient spectrum.[209] This study reported a much lower sensitivity than the other studies and was generally of slightly better quality, although all studies were of reasonable quality.

Scintigraphy

Static renal imaging

The main scintigraphic techniques evaluated for the detection of renal scarring were static renal imaging methods. Seven studies evaluated ^{99m}Tc-DMSA,[133,138,141,193,196,214,217] and one study evaluated ^{123}I-hippuran.[218] Half of these studies did not include an appropriate spectrum of patients. The quality of reporting was poor: the majority of studies (6/8) did not report patient selection criteria;[133,138,141,193,196,214] five studies provided inadequate description of the reference standard and/or the index test;[133,193,196,217,218] half of all studies did not report sufficient information to assess the avoidance of test review bias,[133,141,196,214] and four studies did not report sufficient information to assess avoidance of disease progression bias.[133,214,217,218] Disease progression bias was a problem in two studies[141,196] and partial verification bias in another study.[214] All of the studies used IVP as the reference standard. Given that renal scintigraphy is generally accepted as the reference standard for the detection of renal scarring the appropriateness of this comparison is questionable, and this should be considered when interpreting the results for this section.

When studies reported more than one data set for the same population, only one data set was included in the further analyses, to prevent duplication of participants. In one study, data were reported for patients with UTI and for patients with reflux but no UTI.[196] In this case, data for patients with UTI were used as these were most relevant to the objectives of this review. Another study reported data by patient and by renal unit. The data estimated by renal unit were used as these were deemed more pertinent to the question of detection of renal scarring, and this was the method of reporting in the majority of studies in this category.[138] Two further studies reported data for more than one scintigraphic technique.[133,218] Data from these studies for ^{99m}Tc-DMSA and ^{99m}Tc-DMSA single-photon emission computed tomography (SPECT) were included.

In general, sensitivity and specificity values reported in these studies were high. Sensitivity ranged from 85.3% (specificity 99.5%) to 100% (87.4 and 100%). Specificity ranged from 60.3% (sensitivity 94.1%) to 100% (sensitivity above 85%); specificity was above 85% in all but one study.

Likelihood ratios showed considerable heterogeneity ($p < 0.001$). Positive likelihood ratios ranged from 2.4 (LR– 0.10) to 297.7 (LR– 0.14). Negative likelihood ratios ranged from 0.01 (LR+ 7.9) to 0.15 (LR+ 168.8). The pooled positive likelihood ratio was 27.2 (95% CI 7.5 to 98.9). The pooled negative likelihood ratio was 0.13 (95% CI 0.10 to 0.18). *Figure 28* shows estimates of sensitivity and 1 – specificity plotted in ROC space. The plot suggests one possible outlier, a study with considerably poorer specificity than the other studies.[214] The median positive likelihood ratio was 59.0 (IQR 12.1–129.7) and the median negative likelihood ratio was 0.09 (IQR 0.06–0.13).

These studies were generally of relatively poor methodological quality. Only half of the studies included an appropriate patient spectrum[133,193,196,214] and only two provided sufficient details on selection criteria.[217,218] Half of the studies reported that both test review bias and reference standard review bias had been avoided,[138,193,217,218] but only one study reported that the same clinical information was available when test results were interpreted as would be available in practice.[217] None of the studies provided an explanation as to why some of the children did not receive both the index test and reference standard. The study with a very poor specificity was of similar quality to the other studies and included an appropriate spectrum of patients.[214] However, partial verification bias may have been a problem in this study. There were no other obvious differences between this study and the other studies included in this section.

Three studies compared different scinitigraphic techniques.[117,189,218] One study compared SPECT DMSA with planar DMSA and found complete agreement.[117] The other two studies compared different scintigraphic agents (^{99m}Tc-MAG3 and ^{123}I-hippuran) with scintigraphy using ^{99m}Tc-DMSA. Both studies reported specificities of 100%; sensitivity was 82% in one and 92% in the other.

Dynamic renal imaging

Two studies assessed the diagnostic accuracy of dynamic (including micturating) scintigraphy

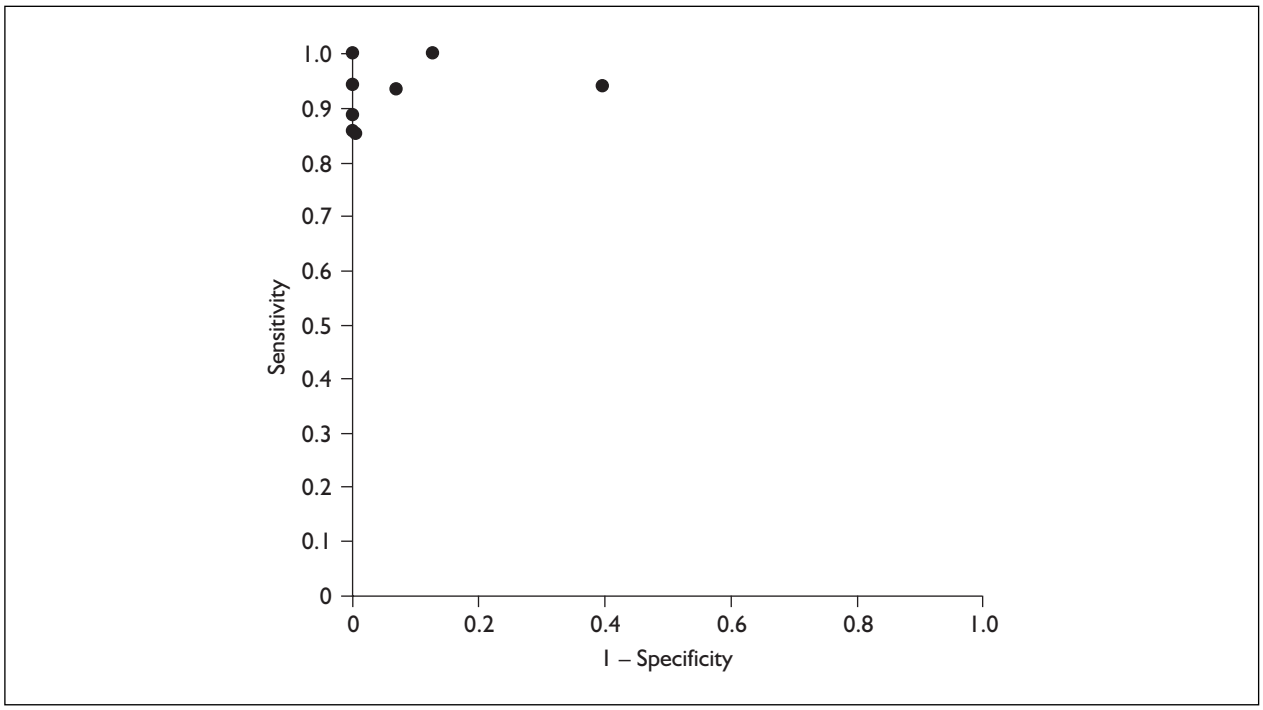

FIGURE 28 *Scintigraphy for the detection of renal scarring: study sensitivity and 1 – specificity plotted in ROC space*

using ^{99m}Tc-MAG3 with ^{99m}Tc-DMSA renal scintigraphy as the reference standard.[147,188] Both studies reported good sensitivity (88 and 82%) and specificity (88 and 95%) for this technique. Only one of these studies used an appropriate spectrum of patients,[147] and neither reported the criteria used to select participants. Disease progression bias was a problem in one study,[188] and insufficient information was reported in the other to allow this to be assessed.

MCUG

Four studies reporting five data sets investigated the presence of reflux, as determined by MCUG, to indicate the presence of renal scarring.[114,135,136,154] The reference standard was IVP in two studies[114,154] and ^{99m}Tc-DMSA-renal scintigraphy in the other two.[135,136] The presence of reflux was a relatively poor indicator of the presence of scarring: sensitivity ranged from 68 to 86%, and specificity from 37 to 82%. These studies were of generally good quality.

Ultrasound

Ten studies reporting 19 data sets investigated the use of standard ultrasound for the detection of renal scarring.[114,120,163,171,172,174,184,198,206,212] The quality of reporting was generally poor. The studies were of variable methodological quality. Only half included an appropriate spectrum of patients,[114,163,172,174,212] and only three provided an adequate description of the selection criteria.[114,174,198] Five studies reported that test review bias had been avoided[114,120,171,172,206] and all but one[172] of these also reported that reference standard review bias had been avoided. Half of the studies did not report sufficient information to assess the avoidance of disease progression bias. Disease progression bias was a problem in one study. Only one study provided details on the clinical information available when test results were interpreted and this study did not provide the same clinical data as would be available when the test results are interpreted in practice.[114] Only three studies reported how uninterpretable results were handled[114,120,206] and only one provided explanations for study withdrawals.[120]

Four studies used IVP as the reference standard[114,163,172,206] and six used ^{99m}Tc-DMSA renal scintigraphy.[120,171,174,184,198,212] One study used a qualitative scale of the probability of abnormality to define the findings of the ultrasound examination.[206] This study used a scale of 'normal, probably normal, uncertain, probably abnormal, and abnormal' to classify study results. The cut-off point of 'probably abnormal, or abnormal' was selected as a cut-off for analyses, as this was most similar to the other studies in this group.

Sensitivity varied greatly, although specificity was generally high. Sensitivity ranged from 3.4% (specificity 97.3%) to 86.5% (specificity 97.7%).

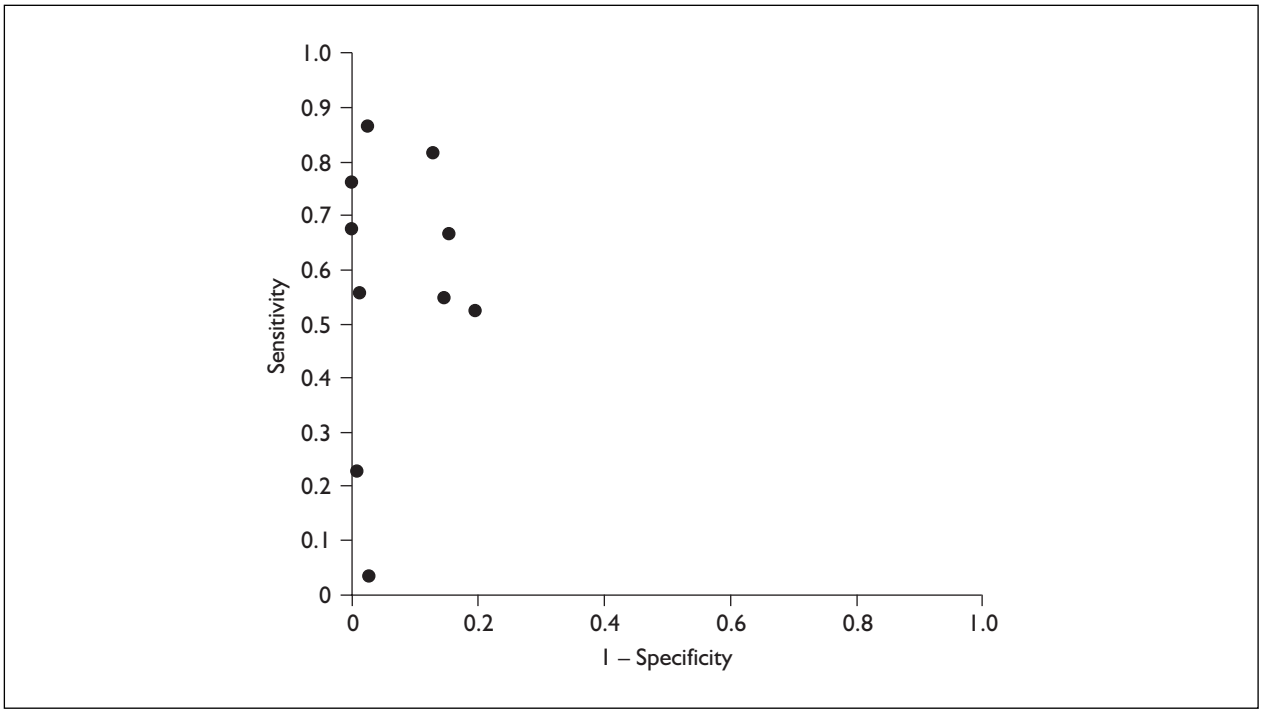

FIGURE 29 *Ultrasound for the detection of renal scarring: study sensitivity and 1 – specificity plotted in ROC space*

Specificity ranged from 80.4% (sensitivity 52.5%) to 100% (sensitivity 67.6 and 76.2%). Likelihood ratios showed considerable heterogeneity ($p < 0.0001$). Positive likelihood ratios ranged from 1.3 (LR– = 0.99) to 38.9 (LR– =0.78). Negative likelihood ratios ranged from 0.14 (LR + = 37.6) to 0.99 (LR+ = 1.3). The pooled positive likelihood ratio was 10.7 (95% CI 4.5 to 25.7) and the pooled negative likelihood ratio was 0.41 (95% CI: 0.19 to 0.86). These estimates should be interpreted with extreme caution owing to the significant heterogeneity present. *Figure 29* shows estimates of sensitivity and 1 – specificity plotted in ROC space. The median positive likelihood ratio was 5.4 (IQR 3.1–29.8) and the median negative likelihood ratio was 0.49 (IQR 0.26–0.73).

Other imaging studies

Two other imaging studies were identified that used ^{99m}Tc-DMSA renal scintigraphy as the reference standard for the detection of renal scarring. One assessed the diagnostic accuracy of various MRI techniques and reported reasonable diagnostic performance.[132] This study was of reasonable quality. However, it was not conducted in an appropriate spectrum of patients, and neither the criteria used to select participants nor the time elapsing between index test and reference standard (disease progression bias) were reported. Sensitivity ranged from 81 to 100% and specificity from 78 to 91%. The second evaluated reflux, identified by indirect voiding radionuclide cystography, to indicate the presence of renal scarring.[135] This study reported a very low sensitivity of 46% and a reasonable specificity of 87%. This study was of generally good quality, but it was not conducted in an appropriate patient spectrum and the criteria used to select participants were not reported. One study that examined the detection of scarring compared a combination of ultrasound and MCUG to IVP.[114] This study reported good diagnostic performance, with a sensitivity of 91% and a specificity of 87%. However, it did not use an appropriate reference standard, and neither the index test nor the reference standard was reported in sufficient detail to permit replication.

Summary

A variety of different techniques has been investigated for the detection of scarring. The most commonly used technique was static renal scintigraphy, which was found to have good diagnostic performance when compared with IVP as the reference standard. The pooled positive likelihood ratio was 27.3 (95% CI 7.5 to 99.4) and the pooled negative likelihood ratio was 0.12 (95% CI 0.09 to 0.18), suggesting that scintigraphy is a good test for ruling in and ruling out disease. However, as renal scintigraphy itself, rather than IVP, is regarded as the appropriate reference standard, this evaluation is of limited value. A further three studies investigated the diagnostic accuracy of different scintigraphic techniques

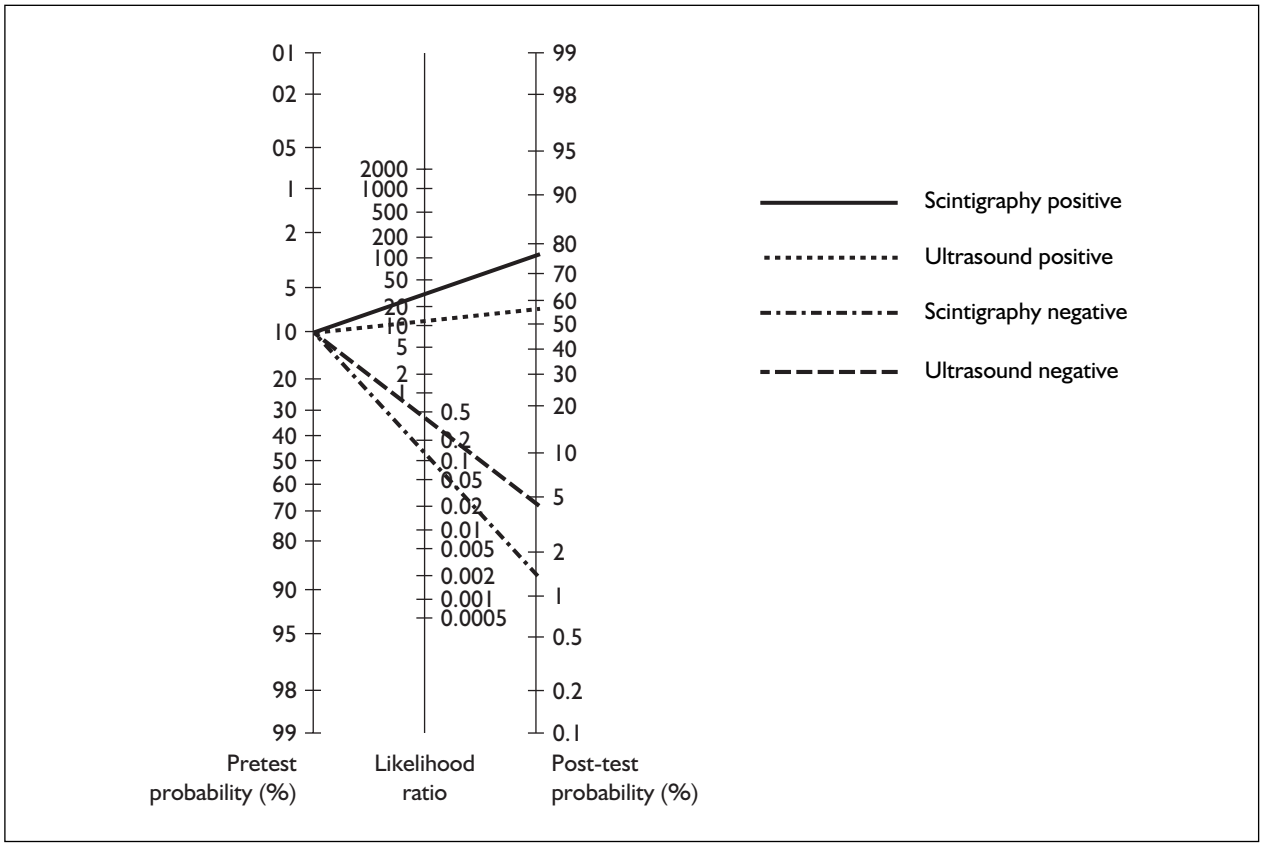

FIGURE 30 *Likelihood ratio monogram for the detection of scarring*

using the appropriate reference standard of ^{99m}Tc-DMSA scintigraphy. These studies reported specificities of 100% and sensitivities ranging from 82 to 100%. Dynamic renal imaging using ^{99m}Tc-MAG3 was investigated in two studies and was found to be a reasonable technique for the detection of scarring, with sensitivities of 82 and 88% and specificities of 88% and 95%.

Ultrasound was investigated in a reasonably large number of studies. Sensitivity showed considerable variation between studies, although specificity was generally high. The pooled positive likelihood ratio was 10.7 (95% CI 4.5 to 25.7) and the pooled negative likelihood ratio was 0.40 (95% CI 0.19 to 0.86). These figures should be interpreted with some degree of caution owing to the significant heterogeneity between the studies. However, they suggest that ultrasound is a reasonably good test for ruling in scarring; that is, a positive ultrasound scan for scarring means that the kidney is likely to be scarred. The negative likelihood ratios suggest that ultrasound is less useful for ruling out disease; that is, if the ultrasound scan is negative, the kidney may still be scarred.

Four studies found a poor association between the detection of reflux using MCUG and the presence of scarring. Sensitivity ranged from 68 to 86% and specificity from 37 to 82%. Other tests investigated 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, ranging between 23 and 86%. Indirect voiding radionuclide cystography was found to be a poor test for detecting scarring, with a very low sensitivity of 46% and a reasonable specificity of 87%. The combination of ultrasound and MCUG was found to be a reasonable test for the detection of scarring, as was MRI. However, as these have been investigated in only one study, further research is needed.

What do these results mean?

The results of studies to detect scarring are summarised in *Table 28*. To help to understand the meaning of these results in practice, an estimate of the pretest probability of scarring and the likelihood ratios were used to calculate the post-test probability of scarring. Published estimates of the prevalence of scarring in children with UTI range from around 5 to 20%.[234–237] The reviewers decided to use 10% as an estimate of the pretest probability of scarring. *Figure 30* shows how the probability of scarring changes after the test has

TABLE 28 *Results of studies to detect scarring*

Study	Test details; time	Definition of positive result	Reference standard; definition of positive result; time	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
IVP vs scintigraphy													
McLorie, 1980[178]	IVP (diatrizoate meglumine and diatrizoate sodium); NS	Presence of renal scarring	Scintigraphy (^{99m}Tc-DMSA); presence of renal scarring; NS	Renal units	24	0 (8)	5	35 (27)	82.8	100.0	316.3	58.8	0.19
Merrick, 1980[180]	IVP; NS	NS	Scintigraphy (^{99m}Tc-glucoheptonate or ^{99m}Tc-DMSA); NS; NS	Renal units	47	0 (2)	8	100 (98)	85.5	100.0	1123.2	171.3	0.15
Pickworth, 1992[188]	IVP; NS	NS	Scintigraphy (dynamic including micturating) (^{99m}Tc-MAG3); presence of renal scarring or reflux; NS	Patients	13	0	9	62	59.1	100.0	177.6	74.0	0.42
Stokland, 1998[209]	IVP; follow-up	Presence of renal scarring	Scintigraphy (^{99m}Tc-DMSA); presence of renal scarring; follow-up	Renal units	14	5	51	244	21.5	98.0	12.5	10.0	0.80
Scintigraphy vs IVP													
Clarke, 1990[133]	^{99m}Tc-DMSA SPECT; NS	NS	IVP; NS; NS	Renal units	14	2	1	27	93.3	93.1	106.3	10.9	0.10
	^{99m}Tc-DMSA planar; NS			Renal units	14	6	1	23 (23)	93.3	79.3	34.9	4.2	0.12
Elison, 1992[138]	^{99m}Tc-DMSA; follow-up	Presence of renal scarring	IVP (Omnipaque); presence of reflux or scarring; follow-up	Renal units	50	46	0	319	100.0	87.4	694.0	7.8	0.01
		Presence of renal scarring		Patients	45	0 (32)	0	163 (131)	100.0	100.0	29757.0	324.4	0.01
Farnsworth, 1991[141]	^{99m}Tc-DMSA; acute and follow-up	Presence of renal scarring	IVP (Omnipaque); abnormal renal tract; NS	Renal units	18	0 (49)	3	176 (127)	85.7	100.0	1865.9	297.7	0.16
Rehling, 1989[193]	^{99m}Tc-DTPA; follow-up	Abnormal urinary tract	IVP (Omnipaque or Telebrix); abnormal urinary tract; follow-up	Patients	17	0 (5)	1	20 (15)	94.4	100.0	478.3	38.7	0.08

continued

TABLE 28 *Results of studies to detect scarring (cont'd)*

Study	Test details; time	Definition of positive result	Reference standard; definition of positive result; time	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Rossleigh, 1990[196]	^{99m}Tc-DMSA; >1 month	Presence of renal scarring	IVP; NS; > 1 month	Patients (UTI)	8	0 (18)	0	41 (23)	100.0	100.0	1411.0	79.3	0.06
				Patients (reflux, no UTI)	0	0 (4)	0	14 (10)		100.0	29.0	15.0	0.52
Verber, 1988[214]	^{99m}Tc-DMSA; NS	Presence of renal scarring	IVP (Hypaque or Niopam); presence of renal scarring; NS	Renal units	32	23	2	35	94.1	60.3	19.6	2.3	0.12
Whitear, 1990[217]	^{99m}Tc-DMSA; NS	NS	IVP; NS; NS	Renal units	162	1 (45)	28	197 (153)	85.3	99.5	750.7	112.9	0.15
Wujanto, 1987[218]	^{123}I-hippuran; NS	Presence of renal scarring	IVP; presence of renal scarring; NS	Renal units	22	0 (2)	4	66 (64)	84.6	100.0	665.0	111.7	0.17
	^{99m}Tc-DMSA; NS				23	0 (3)	3	66 (63)	88.5	100.0	893.0	116.6	0.13
Scintigraphy vs scintigraphy													
Bagni, 1997[117]	Scintigraphy (^{99m}Tc-DMSA SPECT); NS	Presence of renal scarring	Scintigraphy (^{99m}Tc-DMSA planar); presence of renal scarring; NS	Patients	18	0 (16)	0	29 (13)	100.0	100.0	2183.0	58.4	0.03
Piepsz, 1992[189]	Scintigraphy (^{99m}Tc-MAG3); acute	Presence of renal scarring	Scintigraphy (^{99m}Tc-DMSA); renal changes indicative of APN; acute	Renal units	18	0	4	38	81.8	100.0	316.6	62.7	0.20
Wujanto, 1987[218]	Scintigraphy (^{123}I-hippuran); NS	Presence of renal scarring	Scintigraphy (^{99m}Tc-DMSA); presence of renal scarring; NS	Renal units	24	0	2	66	92.3	100.0	1303.4	121.6	0.09
Dynamic vs standard scintigraphy													
Gordon, 1992[147]	Dynamic including micturating (^{99m}Tc-MAG3); follow-up	NS	^{99m}Tc-DMSA; presence of renal scarring; follow-up	Renal units	44	7	6	53	88.0	88.3	48.8	7.1	0.15

continued

TABLE 28 *Results of studies to detect scarring (cont'd)*

Study	Test details; time	Definition of positive result	Reference standard; definition of positive result; time	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Pickworth, 1992[188]	Dynamic including micturating (^{99m}Tc-MAG3); NS	Presence of renal scarring or reflux	^{99m}Tc-DMSA; NS; NS	Patients	14	2	3	36	82.4	94.7	60.5	12.6	0.21
MCUG vs IVP													
Alon, 1986[114]	MCUG (Urographin); 5–7 weeks	Presence of reflux	IVP (Urographin); NS; 5–7 weeks	Patients	15	8	7	51	68.2	86.4	12.5	4.8	0.38
Hellstrom, 1989[154]	MCUG; NS	Presence of reflux	IVP; presence of renal scarring; NS	Patients	8	19	2	55	80.0	74.3	9.7	3.0	0.31
MCUG vs scintigraphy													
De Sadeleer, 1994[135]	MCUG (iodinated contrast material); follow-up	Presence of reflux	^{99m}Tc-DMSA; presence of scarring; acute	Renal units	19	24	7	14	73.1	36.8	1.5	1.1	0.75
Ditchfield, 1994[136]	MCUG; acute	Presence of reflux ≥ grade 2	^{99m}Tc-DMSA; renal changes indicative of APN; acute	Patients	32	18	35	65	47.8	78.3	3.2	2.2	0.67
				Renal units	34	38	54	174	38.6	82.1	2.9	2.1	0.75
Ultrasound vs IVP													
Alon, 1986[114]	Standard; acute	NS	IVP (Urographin); NS; 5–7 weeks	Patients	16	0	5	60	76.2	100.0	363.0	91.5	0.25
Kenda, 1989[163]	Standard; <3 months	NS	IVP; NS; < 3 months	Patients	5	1 (2)	4	91 (90)	55.6	98.9	74.6	34.1	0.46
Lindsell, 1986[172]	Standard; NS	Any abnormality	IVP [Iopamidol (Niopam)]; abnormal renal tract; NS	Patients	25	0	12	63	67.6	100.0	259.1	85.9	0.33
Stokland, 1994[206]	Standard; NS	Renal scarring/renal size: abnormal or probably abnormal, or uncertain, or probably normal	IVU; presence of renal scarring; NS	Renal units	70	49	27	94	72.2	65.7	4.9	2.1	0.43

continued

TABLE 28 *Results of studies to detect scarring (cont'd)*

Study	Test details; time	Definition of positive result	Reference standard; definition of positive result; time	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
		Renal scarring/renal size: abnormal or probably abnormal			53	21	44	122	54.6	85.3	6.9	3.7	0.53
		Renal scarring: abnormal or probably abnormal			44	20	53	123	45.4	86.0	5.0	3.2	0.64
		Renal scarring: abnormal or probably abnormal, or uncertain			52	29	45	114	53.6	79.7	4.5	2.6	0.58
		Renal scarring: abnormal			25	7	72	136	25.8	95.1	6.4	5.0	0.78
		Renal scarring/renal size: abnormal or probably abnormal, or uncertain			62	30	35	113	63.9	79.0	6.6	3.0	0.46
		Renal scarring: abnormal or probably abnormal, or uncertain, or probably normal			57	43	40	100	58.8	69.9	3.3	1.9	0.59
		Renal scarring/renal size: abnormal			32	7	65	136	33.0	95.1	9.0	6.4	0.71
Ultrasound vs scintigraphy													
Barry, 1998[120]	Standard; 1–3 months	Presence of renal scarring	^{99m}Tc-DMSA; NS; follow-up	Renal units	147	11	23	467	86.5	97.7	255.2	35.9	0.14
LeQuesne, 1986[171]	Standard; NS	Presence of renal scarring or signs of reflux	^{99m}Tc-DMSA; NS; NS	Renal units	22	5	5	34	81.5	87.2	25.7	5.8	0.23
MacKenzie, 1994[174]	Standard; acute	Any abnormality	^{99m}Tc-DMSA; renal changes indicative of APN; acute	Patients	32	10 (14)	29	41 (37)	52.5	80.4	4.4	2.6	0.60
Mucci, 1994[184]	Standard; NS	NS	^{99m}Tc-DMSA; NS; NS	Patients	5	1	17	170	22.7	99.4	35.7	27.4	0.77

continued

TABLE 28 *Results of studies to detect scarring (cont'd)*

Study	Test details; time	Definition of positive result	Reference standard; definition of positive result; time	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Scherz, 1994[198]	Standard; NS	Presence of renal scarring	^{99m}Tc-DMSA; presence of renal scarring; follow-up	Patients (asympto-matic)	3	1	0	22	100.0	95.7	105.0	14.0	0.13
				Patients (symptomatic)	9	9	6	33	60.0	78.6	5.2	2.7	0.52
				Patients (all)	12	10	6	55	66.7	84.6	10.2	4.1	0.41
Trave, 1997[212]	Standard; acute	NS	^{99m}Tc-DMSA; renal changes indicative of APN; acute	Renal units	1	1	28	36	3.4	97.3	1.3	1.3	0.99
Other													
Alon, 1986[114]	Combination (ultrasound and MCUG)	Either positive	IVP (Urographin); NS; 5–7 weeks	Patients	19	8	2	52	90.5	86.7	48.2	6.4	0.13
Chan, 1999[132]	MRI (gadolinium-enhanced STIR); follow-up	Presence of renal scarring	Scintigraphy (^{99m}Tc-DMSA); presence of renal scarring; follow-up	Renal units	15	6	1	26	93.8	81.3	42.1	4.6	0.11
	MRI (STIR or T1-W); follow-up				16	7	0	25	100.0	78.1	112.2	4.3	0.04
	MRI (T1-W not gadolinium); follow-up				13	3	3	29	81.3	90.6	32.5	7.5	0.23
De Sadeleer, 1994[135]	Indirect voiding radionuclide cystography (^{99m}Tc-MAG3); acute	Presence of reflux	Scintigraphy (^{99m}Tc-DMSA); presence of scarring; acute	Renal units	12	5	14	33	46.2	86.8	5.3	3.3	0.63

STIR, short T1 inversion recovery. T1-W, T1-weighted.

TABLE 29 *Results of imaging studies with multiple aims*

Study	Aims	Test details; time	Definition of positive result	Reference standard; definition of positive result; time	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
Benigno, 1986[122]	Detection of scarring and reflux	Temperature; acute	≥ 38.5°C	IVP and MCUG; NS; follow-up	Patients	10	7	23	9	30.3	56.3	0.6	0.7	1.24
		CRP; acute	Not clear			25	9	8	7	75.8	43.8	2.4	1.3	0.57
		ESR; acute	25 mm h^{-1}			25	10	8	6	75.8	37.5	1.9	1.2	0.65
		Leucocyte differential	NS			20	6	13	10	60.6	62.5	2.5	1.6	0.64
		Renal concentrating capacity	1025			23	4	10	12	69.7	75.0	6.2	2.6	0.42
Hanbury, 1989[151]	Detection of scarring and reflux	Ultrasound and IVP; NS	NS	IVP and MCUG; NS; NS	Patients	10	18 (19)	4	277 (276)	71.4	93.9	35.0	11.2	0.32
Jequier, 1985[159]	Detection of anatomical abnormalities, reflux and scarring	Ultrasound (standard); acute	NS	IVP and/or MCUG; NS; <5 months after acute phase (usually 3-6 weeks)	Patients	73	25	24	118	75.3	82.5	13.9	4.2	0.30
Johnson, 1985[161]	Identification of treatable urologic problems	Age	<5 years	Ultrasound, IVP (sodium diatrizoate), and voiding cystography (MCUG for boys); presence of reflux or a treatable medical problem; 4–6 weeks	Patients	10	33	1	25	90.9	43.1	5.3	1.5	0.29
		Temperature; acute	>38°C rectally/ >37°C orally			10	14	1	44	90.9	75.9	21.5	3.6	0.17
		Fever, ESR, CRP, DDAVP and renal concentration ability; acute	Any two of the four positive			10	18	1	40	90.9	69.0	15.3	2.8	0.18

continued

TABLE 29 *Results of imaging studies with multiple aims (cont'd)*

Study	Aims	Test details; time	Definition of positive result	Reference standard; definition of positive result; time	Unit of analysis	a	b	c	d	Sensitivity	Specificity	DOR	LR+	LR–
		IVP (sodium diatrizoate); follow-up	Presence of reflux or a treatable medical problem			1	0 (3)	10	58 (55)	9.1	100.0	16.7	14.8	0.88
		MCUG; 4–6 weeks				11	0	0	58	100.0	100.0	2691.0	113.1	0.04
		Ultrasound (standard); 4–6 weeks				1	0	10	58	9.1	100.0	16.7	14.8	0.88
Leonidas, 1985[170]	Detection of urinary tract abnormalities	Ultrasound (standard); NS	Any abnormality	IVP (sodium diatrizoate); abnormal urinary tract; NS	Patients	18	24	0	25 (25)	100.0	51.0	38.5	2.0	0.05
Rickwood, 1992[194]	Detection of anatomical abnormalities, reflux and scarring	Ultrasound (standard); NS	NS	IVU + cystography/ DMSA; NS; >4 weeks DMSA, NS for other exams	Patients	34	1	45	120	43.0	99.2	60.9	35.1	0.58
Smellie, 1995[203]	Detection of scarring and reflux	Ultrasound (standard); <12 months	NS	IVU, MCUG and DMSA; more than one investigation abnormal; <12 months	Patients	17	2	23	16	42.5	88.9	4.9	3.2	0.66
Volti, 1991[216]	Detection of urinary tract abnormalities	Ultrasound (standard); acute	Any abnormality	IVP (Iopamidol); NS; acute	Patients	23	2	7	24	76.7	92.3	30.7	8.2	0.27

DDAVP, desamino-D-arginine vasopressin.

been given to provide a post-test probability of disease in those with a positive and negative ultrasound or scintigraphy scan.

Imaging studies with multiple aims

Eight studies reporting 17 data sets examined the performance of a variety of tests, and combinations of tests, to detect groups of renal and urinary tract pathologies[122,151,159,161,170, 194,203,216] (*Table 29*). The studies used a wide variety of tests and test combinations as reference standards. The diagnostic accuracies reported by studies were generally poor. With the exception of one study,[161] none of the studies reported estimates of both sensitivity and specificity above 80%. The study that reported good diagnostic performance found 100% sensitivity and specificity for MCUG in the diagnosis of the presence of reflux or a treatable medical problem compared with a reference standard consisting of a combination of ultrasound, IVP and voiding cystography.[161] Studies in this section had a variety of quality problems, including reporting quality, disease progression bias and verification bias. The majority of these studies (5/8) did not describe either the index test or the reference standard in sufficient detail to permit replication.

Effectiveness of follow-up

One RCT evaluated the effectiveness of further investigation of UTI using imaging techniques. This study was published as an abstract,[219] and the reviewers were unable to obtain further data.

The objective was to determine whether routine imaging, using ultrasound and MCUG, of low-risk children with their first uncomplicated UTI significantly reduced renal scarring or recurrent UTI. Participants, toilet-trained children aged 2–10 years, with confirmed UTI, were allocated to routine or selected imaging. All children referred for routing imaging received ultrasound and MCUG. Those referred for selected imaging received ultrasound and MCUG only if they had recurrent UTI or persistent problems. Children in both groups diagnosed with reflux were prescribed prophylactic antibiotics for the duration of the study. All children were assessed for renal scars (using DMSA renal scintigraphy) and recurrent UTI after 2 years.

In total, 172 children were enrolled; 22 withdrew immediately following randomisation, 15 because of concerns about imaging. For the remaining 150, no differences in age, gender, duration or type of symptoms were observed between the groups. The median age was 3.9 years, and 21% were boys. All but one of the routine imaging group had both investigations, while only 21% of the selected imaging group had imaging during follow-up. Reflux was diagnosed in 30% of children who had MCUG, with no differences between the groups. More children in the routine imaging group received antibiotic prophylaxis (28% versus 5%, $p < 0.0001$), but there was no significant difference in the proportion of children with recurrent UTIs (26% versus 21%, $p = 0.5618$). Only 65% of the children (62% in the routine imaging group and 67% in the selected imaging group) underwent DMSA after 2 years of follow-up. There was no difference in the rate of renal scars between the two groups (9% versus 9%, $p = 0.6430$).

The authors concluded that routine imaging of toilet-trained preschool and school-aged children with their first uncomplicated UTI leads to higher rates of imaging, identification of reflux and prophylaxis than does selected imaging. However, it does not lead to a reduction in recurrent UTIs or renal scarring, indicating that routine imaging is not worthwhile in this older, low-risk group.

As this study was only published as an abstract limited data were available on the quality of the study. It was not possible to determine whether the randomisation method was appropriate, whether treatment allocation was concealed, whether outcome assessors were blinded to treatment group, and whether groups were treated the same apart from receiving the stated interventions. It is interesting to note that the proportion of children receiving the follow-up DMSA was relatively low in both treatment groups. The authors do not report whether there was any relationship in the selected imaging group between those receiving MCUG and ultrasound and those receiving the follow-up DMSA.

Chapter 6

Economic analysis

Introduction

This chapter presents evidence on the cost-effectiveness of alternative diagnostic strategies for the diagnosis and further investigation of UTI in children under 5. The chapter has the following objectives:

- to identify published economic studies that satisfy the inclusion criteria described in the section 'Economic evaluations' (p. 10)
- to consider the necessary requirements of any economic evaluation seeking to inform decision-making regarding the use of diagnostic strategies in UTI in children
- to structure, populate and analyse a decision-analytic model to assess the cost-effectiveness of alternative strategies in the NHS.

Review of published economic evaluations

Introduction

The search strategies identified 17 papers that were potentially relevant for the economic review, and six were ordered for detailed assessment. Of these, only one study, by Downs,[10] satisfied the inclusion criteria. It consists of a comparison of a large number of diagnostic strategies relating to UTI and the identification of urinary tract abnormalities, and a model that links evidence on diagnostic accuracy with that on therapeutic decisions and hence on health outcomes; and evidence is identified using systematic review methods.

Downs study: overview

The key features and quality of the Downs study are summarised in Appendix 8. The context of the analysis was to help to inform recommendations developed by the Urinary Tract Sub-Committee of the American Academy of Pediatrics Committee on Quality Improvement. The study took the form of a decision-tree model to compare alternative strategies for the diagnosis and management of UTI in children aged between 2 months and 2 years. A conceptual model was developed to inform the development of the decision tree, which had six elements: (1) the prior probability of UTI, which was a function of the clinical presentation and patient demographics; (2) the sensitivity and specificity of the strategy used to diagnose UTI, which revises the probability of a given child having a UTI; (3) short-term treatment for UTI; (4) the prior probability of VUR or obstruction; (5) the use of strategies to diagnose these abnormalities; and (6) the development of renal damage through recurrent UTI.

The analysis used a payer perspective. The time-horizon was not stated explicitly, but appeared to be patient's lifetime as the cost of long-term sequelae such as ESRD were estimated. Health outcomes were presented in natural units in terms of cases of ESRD, hypertension and death. A range of one- and two-way sensitivity analyses was undertaken.

Downs study: options compared

The analysis compared the cost-effectiveness of four options for the diagnosis of UTI:

- doing nothing (anchor option)
- treating all patients
- using a dipstick, confirming positives with culture of suprapubic bladder tap or transurethral catheterisation
- culturing all, suprapubic bladder tap or transurethral catheterisation.

Three imaging strategies were also evaluated:

- renal ultrasonography and MCUG (referred to as VCUG
- renal ultrasonography alone
- no imaging.

Downs study: model structure

The key structural elements of the Downs model are summarised below.

- For positive UTI tests, treatment is initiated and complications (including the rare event of death from anaphylaxis) are accounted for.
- The risk of urosepsis from a UTI is modelled, including the small risk of hospitalisation and of death.
- Patients with positive tests for UTI are then subject to one of the imaging options to identify reflux. False positives for UTI are assumed

TABLE 30 *Key parameter values used in the base case of the Downs model*[10]

Parameter	Base-case value
Prevalence of UTI	0.05
Sensitivity of culture	1.0
Specificity of culture	1.0
Sensitivity of bag culture	1.0
Specificity of bag culture	0.7
Sensitivity of urinalysis	0.92
Specificity of urinalysis	0.70
Probability of complication with antibiotics for UTI	0.10
Risk of death due to complication with antibiotics for UTI	1.0×10^{-4}
Risk of sepsis with UTI	0.09
Efficacy of antibiotics for UTI	0.95
Probability that sepsis will resolve	0.7
Probability of death from sepsis	0.10
Probability of reflux	0.4
Probability of high-grade reflux	0.41
Sensitivity of renal ultrasonography for low-grade reflux	0.14
Sensitivity of renal ultrasonography for high-grade reflux	0.82
Specificity of renal ultrasonography for reflux	1.0
Sensitivity of full imaging evaluation	1.0
Specificity of full imaging evaluation	
Risk of renal scarring without reflux	0.07
Risk of renal scarring with low-grade reflux	0.13
Risk of renal scarring with high-grade reflux	0.53
Risk of hypertension with scarring	0.20
Risk of ESRD after scarring	0.05

never to have urinary tract abnormalities and are not at increased risk of renal damage.
- Children with a true-positive UTI may or may not have reflux and, if they do, this is graded as low or high.
- The sensitivity of imaging is assumed to vary with grade of reflux. All imaging strategies are assumed to be 100% specific.
- Patients identified on imaging as having reflux are assumed to be treated surgically or with prophylactic antibiotics, although this does not seem to have been costed in the analysis.
- The risk of recurrent UTI (assumed to be more than three infections over a 5-year period) is modelled, which is assumed to increase the risk of progressive renal scarring (PRS). The risk of scarring is higher in children with reflux. Surgery for reflux can reduce the risk of scarring. Prophylactic antibiotics can reduce the risk of subsequent UTIs.
- Renal scarring is assumed to increase the risk of ESRD and hypertension.

Downs study: model inputs

Many of the inputs in the model were identified by an extensive systematic literature review process. Separate reviews were undertaken for recognition of children at risk of UTI, diagnosis of UTI, short-term treatment of UTI, imaging evaluation and the costs of chronic hypertension and ESRD. Some degree of quality assessment of studies identified in the review was also undertaken.

The evidence available to parameterise the elements of the model was clearly weaker for some elements than for others. In particular, the link between recurrent UTIs and renal scarring, conditional of reflux grade, and the link between renal scarring and renal damage was not based on longitudinal data. Rather, a series of indirect links in the evidence was used to model the relationship.

Cost data were taken from a single US hospital, with the exception of the costs of long-term complications, which were taken from published literature.

Table 30 presents some of the key parameters (at their base-case values) used in the Downs model.

Downs study: results

With respect to the diagnosis of UTI, the Downs analysis found that a do-nothing option (observation) was the least costly and using culture bag urine the most costly. The incremental cost

per additional case prevented ranged from US$61,000 to 434,000.

For the imaging options, no evaluation represented the cheapest option, followed by renal ultrasonography alone, and renal ultrasonography plus VCUG was the most costly. The incremental cost per additional case prevented from renal ultrasonography alone compared with do nothing was $260,000, and of renal ultrasonography plus VCUG versus renal ultrasonography alone was $353,000.

A range of sensitivity and threshold analyses was undertaken. These used a threshold willingness-to-pay value for an untoward clinical event of $700,000. This is quite arbitrary, and hence the results of the analysis are difficult to interpret.

Implications from the published cost-effectiveness results from the literature

To inform the identification of relevant published economic evaluations, and the development of a decision-analytic model, it is important to establish the key features of an economic evaluation that is informing resource allocation in the NHS. These requirements are discussed in the methodological guidance issued by NICE for economic evaluations undertaken to inform its technology appraisal process.[31]

1. The specification of the decision problem should ideally include the comparison of all diagnostic strategies that could feasibly be used in the NHS. It is recognised, however, that, in practice, these options may be constrained by the availability of evidence and the structural complexity of any model.
2. The analysis should make a clear link between the diagnostic accuracy of a given strategy, the impact on therapy and the ultimate effect on health outcomes. Hence, the effect of each of the four diagnosis groups – true positive, false negative, true negative and false positive – for the selection of therapies needs to be assessed, and the effect of such therapy for outcomes.
3. A lifetime time-horizon is required for any economic evaluation in this area. This is because the sequelae of inappropriately managed UTI in children, in particular with respect to renal disease, can last for the remainder of their life.
4. The ultimate health effects of the alternative diagnostic strategies should be expressed in terms of a generic measure of health such as a quality-adjusted life-year (QALY). This is because it is necessary to assess the value of improved outcomes from more accurate diagnostic tests in units that can be compared with those of programmes and interventions in other specialities and disease areas that are competing for finite healthcare resources.
5. The evidence that is used to establish the cost-effectiveness of the alternative diagnostic strategies needs to be identified systematically and synthesised appropriately. The evidence will relate both to the diagnostic accuracy of the tests and to parameters related to, for example, the effectiveness of treatment and the quality of life impact of UTI.
6. The evidence used to estimate cost-effectiveness should be relevant to patients and clinical practice in the UK health service.
7. The uncertainty in the evidence base needs to be reflected in the model. To assess simultaneously the implications of uncertainty in all elements of evidence, probabilistic analysis should be used to establish the decision uncertainty associated with each diagnostic strategy being compared.[238–240] This informs decision-makers about the probability of each strategy being the most cost-effective conditional on the value that the decision-maker places on a unit of health gain. Such methods can be used to provide an opportunity to use value of information methods to inform priority setting in research.[241,242]

The Downs study has four major limitations for NHS decision-making. First, it does not express outcomes in terms of a generic measure of health such as a QALY. Secondly, it is a US study seeking to inform decision-making in the USA and, as such, some of its parameter estimates and assumptions may be inappropriate for NHS practice. Thirdly, it does not undertake probabilistic sensitivity analysis to translate the uncertainty in the evidence base (in the form of the parameter values in the model) into decision uncertainty about the diagnostic strategies. Finally, it considers only a few of the possible diagnostic and imaging strategies that are relevant for UK decision-making.

Despite its limitations and the fact that it is unlikely to provide a direct guide to resource allocation decisions in the NHS, it provides important insights about the modelling of UTI in children and the interpretation of the evidence base that are valuable for the modelling undertaken as part of this study.

Development of a cost-effectiveness model

Introduction

Given the limitations of the available literature on the cost-effectiveness of alternative diagnostic strategies for UTI in children, it is necessary to develop a new analysis to inform resource allocation decisions. The new analysis is in the form of a decision-analytic model that synthesises available data to identify the optimal strategy regarding the diagnosis and further investigation of children under 5 with suspected UTI. The analysis seeks to satisfy the requirements for decision-making described in the previous section.

Methods

Overview

The objective of the model is to estimate, based on all available data, long-term costs and health outcomes from a range of alternative strategies for the diagnosis and further investigation of UTI in children under 5 years of age. The model starts at the point at which the first UTI is suspected and the diagnostic process is enacted. Costs are estimated from the perspective of the NHS and include short-term diagnostic and treatment costs, and also the cost of long-term complications. Health outcomes are expressed in terms of reductions in patients' quality-adjusted life expectancy (decrements in QALYs). The QALY decrements represent the effects of treated and untreated UTIs and the impact of renal scarring on the risk of end-stage renal failure. All costs and benefits are discounted at 6% and 2% per annum, respectively.[243] The optimal strategy is the one with the highest expected net monetary benefit, and probabilistic sensitivity analysis was undertaken to establish the probability with which each strategy is optimal.

Model structure

The model is made up of two parts, short-term and long-term. The short-term model is shown in *Figure 31* in the form of a decision tree, and relates to the process of diagnostic testing for UTI and reflux. The first branches indicate the (unknown) UTI status of the child. The second set of branches shows the result of the UTI test. The third set of branches relates to a child's reflux status. For children without UTI, their reflux status and future prognosis are not considered. For children who are sent for further investigation, the third set of branches shows the results of that test. This is a generic decision tree in that some groups of branches will not be used for particular strategies; for example, if imaging is not undertaken as part of a particular strategy, then the imaging branches will not be part of the tree.

The second element of the model is the long-term model, which estimates long-term cost and QALY decrements for children depending on which pathway of the decision tree they pass along. The long-term model is shown in *Figure 32*. Again this model is generic, and the specific version used for a given pathway in the short-term model may differ. This model takes the form of a Markov process, and is developed from earlier work looking at the cost-effectiveness of prophylactic antibiotics in children with recurrent UTIs.[244] For a given pathway, the model provides a link between the number of UTI attacks that a child will experience, which varies from none to four, the proportion of these that are pyelonephritic, the probability of progressive renal scarring, the risk of ESRD and the form of management for that disease.

Structural assumptions

In structuring the short- and long-term models, and in relating the two, a series of assumptions has been made. Within the short-term model, the first assumption is that, following a positive test for a UTI, children will be treated with a course of antibiotics. Each time a recurrent UTI occurs there is a chance that this will be a pyelonephritic attack, which has different cost and quality of life implications from a UTI attack. Antibiotic treatment is assumed to resolve the infection after 3 days for a lower tract infection and 10 days for a pyelonephritic attack.[245,246] During these attacks, children will experience a decrement in their health-related quality of life, which is expressed as a reduction in utility measured on a 0 (equivalent to death) to 1 (equivalent to good health) utility or preference scale. The second assumption is that false-negative UTI tests will result in a UTI not being treated and hence not resolving for 7 days following a lower tract infection and 14 days after a pyelonephritic attack.

The third structural assumption is that children who are confirmed positive for reflux will be treated either surgically or using long-term low-dose prophylactic antibiotics. The effect of these treatments is to reduce the risk of the child experiencing recurrent UTIs. The evidence on the comparative effectiveness of these interventions is weak. Given this inconclusive evidence, it is assumed that surgery and long-term low-dose prophylactic antibiotics would have the same treatment effect.[247] The purpose of the model is not, however, to identify optimal treatment for

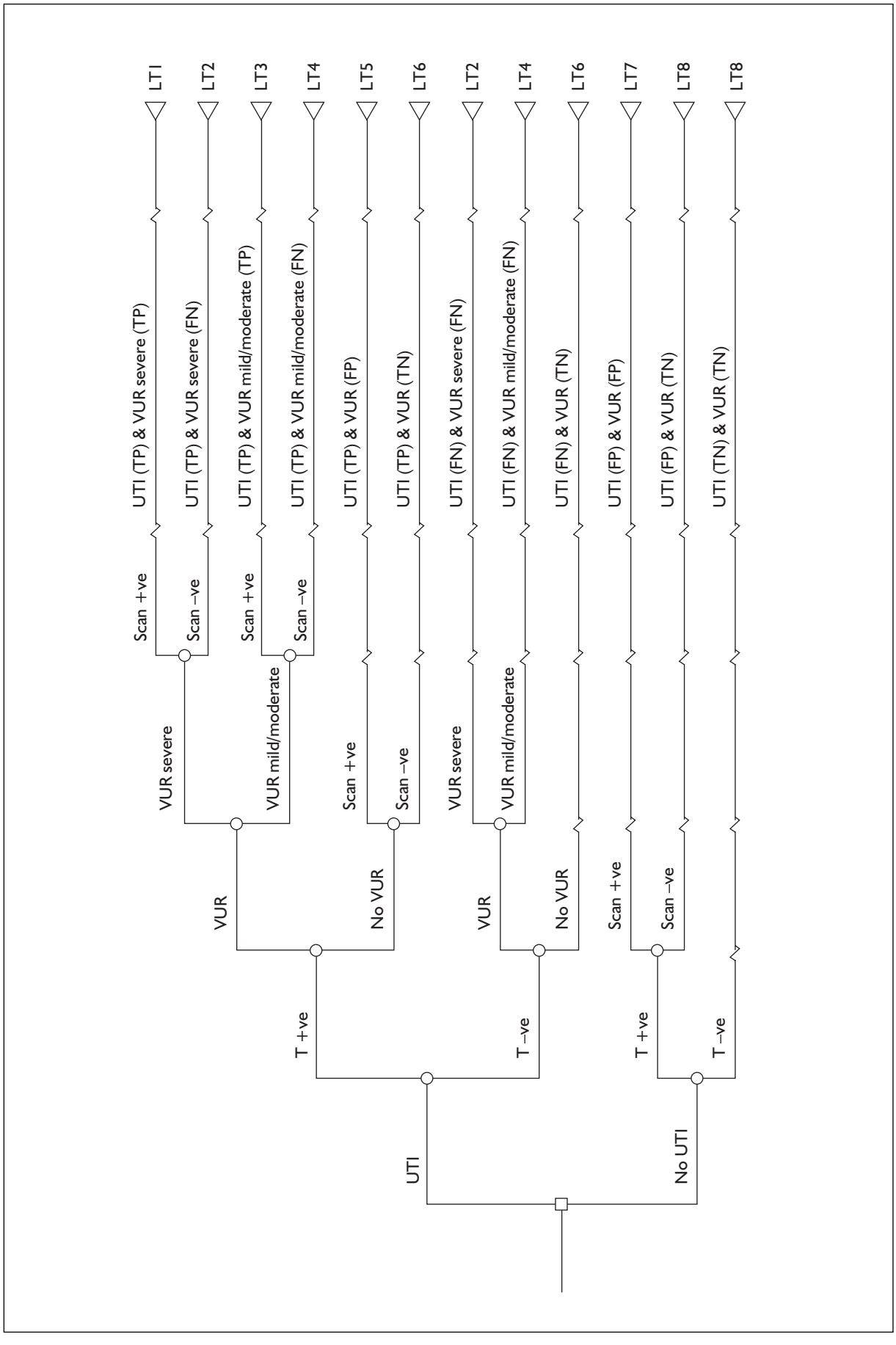

FIGURE 31 *Structure of the short-term model in the form of a generic decision tree. FN, false negative; FP, false positive; TN, true negative; TP, true positive.*

reflux. The model is, therefore, structured to reflect the prevailing opinion in clinical practice regarding the use of these interventions. Hence, it is assumed that treatment will be surgical if the child has severe reflux, and through the use of prophylactic antibiotics if they have mild to moderate reflux.

The fourth assumption is that, following a positive ultrasound (standard or contrast-enhanced) for reflux, children will be undergo MCUG, which is assumed to be a definitive test. The rationale for this is that it was considered implausible that children would be referred for surgery solely on this basis of ultrasound.

A series of assumptions is also made in the long-term models. The first relates to how children are treated if they present with recurrent UTIs. Following consultation with clinical experts, it is assumed that if a child experiences two recurrent UTIs (i.e. three in total), they will undergo MCUG. If that test indicates that they have severe reflux they will be treated surgically; otherwise, they will receive prophylactic antibiotics. These treatments will reduce the risk of all subsequent UTIs. All subsequent UTIs are assumed to be correctly detected in general practice and are appropriately treated. Treatment costs and utility decrements are attributed as appropriate.

A second structural assumption of the long-term model is that a relationship exists between the number of pyelonephritic attacks that a child experiences and the risk of PRS. The associated cost and health decrements are a result of the elevated risk of ESRD needing treatment. The causal relationship between renal scarring and ESRD is assumed to work through both renal insufficiency and hypertension.[19] In the absence of an appropriate longitudinal study, the quantification of this relationship relies on indirect evidence, so there is considerable uncertainty associated with this structural assumption.

On the basis of these assumptions, it is possible to describe the management of children moving along each of the pathways shown in *Figure 31*.

- For pathway LT1, children are tested true positive for UTI (and are treated for the infection) and for severe reflux. All children testing positive for reflux are assumed to undergo definitive MCUG. These children receive surgery, which is assumed to reduce the risk of recurrent UTIs. Prognosis is estimated as shown in *Figure 32*.
- For pathway LT2, patients are tested true positive for UTI (and are treated for the infection) but false negative for reflux and, as a result, undergo no treatment for reflux. A proportion of children (higher than in LT1 because the latter have surgery) will experience two recurrent UTIs and undergo MCUG, at which point they will undergo surgery, the treatment effect of which will be to reduce the risk of further UTIs. Again, prognosis is determined by the long-term model.
- For pathway LT3, children are true positive for UTI (and are treated for the infection) and for reflux, where the latter is mild/moderate in severity. These children are treated with prophylactic antibiotics, which are assumed to reduce the risk of recurrent UTIs. Prognosis in terms of PRS and renal disease is determined by the long-term model.
- For pathway LT4, children are true positive for UTI (and are treated for the infection) but false negative for mild/moderate reflux which, as a result, is not treated. A proportion of these patients will go on to experience two recurrent UTIs and, as a result, will undergo testing with MCUG and receive treatment with prophylactic antibiotics, the treatment effect of which is to reduce the risk of future UTIs. Again, the long-term model estimates the prognosis of these children.
- For pathway LT5, children are tested true positive for UTI (and are treated for the infection) but false positive for reflux. As they have had positive scans, these children are assumed to undergo MCUG, which indicates their false-positive status; hence, they do not receive inappropriate treatment. A proportion of these children will go on to experience two recurrent UTIs, at which time they will again be tested with MCUG; and they are assumed to be treated with prophylactic antibiotics, which will reduce the risk of further UTIs. The long-term model, again, determines their future prognosis.
- For pathway LT6, children are true positive for UTI (and are treated for the infection) and true negative for reflux. Hence, they receive no MCUG or treatment for reflux. Some children will go on to experience two recurrent UTIs, in which case they will be given MCUG and treatment with prophylactic antibiotics. The long-term model again determines their future prognosis.
- For pathway LT7, children are false positives for UTI (and are treated for the infection). This positive UTI test means that they are scanned

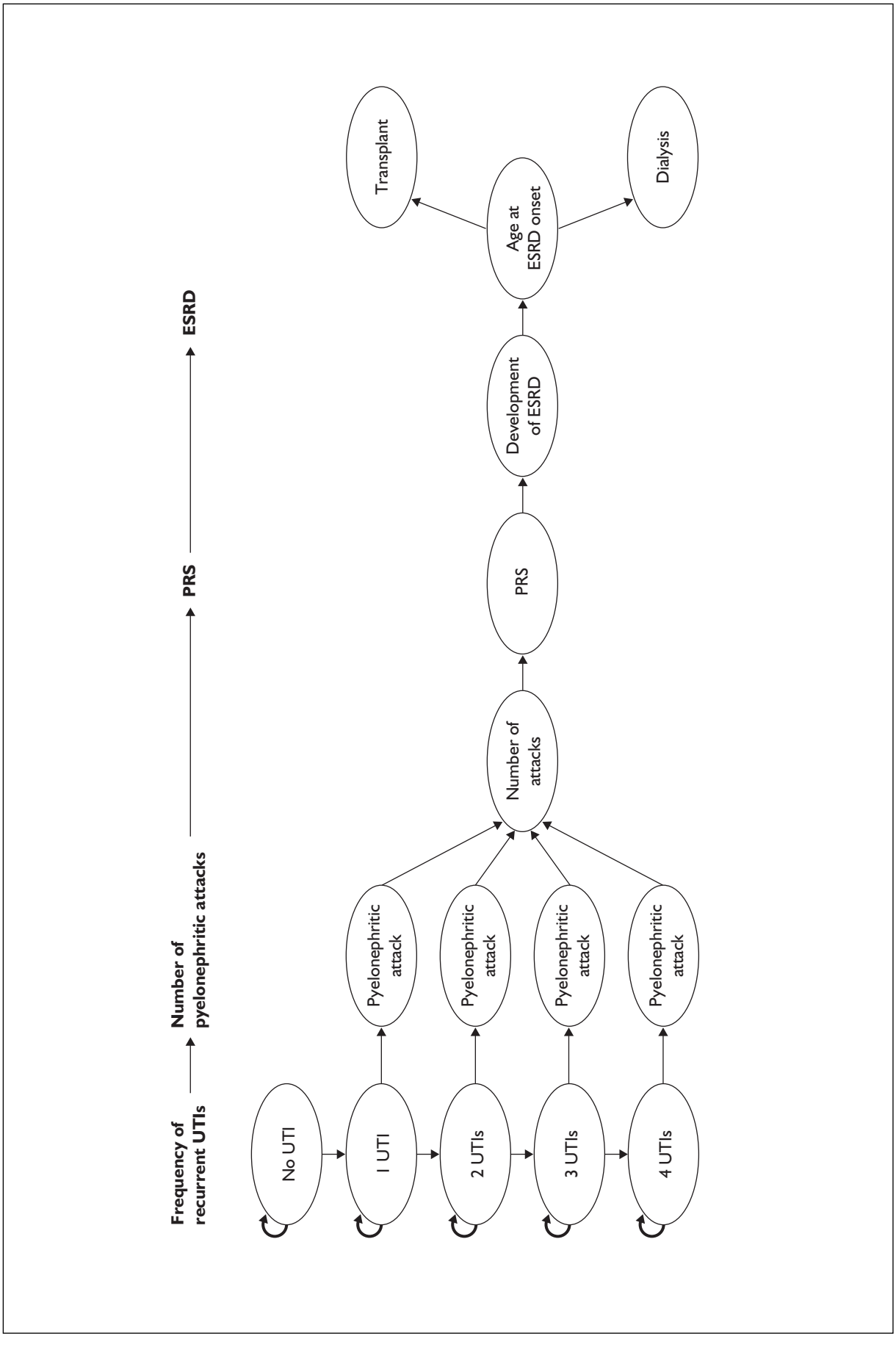

FIGURE 32 *Structure of the long-term model*

TABLE 31 *Prevalence data*

Initial diagnosis of UTI	Log odds	SE	Distribution	Mean	Source
Prevalence of UTI	–1.216	0.281	Normal	0.229	Meta-analysis of clinical data
Presence of reflux	–0.904	0.059	Normal	0.288	Meta-analysis of Downs[10]
Proportion of reflux classified as mild/moderate	1.961	0.175	Normal	0.877	
Recurrence of UTI	**Value**		**Distribution**		**Source**
Girls < 1 year	0.194		β (14,58)		Panaretto *et al*.[248]
Girls 1–2 years	0.166		β (5,25)		
Girls 2–3 years	0.2		β (4,16)		
Girls over 3 years	0.08		β (2,23)		
Boys < 1 year	0.144		β (18,107)		
Boys 1–2 years	0.125		β (2,14)		
Boys 2–3 years	0.133		β (2,13)		
Boys over 3 years	0.006		β (1,14)		

for reflux. They are false positives for reflux and as such are undergo MCUG. This MCUG reveals their true reflux status and as such they are not treated with prophylactic antibiotics.

- For pathway LT8, children are either false positives for UTI (and are treated for the infection) who test negative for reflux, or true negatives for UTI who do not go on to have subsequent imaging. Neither of these two groups goes on to have confirmatory MCUG or to receive prophylactic antibiotics.

Model inputs

The parameter inputs incorporated into the model are detailed in *Tables 31–36*.

Prevalence

Table 31 shows the prevalence (prior probability) estimates in the model. In the case of the prevalence of UTI, this relates to the prior probability of an infection in children presenting to primary care with symptoms suggesting UTI. This parameter is highly variable, depending on the clinical screening used by the physician. The preferred estimate in this study is to pool the prevalence results from all studies used to estimate the test characteristics of tests for the diagnosis of UTI that included an appropriate patient spectrum (see the section 'Accuracy of tests used to diagnose UTI', p. 29). This pooling has been based on a random effects meta-analysis on a log odds outcome scale.[249] The results of these were then transformed to the proportion scale for ease of interpretation. Based on this approach, a mean prevalence of UTI of 22.9% was estimated. The prevalence of reflux, in total and by severity, was derived from the studies identified in a previous systematic review reported in the analysis undertaken by Downs.[10] These studies were pooled using the same approach applied to estimate the prevalence of UTI. The results of the meta-analysis estimated that the prevalence of reflux in children with UTI was 28.8% (of which 87.7% were categorised as mild/moderate).

Test characteristics

Table 32 details the sensitivities and specificities of the tests for UTI considered in the model. These parameters are also pooled from those studies identified in the section 'Accuracy of tests used to diagnose UTI' (p. 29). Random effects meta-analysis is again used to pool these data, with estimates shown on the log-odds scale. Within the model, these parameters are incorporated as normal distributions, and the mean of this distribution is shown on the transformed scale. This can be interpreted as the pooled sensitivity or specificity.[249] *Table 33* shows test characteristics of imaging techniques. These data have been pooled as for the diagnostic tests, and can be interpreted on a similar basis. A series of separate regression analyses was undertaken to establish the relationship between the log odds of sensitivity and specificity for each individual diagnostic test. Cholesky decomposition was then used to allow for correlation when generating the random normal variates for sensitivity and specificity in the probabilistic simulation.[250]

Costs

The cost inputs used in the model are detailed in *Table 34*. It was not possible to identify the required costs from a single source and, as a result, the table shows the range of sources from which

TABLE 32 *Performance of diagnostic tests*

Type	Test name	Sensitivity				Specificity				Source
		Log odds	SE	Distribution	Mean	Log odds	SE	Distribution	Mean	
Dipstick	Nitrite	0.026	0.171	Normal	0.506	3.551	0.228	Normal	0.972	Meta-analysis
	LE	1.334	0.220	Normal	0.792	1.854	0.190	Normal	0.865	Meta-analysis
	Glucose	2.563	1.047	Normal	0.928	4.291	0.591	Normal	0.986	Meta-analysis
	Nitrite and LE	0.895	0.429	Normal	0.710	3.753	0.247	Normal	0.977	Meta-analysis
	Nitrite or LE	1.659	0.167	Normal	0.840	1.837	0.230	Normal	0.863	Meta-analysis
Microscopy	Pyuria	1.226	0.228	Normal	0.773	1.943	0.186	Normal	0.875	Meta-analysis
	Bacteriuria	1.654	0.153	Normal	0.839	2.794	0.298	Normal	0.942	Meta-analysis
	Pyuria and Bacteriuria	1.355	0.304	Normal	0.795	3.924	0.655	Normal	0.981	Meta-analysis
	Pyuria or Bacteriuria	2.493	0.304	Normal	0.924	1.190	0.342	Normal	0.767	Meta-analysis
Culture	Conventional	1.291	0.295	Normal	0.784	2.871	0.448	Normal	0.946	Meta-analysis
	Laboratory	NA	NA	Fixed	1	NA	NA	Fixed	1	

NA, not applicable.

TABLE 33 *Performance of imaging techniques*

Type	Sensitivity				Specificity				Source
	Log odds	SE	Distribution	Mean	Log odds	SE	Distribution	Mean	
Conventional ultrasound	–0.243	0.235	Normal	0.440	1.238	0.286	Normal	0.775	Meta-analysis
Contrast-enhanced ultrasound	1.627	0.237	Normal	0.836	2.604	0.178	Normal	0.931	Meta-analysis
MCUG	NA	NA	Fixed	1	NA	NA	Fixed	1	

TABLE 34 *Cost inputs into the model*

Name	Details	Value (£)	Distribution	Source
Diagnostic tests	Nitrite	0.13	Fixed	BNF[251]
	LE	0.13	Fixed	BNF[251]
	Glucose	0.13	Fixed	BNF[251]
	Nitrite/LE	0.15	Fixed	BNF[251]
	Pyuria	8	Fixed	Molyneux[252]
	Bacteriuria	8	Fixed	Molyneux[252]
	Pyuria/bacteriuria	16	Fixed	Molyneux[252]
	Dipslide culture	2.60	Fixed	Fenwick[246]
	Laboratory culture	2.60	Fixed	Fenwick[246]
Imaging	Conventional ultrasound	25.84	Fixed	York Hospital
	Contrast-enhanced ultrasound	124.05	Fixed	York Hospital
	MCUG	124.05	Fixed	York Hospital
Administration of tests	GP-administered tests (GP time)	6.77	Fixed	PSSRU[253]
	Hospital-based tests (outpatient visit)	86.00	Fixed	PSSRU[253]
Costs of treatment	Cost of low-dose long-term prophylaxis (per month)	2.43	Fixed	BNF[251]
	Cost of acute antibiotic treatment	6.58	Fixed	Fenwick[246]
	Additional cost of pyelonephritic treatment	17.256	Fixed	Claxton[244]
	Cost of UTI untreated	18	Fixed	PSSRU[253]
	Cost of pyelonephritic attack untreated	125	Fixed	Claxton[244]

PSSRU, Personal Social Services Research Unit.

TABLE 35 *Utility estimates used in the model*

Utility decrement of UTIs	Value	Duration	Distribution	Source
Utility decrement of treated UTI	0.001392	3 days	Fixed	Barry[245]
Utility decrement of treated pyelonephritic attack	0.010225	10 days	Fixed	Barry[245]
Utility decrement of untreated UTI	0.003248	7 days	Fixed	Barry[245]
Utility decrement of untreated pyelonephritic attack	0.014315	14 days	Fixed	Barry[245]

data were extracted. Drug costs were taken from the British National Formulary (BNF, issue 43).[251] Other sources included specific NHS hospitals, earlier published estimates in the area and national unit cost databases. All costs were adjusted to a 2003 price basis as appropriate.

Utilities

To reflect the implications of infections (lower UTI and pyelonephritis) when treated and untreated in the estimates of differential QALYs between the strategies, estimates of utility decrements associated with different types of infections are shown in *Table 35*. A study was not available that estimated the disutility of infections in children; therefore, these data are taken from a single source that looked at the cost-effectiveness of various treatment strategies for women with suspected UTIs.[245] Utilities were obtained from the Index of Well-Being, a multiattribute health scale that takes into account patient mobility, social activity and symptoms. The assumed durations of different attacks are shown in *Table 35*. These are also taken from Barry and colleagues.[245]

Long-term costs and effects from the long-term model

The long-term model is a development from one published earlier,[244] and the input parameters in the model are detailed in *Table 36*.

In the long-term model the frequency of recurrent UTI is modelled using a Markov process to establish the impact on quality of life and associated costs. The period for which a child is at risk of recurrent UTI was assumed to be 3 years.[10] A proportion of recurrent UTI episodes is assumed to be pyelonephritic attacks,[256] which have an additional (negative) impact on quality of life and costs. The probability of a UTI being pyelonephritic (by age and gender) and the probability of PRS given the cumulated number of pyelonephritic attacks (by VUR status) were based

TABLE 36 *Additional input parameters into the long-term model based on Claxton and colleagues*[244]

	Value	Distribution	Source
Recurrence of UTI			
Girls < 1 year	0.194	β (14,58)	Panaretto[248]
Girls 1–2 years	0.166	β (5,25)	Panaretto[248]
Girls 2–3 years	0.2	β (4,16)	Panaretto[248]
Girls > 3 years	0.08	β (2,23)	Panaretto[248]
Boys < 1 year	0.144	β (18,107)	Panaretto[248]
Boys 1–2 years	0.125	β (2,14)	Panaretto[248]
Boys 2–3 years	0.133	β (2,13)	Panaretto[248]
Boys > 3 years	0.06	β (1,14)	Panaretto[248]
Log RR of prophylaxis RR (transformed)	−1.0788 0.340003	Normal (SE 0.4384)	Smellie,[254] Stansfeld[255]
Probability of renal scarring (severe VUR)	No pyl = 0.27 1 pyl = 0.44 2 pyl = 0.83 3 pyl = 1 4 pyl = 1	β (2.18,5.88) β (10,10.94) β (4.68,0.92) Fixed Fixed	Jodal[256]
Probability of ESRD, given renal scarring	0.05	β (10.2,193.1)	Alexander[257]
Mean age at onset of ESRD	13.67 years	Triangular (7,24)	Mean from Arant;[258] range from Jacobson[259]
Cost of dialysis per year	£19,871	Fixed	Mowatt[260]
Cost of renal transplant	£6212.44	Fixed	CIPFA[261]
Mean survival duration on dialysis	12.25 years	Fixed	Mowatt[260]
Mean survival duration without ESRD	68.14 years	Uniform (66.65,69.62)	ONS[262]
Utility on dialysis	0.43	Normal (SE 0.26)	Claxton[244]
Utility following transplant	0.84	Normal (SE 0.24)	Claxton[244]
Proportion of pyelonephritic attacks	*P*(pyelonephritic|UTI) 1 year	0.83	β (110,22)[256]
	P(pyelonephritic|UTI) 2 years	0.51	β (66,63)[256]
	P(pyelonephritic|UTI) >3 years	0.51	β (66,63)[256]

CiPFA, Chartered Institute of Public Finance and Accountancy; ONS, Office for National Statistics; pyl, pyelonephritic attack; RR, relative risk.

on natural history evidence,[256] with beta distributions assigned reflecting the number of observations in each case. In addition, reflux status and the cumulative number of pyelonephritic attacks are important determinants of the risk of developing PRS.[10] Those children who develop PRS face a risk that this will lead to ESRD at some time in the future. The probability of a UTI being pyelonephritic (by age and gender) and the probability of PRS given the cumulated number of pyelonephritic attacks (by VUR status) were based on natural history evidence, with beta distributions assigned reflecting the number of observations in each case.[256] The probability of developing ESRD for children who experienced PRS was based on registry data.[257] The age at which children will develop ESRD is uncertain; therefore, this is represented by a distribution of ages (7–24 years) based on evidence reported in two observational studies.[258,259] Two consequences of ESRD are considered, transplant and long-term home dialysis. ESRD, whether managed by transplant or by long-term home dialysis, is associated with a reduction in quality-adjusted life expectancy as well as resource costs.

The benefit of identifying patients at their first UTI episode and correctly identifying the presence of reflux is, therefore, established through a series of links of evidence. Following a

correct diagnosis of VUR, patients are assumed to receive the appropriate management (either surgery or long-term antibiotics depending on the severity), resulting in a reduction in the frequency of recurrent UTIs. In turn, this is assumed to reduce the number of pyelonephritic attacks, which may also reduce the risk of PRS and the development of ESRD in later life. In this manner it is possible to estimate both the short-term and longer term impacts on quality-adjusted life expectancy and resource use of the alternative diagnostic and imaging strategies.

Data on the risk of recurrent urinary tract infection were identified from a previous systematic review of studies.[263] Although a number of separate sources was identified, these were largely based on studies undertaken during the 1970s. Furthermore, data on the rate of recurrence was not reported for a consistent period across these studies and separate estimates were not reported according to different patient characteristics (e.g. age and gender). However, one contemporary study was identified that reported the long-term follow-up of 290 children under 5 years of age with a first symptomatic UTI.[248] Separate recurrence rates were also reported by both age and gender. In general, rates of recurrence were reported to be highest in children under 1 year and lowest in children aged 3–5 years. Owing to the nature of the study, all children were managed according to standard clinical practice. Consequently, all children with reflux received prophylactic antibiotics and/or assessment for surgery. To determine the prognosis of those children in which reflux status was not identified following the first attack (i.e for those strategies in which imaging was not included and for false-negative results based on imaging for reflux), it was necessary to adjust the estimates reported in Panaretto.[248]

The assumed treatment effect of the appropriate management of reflux was based on evidence identified from a previous systematic review[264] for the effectiveness of long-term cotrimaxazole for the prevention of recurrent UTI. The treatment effect applies to all children diagnosed with recurrent UTI (second UTI for those diagnosed correctly at first UTI, and third UTI for those incorrectly diagnosed at first UTI) at any stage in the model. The treatment effect applies to all subsequent episodes of UTI. The reported treatment effect was then used to adjust the data reported in Panaretto[248] to obtain the likely recurrence rates in the absence of appropriate treatment for reflux. In this way the model uses RCT evidence for the effect of treatment on the frequency of recurrent UTI, combined with observational evidence, to estimate the longer term impacts on quality-adjusted life expectancy and resource use, for the alternative pathways outlined previously.

Given these inputs and the structure shown in *Figure 32*, the long-term costs and QALY decrements are shown in *Tables 44* (girls) and *45* (boys) of Appendix 9 for each of the pathways in the short-term model. In each case, pathway-specific costs and effects are shown for four subgroups of children: those aged less than 1 year, those aged 1–2 years, those aged 2–3 years and those over 3 years.

Strategies evaluated

Table 37 lists the 79 diagnostic strategies evaluated in the model. In principle, there are thousands of alternative strategies that could be compared, representing different tests, sequences and combinations relating to UTI and reflux. The guiding principle in selecting strategies for evaluation is those for which evidence on test performance exists. The large number of sequential tests has not been considered given the limited evidence reported on test combinations in the literature. In the absence of adequate data on the correlation between tests results it would represent an unrealistic assumption to assume that test results are independent. The only sequential strategy that is therefore considered in the strategies is where the second test/scan can reasonably be assumed to be definitive (e.g. laboratory culture for UTI or MCUG for reflux).

A range of possible strategies exists where, for tests with multiple elements such as a dipstick with nitrite and LE, those with indeterminate tests results (e.g. positive for nitrate and negative for LE) go on to receive subsequent tests. Estimating the number of indeterminates for all studies as in *Figure 13* would not have been possible for a probabilistic model, owing to the correlation between positive/positive tests and negative/negative results. Therefore, for a certain combination of positive/positive and negative/negative results the resulting number of indeterminate (negative/positive or positive/negative) test results may have been negative, which would not make clinical sense. Defaulting to a deterministic model, to allow the calculation of indeterminate test results from all studies, would not allow the full characterisation of uncertainty in model input parameters. For decision models in which there is a non-linear

TABLE 37 *List of strategies evaluated in the model*

Strategy no.	Strategy detail	Diagnostic tree
1	Treat none (no diagnostic)	Treat none
2	Treat all (no diagnostic)	Treat all
3	Treat all, followed by ultrasound	Treat all/image
4	Treat all, followed by enhanced ultrasound	Treat all/image
5	Treat all, followed by MCUG	Treat all/image
6	Nitrate	Single/no image
7	LE	Single/no image
8	Glucose	Single/no image
9	Nitrate and LE	Single/no image
10	Nitrate or LE	Single/no image
11	Pyuria	Single/no image
12	Bacteriuria	Single/no image
13	Pyuria and bacteriuria	Single/no image
14	Pyuria or bacteriuria	Single/no image
15	Culture	Single/no image
16	Laboratory culture	Single/no image
17	Nitrate, followed by ultrasound	Single diagnostic
18	LE, followed by ultrasound	Single diagnostic
19	Glucose, followed by ultrasound	Single diagnostic
20	Nitrate and LE, followed by ultrasound	Single diagnostic
21	Nitrate or LE, followed by ultrasound	Single diagnostic
22	Nitrate, followed by enhanced ultrasound	Single diagnostic
23	LE, followed by enhanced ultrasound	Single diagnostic
24	Glucose, followed by enhanced ultrasound	Single diagnostic
25	Nitate and LE, followed by enhanced ultrasound	Single diagnostic
26	Nitrate or LE, followed by enhanced ultrasound	Single diagnostic
27	Nitrate, followed by MCUG	Single diagnostic
28	LE, followed by MCUG	Single diagnostic
29	Glucose, followed by MCUG	Single diagnostic
30	Nitrate and LE, followed by MCUG	Single diagnostic
31	Nitrate or LE followed by MCUG	Single diagnostic
32	Pyuria, followed by ultrasound	Single diagnostic
33	Bacteriuria, followed by ultrasound	Single diagnostic
34	Pyuria and bacteriuria, followed by ultrasound	Single diagnostic
35	Pyuria or bacteriuria, followed by ultrasound	Single diagnostic
36	Culture, followed by ultrasound	Single diagnostic
37	Laboratory culture, followed by ultrasound	Single diagnostic
38	Pyuria, followed by enhanced ultrasound	Single diagnostic
39	Bacteriuria, followed by enhanced ultrasound	Single diagnostic
40	Pyuria and bacteriuria, followed by enhanced ultrasound	Single diagnostic
41	Pyuria or bacteriuria, followed by enhanced ultrasound	Single diagnostic
42	Culture, followed by enhanced ultrasound	Single diagnostic
43	Laboratory culture, followed by enhanced ultrasound	Single diagnostic
44	Pyuria, followed by MCUG	Single diagnostic
45	Bacteriuria, followed by MCUG	Single diagnostic
46	Pyuria and bacteriuria, followed by MCUG	Single diagnostic
47	Pyuria or bacteriuria, followed by MCUG	Single diagnostic
48	Culture, followed by MCUG	Single diagnostic
49	Laboratory culture, followed by MCUG	Single diagnostic
50	Nitrate/laboratory, followed by ultrasound	Multiple diagnostic[a]
51	LE/laboratory culture, followed by ultrasound	Multiple diagnostic
52	Glucose/laboratory, followed by ultrasound	Multiple diagnostic
53	Nitrate and LE/laboratory culture, followed by ultrasound	Multiple diagnostic
54	Nitrate or LE/laboratory culture, followed by ultrasound	Multiple diagnostic
55	Nitrate/laboratory culture, followed by enhanced ultrasound	Multiple diagnostic
56	LE/laboratory culture, followed by enhanced ultrasound	Multiple diagnostic
57	Glucose/laboratory culture, followed by enhanced ultrasound	Multiple diagnostic
58	Nitrate and LE/laboratory culture, followed by enhanced ultrasound	Multiple diagnostic

continued

TABLE 37 *List of strategies evaluated in the model (cont'd)*

59	Nitrate or LE/laboratory culture, followed by enhanced ultrasound	Multiple diagnostic
60	Nitrate/laboratory culture, followed by MCUG	Multiple diagnostic
61	LE/laboratory, followed by MCUG	Multiple diagnostic
62	Glucose/laboratory, followed by MCUG	Multiple diagnostic
63	Nitrate and LE/laboratory culture, followed by MCUG	Multiple diagnostic
64	Nitrate or LE/laboratory culture, followed by MCUG	Multiple diagnostic
65	Pyuria/laboratory culture, followed by ultrasound	Multiple diagnostic
66	Bacteriuria/laboratory culture, followed by ultrasound	Multiple diagnostic
67	Pyuria and bacteriuria/laboratory culture, followed by ultrasound	Multiple diagnostic
68	Pyuria or bacteriuria/laboratory culture, followed by ultrasound	Multiple diagnostic
69	Culture/laboratory culture, followed by ultrasound	Multiple diagnostic
70	Pyuria/laboratory culture, followed by enhanced ultrasound	Multiple diagnostic
71	Bacteriuria/laboratory culture, followed by enhanced ultrasound	Multiple diagnostic
72	Pyuria and bacteriuria/laboratory culture, followed by enhanced ultrasound	Multiple diagnostic
73	Pyuria or bacteriuria/laboratory culture, followed by enhanced ultrasound	Multiple diagnostic
74	Conventional/laboratory culture, followed by enhanced ultrasound	Multiple diagnostic
75	Pyuria/laboratory culture, followed by MCUG	Multiple diagnostic
76	Bacteriuria/laboratory culture, followed by MCUG	Multiple diagnostic
77	Pyuria and bacteriuria/laboratory culture, followed by MCUG	Multiple diagnostic
78	Pyuria or bacteriuria/laboratory culture, followed by MCUG	Multiple diagnostic
79	Culture/laboratory culture, followed by MCUG	Multiple diagnostic

[a] Multiple diagnostic refers to diagnostic test followed by a confirmatory laboratory culture for all positive results.

relationship between inputs and outputs (e.g. Markov models), probabilistic methods provide the only reliable method of estimating mean costs and outcomes.[31]

Incorporating the correlation between positive/positive and negative/negative tests in the probabilistic model would have restricted the analysis to the subset of studies for which the test performance was reported in sufficient detail to identify separately the indeterminate tests. To have assessed the strategies using data from only a subgroup of studies would have risked conflicting results for those strategies that have been evaluated based on all available strategies, making interpretation and conclusions difficult. Furthermore, none of the dipstick combinations separately specified which element of the test (e.g. nitrate or LE) was positive. Therefore, although these strategies are highly policy relevant, the evidence base was considered too limited to evaluate them.

The strategies differ in terms of which diagnostic elements they include, and this is illustrated with a series of specific diagnostic trees, presented as *Figures 34–37* in Appendix 9. Strategy 1 represents a 'boundary' scenario in that, although it is unlikely to be considered in routine practice, the cost-effectiveness of the strategy helps in the interpretation of other results in the analysis. With this strategy, no diagnostic tests are performed for UTI and no treatment is offered for UTI; no scanning for reflux is undertaken. Strategy 2 also represents a boundary strategy, where no tests are undertaken and every presenting child is treated for UTI, and no scanning is undertaken.

Strategies 3–5 represent strategies where no diagnostic tests are undertaken for UTI and all children are treated with short-term antibiotics, but all children go on for scanning for reflux with alternative tests. Strategies 6–16 represent variations on a strategy where a test for UTI is undertaken (with one of several technologies) and treatment offered to positive cases, but no scanning is undertaken for reflux.

Strategies 17–49 include tests for UTI, with positive cases treated with antibiotics and going on for a scan for reflux. The remainder of the strategies in *Table 37* are the same as for strategies 17–49, but all positive tests for UTI are then sent for a confirmatory laboratory-based culture.

Given that the studies used to derive the estimates of sensitivity and specificity for the glucose tests were of poor quality and dated (see the section 'Glucose', p. 35), and the test considered is not currently commercially available in the NHS, the main results are reported excluding glucose strategies. However, since the results for the glucose studies were reported to have the highest

accuracy in terms of both ruling in and ruling out disease, a separate sensitivity analysis was undertaken including glucose strategies. Owing to the practical limitations of performing this test in very young children (see the section 'Glucose', p. 35), the results are only presented in children aged 3 years and over.

Analytical methods

All models were implemented in Microsoft Excel 2000. Each strategy is evaluated in terms of its expected costs and decrements in QALYs. Expected costs include the costs of testing and treatment in the short-term model, as well as the costs of treatment and complications in the long-term model. Expected decrements in QALYs include the health effects of UTIs (differentiating between lower UTIs and pyelonephritic attacks, and between those that are treated and those that are not) and the health implications of long-term renal complications.

To rank the large number of strategies in terms of cost-effectiveness, the concept of net monetary benefit (NMB)[265] is used, rather than standard decision rules based around the incremental cost-effectiveness ratio. The NMB is simply a construct to place the expected costs and QALYs of a strategy on the same scale, and this is achieved by translating expected QALYs into a monetary value using a figure for the maximum willingness to pay for a QALY. If this maximum willingness to pay is termed λ, and ex QALY and ex cost are the expected values of QALYs and costs, respectively, from the model, then the expected NMB for strategy i is:

$$\text{Ex NMB}_i = [\text{Ex QALY}_i \times \lambda] - \text{Ex Cost}_i$$

Given that the maximum willingness to pay for an additional QALY is unknown, results are presented for a range of these values (£0–50,000 in £1000 intervals). On this basis, each of the 78 strategies is ranked in terms of their expected NMB.

To allow for the uncertainty associated with the parameter inputs in the model, probabilistic sensitivity analysis[240] is used. The distributions used in the probabilistic analysis are defined in the respective input tables (*Tables 31–36*). Second order Monte Carlo simulation is used to identify the proportion of simulations for which a given strategy is ranked first out of all those evaluated.

These results are conditional on the structure of the model and the data inputs used. Sensitivity analysis is also undertaken to rerun the analysis including glucose tests. This is because the evidence base relating to these tests is old and its quality is doubtful.

Results

Tables 38 and *39* provide full details of the ranking (in terms of expected NMB) of each of the strategies, excluding the glucose-based strategies for the base-case analysis.. In the base-case analysis separate results are provided for each of the eight sub-groups of children. In addition, a separate sensitivity analysis including glucose strategies is reported in Appendix 9 (*Tables 46* and *47*). The sensitivity analysis is constrained to children aged over 3 years. In addition, the tables presented in Appendix 10 show the results of the Monte Carlo simulation in terms of the probability that each of the strategies is ranked first.

Base-case analysis (excluding glucose strategies)

Table 38 provides the summary results for strategies that exclude a glucose test for UTI. The results are shown for each subgroup of girls by indicating the optimal strategy (the strategy ranked first in terms of expected NMB) and the probability of that strategy being ranked first. A number of strategies is identified as optimal over the range of willingness-to-pay thresholds. Strategy 2 (treating all children without any prior diagnostic test) is the most cost-effective at lower threshold values: until £7000, £10,000, £13,000 and £27,000 for the subgroups less than 1 year, 1–2, 2–3 and over 3 years, respectively. The next most optimal strategy is 30 (nitrite and LE, followed by MCUG) until thresholds of £17,000, £24,000, £29,000 and £49,000 for the four subgroups, respectively. Strategy 48 (laboratory culture, followed by MCUG) is the most cost-effective as the willingness-to-pay threshold becomes yet higher, although for relatively few threshold values: until £18,000, £27,000, £35,000 and £50,000 for the four subgroups, respectively. As the willingness-to-pay thresholds move up towards £50,000 per additional QALY (the maximum value considered here), strategy 64 (nitrite or LE, followed by laboratory culture for positive tests, followed by MCUG) becomes the most cost-effective strategy for girls aged 3 years or less. For all subgroups of girls, the probability of the strategy with the highest expected NMB being the optimal strategy varies widely, between 0.12 and 0.67.

Table 39 provides, on a similar basis, the summary results for each subgroup of boys. Strategy 2 (treating all children without any prior diagnostic test) is optimal until a willingness-to-pay threshold

TABLE 38 *Optimal strategy (and probability ranked first) for each subgroup of girls as a function of the threshold willingness to pay for an additional QALY, excluding glucose test*

Threshold value (£)	Girl < 1 year		Girl 1–2 years		Girl 2–3 years		Girl > 3 years	
	Optimum	Prob. ranked 1st	Optimum	Prob. ranked 1st	Optimum	Prob. ranked 1st	Optimum	Prob ranked 1st
0	2	0.677	2	0.682	2	0.712	2	0.665
1,000	2	0.757	2	0.759	2	0.784	2	0.739
2,000	2	0.813	2	0.818	2	0.83	2	0.801
3,000	2	0.807	2	0.862	2	0.858	2	0.857
4,000	2	0.755	2	0.874	2	0.864	2	0.892
5,000	2	0.679	2	0.843	2	0.838	2	0.919
6,000	2	0.594	2	0.794	2	0.8	2	0.918
7,000	2	0.522	2	0.753	2	0.744	2	0.922
8,000	30	0.299	2	0.697	2	0.691	2	0.908
9,000	30	0.313	2	0.619	2	0.638	2	0.901
10,000	30	0.315	2	0.575	2	0.588	2	0.889
11,000	30	0.324	30	0.257	2	0.548	2	0.873
12,000	30	0.323	30	0.27	2	0.505	2	0.857
13,000	30	0.326	30	0.278	2	0.462	2	0.841
14,000	30	0.323	30	0.292	30	0.222	2	0.826
15,000	30	0.316	30	0.306	30	0.222	2	0.804
16,000	30	0.305	30	0.31	30	0.227	2	0.791
17,000	30	0.294	30	0.302	30	0.222	2	0.776
18,000	48	0.188	30	0.295	30	0.219	2	0.758
19,000	64	0.157	30	0.303	30	0.215	2	0.745
20,000	64	0.169	30	0.296	30	0.213	2	0.722
21,000	64	0.172	30	0.297	30	0.218	2	0.709
22,000	64	0.173	30	0.287	30	0.215	2	0.696
23,000	64	0.177	30	0.275	30	0.21	2	0.686
24,000	64	0.178	30	0.277	30	0.212	2	0.674
25,000	64	0.185	48	0.189	30	0.205	2	0.661
26,000	64	0.189	48	0.196	30	0.208	2	0.648
27,000	64	0.197	48	0.202	30	0.205	2	0.633
28,000	64	0.199	64	0.137	30	0.2	30	0.167
29,000	64	0.202	64	0.14	30	0.193	30	0.161
30,000	64	0.207	64	0.144	48	0.14	30	0.167
31,000	64	0.204	64	0.147	48	0.143	30	0.165
32,000	64	0.209	64	0.149	48	0.143	30	0.167
33,000	64	0.21	64	0.153	48	0.143	30	0.171
34,000	64	0.211	64	0.153	48	0.141	30	0.174
35,000	64	0.214	64	0.159	48	0.141	30	0.171
36,000	64	0.213	64	0.163	64	0.137	30	0.17
37,000	64	0.214	64	0.168	64	0.139	30	0.168
38,000	64	0.217	64	0.164	64	0.142	30	0.166
39,000	64	0.218	64	0.166	64	0.147	30	0.165
40,000	64	0.214	64	0.17	64	0.148	30	0.16
41,000	64	0.213	64	0.17	64	0.148	30	0.167
42,000	64	0.209	64	0.174	64	0.151	30	0.167
43,000	64	0.208	64	0.178	64	0.152	30	0.169
44,000	64	0.207	64	0.18	64	0.15	30	0.173
45,000	64	0.208	64	0.183	64	0.149	30	0.169
46,000	64	0.204	64	0.18	64	0.147	30	0.169
47,000	64	0.196	64	0.175	64	0.148	30	0.166
48,000	64	0.196	64	0.179	64	0.152	30	0.161
49,000	64	0.193	64	0.176	64	0.154	30	0.163
50,000	64	0.19	64	0.176	64	0.152	48	0.12

Details of strategies are given in *Table 37*.
Prob, probability.

TABLE 39 *Optimal strategy (and probability ranked first) for each subgroup of boys as a function of the threshold willingness to pay for an additional QALY, excluding glucose tests*

Threshold value	Boy < 1 year		Boy 1–2 years		Boy 2–3 year		Boy > 3 years	
	Optimum	Prob. ranked 1st	Optimum	Prob. ranked 1st	Optimum	Prob. ranked 1st	Optimum	Prob. ranked 1st
0	2	0.683	2	0.7	2	0.684	2	0.685
1,000	2	0.759	2	0.765	2	0.743	2	0.748
2,000	2	0.809	2	0.814	2	0.802	2	0.809
3,000	2	0.867	2	0.849	2	0.838	2	0.853
4,000	2	0.903	2	0.882	2	0.878	2	0.885
5,000	2	0.9	2	0.893	2	0.902	2	0.907
6,000	2	0.891	2	0.9	2	0.921	2	0.932
7,000	2	0.872	2	0.89	2	0.93	2	0.936
8,000	2	0.831	2	0.874	2	0.928	2	0.936
9,000	2	0.774	2	0.848	2	0.928	2	0.927
10,000	2	0.723	2	0.82	2	0.921	2	0.914
11,000	2	0.681	2	0.776	2	0.912	2	0.907
12,000	2	0.636	2	0.754	2	0.898	2	0.902
13,000	2	0.599	2	0.718	2	0.885	2	0.893
14,000	2	0.562	2	0.69	2	0.873	2	0.875
15,000	30	0.253	2	0.65	2	0.854	2	0.861
16,000	30	0.27	2	0.622	2	0.83	2	0.843
17,000	30	0.263	2	0.595	2	0.809	2	0.83
18,000	30	0.274	2	0.575	2	0.79	2	0.821
19,000	30	0.283	30	0.213	2	0.776	2	0.804
20,000	30	0.274	30	0.22	2	0.755	2	0.79
21,000	30	0.279	30	0.223	2	0.736	2	0.778
22,000	30	0.279	30	0.225	2	0.71	2	0.768
23,000	30	0.281	30	0.23	2	0.698	2	0.759
24,000	30	0.282	30	0.231	2	0.689	2	0.751
25,000	30	0.279	30	0.229	2	0.672	2	0.745
26,000	30	0.271	30	0.228	2	0.662	2	0.737
27,000	30	0.268	30	0.223	2	0.648	2	0.728
28,000	30	0.261	30	0.227	2	0.639	2	0.721
29,000	30	0.263	30	0.229	2	0.628	2	0.715
30,000	30	0.258	30	0.229	2	0.618	2	0.709
31,000	30	0.253	30	0.224	2	0.609	2	0.704
32,000	48	0.179	30	0.225	2	0.599	2	0.694
33,000	48	0.186	30	0.219	2	0.588	2	0.683
34,000	48	0.189	30	0.209	2	0.582	2	0.678
35,000	48	0.192	30	0.206	30	0.155	2	0.668
36,000	48	0.19	30	0.203	30	0.164	2	0.663
37,000	64	0.125	48	0.185	30	0.164	2	0.656
38,000	64	0.13	48	0.185	30	0.169	2	0.65
39,000	64	0.133	48	0.185	30	0.17	2	0.648
40,000	64	0.136	48	0.187	30	0.172	2	0.643
41,000	64	0.139	48	0.187	30	0.173	2	0.63
42,000	64	0.147	48	0.187	30	0.173	2	0.62
43,000	64	0.149	48	0.189	30	0.173	30	0.111
44,000	64	0.158	48	0.187	30	0.175	30	0.109
45,000	64	0.159	48	0.188	30	0.174	30	0.108
46,000	64	0.163	48	0.191	30	0.176	30	0.108
47,000	64	0.165	64	0.141	30	0.177	30	0.106
48,000	64	0.168	64	0.141	30	0.174	30	0.104
49,000	64	0.171	64	0.144	30	0.171	30	0.108
50,000	64	0.173	64	0.146	30	0.172	30	0.11

Details of the strategies are given in *Table 37*.

of £14,000, £18,000, £34,000 and £42,000 per additional QALY for the subgroups less than 1 year, 1–2, 2–3 and over 3 years, respectively. Strategy 30 (nitrate and LE, followed by MCUG) becomes optimal at the next step up of the willingness-to-pay threshold: until £31,000 and £36,000 for the less than 1 year and 1–2 years subgroups, and all the way to £50,000 (the maximum threshold considered here) for the two oldest subgroups. For the two youngest subgroups, strategy 48 (laboratory culture, followed by MCUG) is optimal until maximum willingness-to-pay thresholds of £36,000 and £46,000 for the less than 1 year and 1–2 years subgroups, respectively. Strategy 64 (nitrite or LE, followed by laboratory culture for positive tests, followed by MCUG) is cost-effective for the two youngest subgroups from these thresholds to £50,000, the maximum considered here. For all subgroups of boys, the probability of the strategy with the highest expected NMB being the optimal strategy varies widely, between 0.11 and 0.68.

Sensitivity analysis (including glucose strategies in children aged over 3)

Table 46 in Appendix 9 provides a summary of the results for girls, using a similar approach to the last tables, but including glucose strategies. The table indicates that strategy 2 (treating all children without any prior diagnostic test) is the optimal strategy at lower levels of willingness to pay for a QALY. This is the optimal strategy until £24,000 per QALY for girls aged over 3 years. After these thresholds, strategy 29 (the use of a glucose test followed by MCUG) is optimal for all subgroups up to the maximum willingness-to-pay threshold considered (£50,000 per QALY). For all subgroups, the probability of the strategy with the highest expected NMB being the optimal strategy varies widely, between about 0.3 and 0.9 depending on how close the threshold is to the point at which the optimal strategy will change.

Table 47 in Appendix 9 shows the summary results for the subgroup of boys aged over 3 years. A similar pattern to that for girls emerges. As for the girls, the table indicates that strategy 2 (treating all children without any prior diagnostic test) is the optimal strategy at lower levels of willingness to pay for a QALY, but strategy 29 (the use of a glucose test followed by MCUG) is optimal for higher levels of willingness to pay. Compared with the results for girls, however, the willingness-to-pay threshold at which the optimal strategy shifts from 2 to 29 is higher for the boys in this age group. Strategy 2 is optimal until the threshold is £40,000 per QALY for boys aged over 3 years.

Chapter 7
Discussion

Systematic review methodology

Literature searches

Extensive literature searches were conducted in an attempt to locate all relevant studies that met inclusion criteria. These included electronic searches in a wide variety of databases, scanning the references of included studies, contacting experts in the field and handsearching. Diagnostic accuracy studies are very difficult to identify from electronic databases as there are no specific indexing terms for diagnostic accuracy studies.[266] Therefore, very sensitive searches were carried out to ensure that relevant studies were not missed. Attempts were also made to identify unpublished studies. These included contacting experts in the field and searching research registers, conference proceedings, grey literature and the Internet. It is unlikely that any relevant published studies have been missed, although it is possible that some unpublished studies were not identified.

The possibility of publication bias remains a potential problem in this review. However, although publication bias is likely to exist for studies of diagnostic accuracy, it may be less of a problem than for intervention studies. For intervention studies there is a clear cut-off defining whether an intervention works; that is, whether there is a significant difference in outcome between the treatment and control, and whether this difference favours the intervention. This is not the case for studies of diagnostic accuracy, where studies evaluate the agreement between the results of the index test and a reference standard. It is possible, and indeed likely, that studies that report higher estimates of test performance are more likely to be published, but the extent to which this occurs is unclear. There is evidence that publication bias is a particular problem for studies of small sample size, although these data are general and do not come from the diagnostic literature.[267,268] This review was restricted to studies that included at least 20 children, meaning that this type of publication bias is less likely to be a problem. The authors are unaware of any publications on publication bias in diagnostic tests or on methods to assess formally publication bias in a diagnostic systematic review.

Inclusion assessment and data extraction

Clear inclusion and exclusion criteria were set out in the protocol for this review. It is therefore explicit exactly which studies were eligible and which were not. A list of studies that appeared initially relevant but that did not meet all of the inclusion criteria for the review is provided in Appendix 7. Anyone who feels that a particular study may have been missed can therefore refer to this list to see whether it was identified and, if so, why it was not included in the review.

Data extraction was carried out using structured forms to ensure that information was extracted in a standardised manner. Detailed data extraction tables are provided in Appendix 5 and 6 to give readers further information on any studies in which they may be particularly interested. Inclusion assessment and data extraction were both performed by one reviewer and checked by a second. This helps to minimise errors in these processes and to ensure that they are carried out in an objective and reproducible manner.

Quality assessment

This review was used as an opportunity to pilot the QUADAS tool. This tool was developed as part of a previous HTA project[27] to be used in systematic reviews of diagnostic tests to assess the quality of diagnostic accuracy studies. The use of QUADAS in this review allowed the quality of the studies to be assessed using criteria developed though an evidence-based method.

Regression analysis was used to investigate the impact of components of study quality on diagnostic accuracy. This analysis was severely limited, however, by the completeness of reporting and the differences between the index tests evaluated and the reference standards used to confirm diagnoses in the primary studies. Few tests were evaluated by sufficient studies to allow meaningful use of meta-analytical pooling and investigation of heterogeneity. In addition, the analyses did not differentiate between the presence of a particular methodological bias and the absence of sufficient information to determine whether or not bias was present.

Studies of different diagnostic tests are likely to vary in their openness to the effects of different sources of methodological bias. This review included both tests with largely objective methods of interpretation (e.g. dipstick tests, microscopic cell counts) and tests with more subjective interpretations (e.g. imaging techniques). It might be expected that tests with more subjective interpretations would be more vulnerable to biases associated with inadequate blinding. This hypothesis was not supported by the results of the regression analysis. Disease progression bias would be unlikely to be a factor in studies of tests to diagnose UTI, where index test and reference standard are usually conducted on the same urine sample. Conversely, in the further investigation of UTI a clinically significant gap may occur between index test and reference standard (particularly in the case of imaging studies, where appointment scheduling may be difficult). Verification bias may therefore be expected to be more significant in the evaluation of imaging studies, as was observed in studies evaluating ultrasound for the detection of reflux. For tests to diagnose UTI, analyses showed an association between a number of variables potentially relating to quality of reporting and diagnostic accuracy (well-reported studies had higher DORs). One might expect this association to extend to diagnostic accuracy studies of all types of tests, but the data derived from this review are not adequate to demonstrate this.

Given the limitations described, the results of the regression analyses should be treated as hypothesis generating. Large data sets of well-reported primary studies are required to elucidate the influence of components of the methodological quality of primary studies on the results of any diagnostic meta-analysis. Without significant improvements in the reporting of primary studies, progress in this area will be limited. The components of quality assessment should always be reported, and their impact on summary outcome measures investigated, individually rather than as summary quality scores.

Synthesis

All measures of diagnostic accuracy rely on the assumption that diagnoses made using the reference standard are 100% correct. As only three studies of tests used to diagnose UTI reported reference standards other than culture, and culture is generally regarded as the reference standard of diagnosis in current practice, the analyses were conducted using culture as the reference standard. It remains a possibility, as with most reference standards, that diagnosis is not perfect and as such a better reference standard may be available. However, there is currently insufficient primary research available to shed light on this question. There is a general variation in the reference standards of diagnosis used by studies evaluating tests for the further investigation of UTI. Given the lack of consensus around appropriate reference standards in this area, the authors chose to consider all reported combinations of index tests and reference standards. This is a particular issue in tests used to detect renal scarring; four studies evaluated IVP (the historical reference standard) in comparison with ^{99m}Tc-DMSA renal scintigraphy (the currently excepted reference standard), and seven studies evaluated ^{99m}Tc-DMSA renal scintigraphy in comparison with IVP. The data from these studies indicate incomplete agreement between the two tests, but do not provide information on their relative contributions to the assessment of renal scarring. In this situation, the multivariable prediction modelling approach (described in the section 'Diagnosis of UTI', p. 139) may represent a more useful design for primary research than diagnostic accuracy studies.

Sensitivity, specificity and likelihood ratios were chosen to summarise estimates of test performance. Ranges in sensitivity and specificity were reported, as these were the measures most commonly reported by the primary studies and they can be used to produce ROC plots. ROC plots provide an easy-to-interpret visual summary of all the studies included in a review. They enable the reader quickly to assess the variability between studies, the accuracy of the test and whether there appears to be a threshold effect.

Likelihood ratios were chosen as the primary effect measure as these are the measure that physicians find easiest to interpret.[269] Pooled likelihood ratios and estimates of the pretest probability of disease were used to calculate estimates of the post-test probability of disease. These measures provide a simple illustration of how the results of a test change the probability of disease and help the reader to determine how useful a test is likely to be in practice. Where possible, estimates of the pretest probability of disease were taken from published data. When these data were not available, estimates from the studies included in this review were used. The main limitation of this approach was the considerable heterogeneity in pooled likelihood ratios; it is debatable whether it is appropriate to

pool these estimates. It is important that pooled estimates are interpreted with caution and that the heterogeneity between studies is considered when interpreting these results. A further problem with this analysis is that positive and negative likelihood ratios were pooled individually. These measures are likely to be correlated within an individual study and so ignoring this correlation may be problematic.[270]

A regression analysis was conducted to investigate possible explanations for the observed heterogeneity. This analysis was carried out according to standard methods for pooling studies of diagnostic accuracy using the summary ROC approach.[36] This method takes the DOR as the dependent variable. The DOR is used as a single indicator of test performance and shows how much more often a positive test result occurs in a person with the condition of interest than in one without the condition.[270] Using the DOR for further investigation of heterogeneity means that one can only assess whether the factors investigated are associated with the DOR and not with sensitivity and specificity, or with positive and negative likelihood ratios. Often factors that lead to an increase in sensitivity will lead to a decrease in specificity and vice versa. Factors that lead to such changes in sensitivity and specificity may have no effect on the DOR. Using the DOR for further investigation of heterogeneity may thus miss relevant clinical associations. Recently, a new method for pooling sensitivity and specificity has been developed. This method is known as the bivariate model.[270] It preserves the underlying two-dimensional nature of the data and produces direct pooled estimates of sensitivity and specificity, incorporating any correlation that may exist between these two measures. The model can be extended to include explanatory variables leading to separate effects on sensitivity and specificity. This method has two advantages over the standard methods used in this review: first, the pooled estimates of sensitivity and specificity take into account the correlation between these two measures; and secondly, the effect of possible sources of heterogeneity on both sensitivity and specificity can be investigated in a single model rather than just looking at the effect of these variables on a single measure of test performance, the DOR. These methods have only been very recently developed and were not available at the time the analysis for this review was conducted. If the analysis were to be repeated it would be interesting to compare the results obtained using the bivariate model to the results presented here.

Economic modelling (methodology and findings)

The cost-effectiveness modelling presented here has found that the optimal diagnostic strategy for children presenting with symptoms suggestive of UTI depends on two key factors. The first is the relevant subgroup of children concerned, in terms of gender and age. The second is the health service's maximum willingness to pay for an additional QALY. With respect to the first issue, the willingness-to-pay threshold at which the optimal strategy switches to the use of diagnostic tests for confirming UTI and imaging for reflux differs markedly across these subgroups. The differences reported between girls and boys can largely be accounted for by the higher probability of recurrent UTI in girls (and hence a higher number of accumulated pyelonephritic attacks which are linked to the development of renal scarring and ultimately ESRD). Similarly, since the probability of recurrent UTI appears higher in children aged under 3 years (and, in particular, for boys aged under 1 year), the use of diagnostic tests and imaging appears more cost-effective than in children aged 3–5 years.

With respect to the issue of the appropriate value for the health service's willingness to pay for a QALY, this threshold should inevitably depend on the opportunity cost of introducing a new service; that is, the benefits forgone by reducing and eliminating other services and programmes to free up the resources to fund a more expensive approach to UTI in children. This opportunity cost is likely to vary markedly in different settings, and providing a good estimate of its value will depend on acquiring more evidence on the costs and benefits of those services (across a range of specialities and sectors) that the healthcare system currently provides.

Given the difficulty in quantifying the opportunity cost of a new service, many decision-makers have adopted the rule of thumb that they are willing to pay for a new service, programme or intervention if its incremental cost per additional QALY is less than some willingness-to-pay threshold. In the UK, NICE has indicated that, although it has no fixed threshold, a threshold of around £30,000 per QALY gained is likely to be used in practice.[31]

The base-case analysis suggests that the most cost-effective strategy for girls would be a combined dipstick test, with a positive defined with either nitrite or LE, followed by laboratory culture for positive tests, followed by MCUG for reflux. This

conclusion would apply to the two youngest subgroups of girls (less than 1 year and 1–2 years). This strategy would be cost-effective in the subgroup aged 2–3 years as long as the health service was willing to pay up to £36,000 per QALY. For the oldest subgroup (girls aged over 3 years), a laboratory culture, followed by MCUG, would appear to be the most cost-effective.

For boys, the most cost-effective strategy, assuming a threshold willingness to pay of £30,000 per QALY, will vary by subgroup. In boys aged less than 1 year, a strategy of using a dipstick test and defining a positive as a positive nitrate and LE, followed by MCUG, would be cost-effective, although a laboratory culture followed by MCUG may also be considered. For boys aged 1–2 years, a positive nitrite and LE followed by MCUG would be the most cost-effective. For the two oldest subgroups, treating all children without any prior diagnostic test would appear to be the most cost-effective strategy.

A separate sensitivity analysis was undertaken to include a series of additional strategies based on glucose testing for the detection of UTI. This analysis was confined to children aged over 3 years owing to the practical difficulties associated with this test in younger children. The results from this analysis suggest that glucose testing may be cost-effective in girls in this older age group. In girls aged over 3 years, use of a glucose dipstick test followed by MCUG appears to be the most cost-effective strategy. The base-case results, for boys aged over 3 years, did not alter when the glucose strategies were included. Consequently, treating boys in this age group without any prior diagnostic test remained the most cost-effective strategy. Despite this finding in girls aged over 3 years, these results should be interpreted with some caution given the poor quality of the studies considered in this area. Furthermore, since the glucose test considered is not commercially available in the NHS, the strategies including this test may not be considered relevant alternatives. In the absence of a suitable estimate for the cost of a glucose test, the assumption was made that this would be similar to other dipstick tests. However, the provisional findings from both the effectiveness review and the economic analysis suggest that further research may be worthwhile with respect to this test.

The decision model presented here provides the first comprehensive cost-effectiveness analysis of alternative diagnostic strategies for UTI in children of which the reviewers are aware. The comprehensive nature of the analysis comes from the fact that the available evidence base has been synthesised and incorporated into the model as fully as possible. It also reflects the fact that, by mapping short-term outcomes to a prognostic model, it has been possible to express the cost-effectiveness of alternative diagnostic strategies in terms of QALYs. The use of this generic measure provides decision-makers with a basis of comparison with other uses of healthcare resources in a range of sectors, disease areas and specialities. The only other decision model assessing the cost-effectiveness of alternative diagnostic strategies for UTI in children was by Downs[10] and was fully reviewed in this document. However, it is difficult to compare Downs' results with those presented here. This is because the present review has considered a much fuller range of strategies, and expressed cost-effectiveness in terms of cost per QALY rather than using event-specific outcomes.

The main weakness of the modelling work is that it can only assess those diagnostic strategies for which evidence exists on test performance. In principle, there is a huge number of additional strategies, over and above those considered here, that may be used in practice. These include sequential strategies where one diagnostic test is used on all children and then another is used for positive or indeterminate cases. The only sequential test of this type that has been evaluated here is where children with a positive UTI test are sent for a confirmatory laboratory culture, which is assumed to be 100% sensitive and specific. For tests that cannot be assumed to be 100% accurate, it was not possible to model sequences as no evidence was identified on the correlations between test results; for example, it may be the case that a negative result on the first test in the sequence is likely to be followed by a negative test on the second test because their results are correlated. In the absence of evidence on correlations, it would have been unrealistic to model sequential strategies assuming independence in test results.

A second set of strategies for which the evidence base is too limited for their inclusion in the probabilistic model relates to indeterminate test results for those tests, such as nitrite and LE dipsticks and microscopy, producing two results. In practice, if one part of the test is positive and one negative, a clinician may choose to use another sequential test in those indeterminate cases. An additional weakness is that there are only limited data on correlations between test results. Furthermore, the performance and cost-

effectiveness of strategies involving further tests for children with indeterminate tests would depend on which part of the test gave the positive result (e.g. nitrite or LE in the dipstick) because, as this document has shown, the performance of these separate elements differs.

In the absence of adequate longitudinal data on the effects of imaging investigations for children with UTI on patient outcome, data for the long-term component of the economic model were assembled using a series of indirect evidence links based on the relationship reported between (1) reflux/recurrent infections and scarring, and (2) renal scarring and ESRD.[10] Although the parameterisation of the model reflects that there is significant uncertainty associated with this structural assumption, the lack of direct evidence concerning the existence of this link and hence the benefits of imaging have been brought into question.[29]

The existing evidence on the effectiveness of prophylactic antibiotics in preventing recurrent UTIs has also been reported to be weak,[264] owing to the small number and poor quality of studies in this area. In the absence of large, properly randomised, double-blinded trials to establish the efficacy of the long-term antibiotics, it is possible that the treatment effects reported in existing trials may have overestimated the true treatment effect.[264] Consequently, the model may have overestimated the benefit of imaging in terms of its impact on guiding subsequent treatments for VUR. However, it should be acknowledged that further investigation of children with UTI may also be used to detect other conditions that were not evaluated by any of the studies included in the review. For example, ultrasound may be used for other conditions, including hydronephrosis, abscess or calculus and duplex kidneys.[6] Similarly, contrast-enhanced ultrasound and MCUG may be used to detect urethral valves. Owing to the lack of data reported on these specific conditions, it was not possible to incorporate any associated benefits of the alternative imaging strategies in the management of conditions other than VUR. This could have implications for the both the absolute level of benefit and the relative benefits between the alternative imaging strategies considered in the model.

In the absence of appropriate utility data for children under 5, the review had to use data from women with suspected UTIs. This is not ideal as children may potentially have different utilities than women, and these may differ between boys and girls.

Results of the systematic review of diagnostic accuracy and clinical effectiveness

Diagnosis of UTI

An accurate and prompt diagnosis is essential to inform patient management decisions in young children with suspected UTI. The first step in the diagnostic process is to identify children presenting to the GP's surgery who may have a UTI. This will inevitably involve a clinical assessment. Only two of the studies included in this review looked at how good a clinical examination was at identifying children with possible UTI. One study found that a combination of age, race, temperature, presence of fever and absence of another source of fever was a good test for ruling out disease in children aged less than 2 years with fever of unknown source. A second study looked only at temperature and found that this was poor for both ruling in and ruling out disease. It is unsurprising that so few studies looked at the clinical identification of children who may have a UTI, as the design of any such study would present considerable problems. It is very difficult, if not impossible, to capture the signs and symptoms that a GP would use when deciding to test a child for a UTI. The difficulty in identifying children with possible UTI is supported by the wide range of prevalences of UTI in the studies included in this review. Even in those studies that included an appropriate spectrum of children (i.e. children suspected of having a UTI), prevalence of UTI ranged from 3 to 73%. This may be due to underlying differences in the true prevalence of UTI in these populations, for example due to age and gender differences. However, the criteria used to select children for testing for UTI are also likely to be an important source of variation. Further research may be useful to define from which children urine samples should be taken to test for UTI.

A small number of studies was included in the review that looked at clinical features for the diagnosis of UTI. The clinical tests investigated included urine cloudiness, urine odour and a combination of clinical symptoms. Urine cloudiness was a reasonable test for the presence of UTI, but all other clinical indices were found to be poor tests for UTI. As these studies only investigated a limited number of clinical features it is not possible to draw overall conclusions about

the accuracy of clinical information for the diagnosis of UTI. This is an area where further research may be useful.

In children with suspected UTI, the next step is to collect a suitable urine sample to test for the presence of infection. Different methods of urine sampling may be differently susceptible to contamination; this is associated with false-positive results, whatever method is used to test the urine. The issue of appropriate urine sampling techniques is of particular concern in young children, where the collection of a sterile, midstream sample can be problematic. Suprapubic aspiration has been regarded as the reference standard collection method. This procedure is invasive and may require the use of ultrasound guidance to ensure that the needle is inserted into the bladder. The identification of an alternative sampling method with acceptable diagnostic performance, which can readily be applied in the GP's surgery, and which is more acceptable to children and parents, is therefore desirable.

The studies on urine sampling included in this review looked at a number of different techniques and reported heterogeneous results. The only method for which a reasonable amount of data was available was CVU. Culture of CVU samples showed reasonably good agreement with culture of SPA samples, suggesting that this may be an appropriate routine method of urine collection. CVU samples are difficult to collect in young children who are not potty trained. Several alternative collection methods have been developed, including bag, pad and nappy specimens. This review identified only two studies of bag specimens. Although both studies reported excellent specificity, and sensitivity was also excellent in one study, sensitivity was very poor in the other. This study was published in 1976 and so the results may not be generalisable to today. One further study was identified that compared culture of a nappy/pad specimen to culture of an SPA sample. This study found excellent agreement between the two urine sampling techniques. Only limited data were available on these non-invasive methods of urine collection for young children. It is therefore difficult to draw overall conclusions regarding the appropriateness of their routine use. However, the limited data suggest that both bag and nappy/pad specimens may be suitable substitutes for SPA. Further work is needed to confirm this.

The main types of urine testing evaluated for the diagnosis of UTI were dipstick and microscopy. Culture is generally considered to be the reference standard for UTI diagnosis. The validity of culture as a reference standard was investigated in one study in which the results of culture were compared with a combined reference standard of culture and microscopy. This study reported excellent agreement, with 100% sensitivity and 93% specificity. The logistics of urine culture represent a significant drawback; culture takes approximately 48 hours to give a result, is generally performed in the laboratory and is more expensive than other methods. For this reason alternative, more rapid tests are needed to guide the prompt initiation of treatment. Dipsticks have the advantages of providing an immediate result, and of being both cheap and easy to perform. The studies of dipstick tests showed considerable heterogeneity and so the results should be interpreted with caution. The results suggest that a dipstick test that is positive for both LE and nitrite is good for ruling in disease, while one that is negative for both LE and nitrite is good for ruling out disease; that is, if a child has a dipstick positive for both nitrite and LE there is a very high likelihood that they have a UTI, whereas if they test negative for both nitrite and LE then the likelihood of having a UTI is small. A test combination that showed promise for both ruling in and ruling out disease was that of a dipstick test positive for nitrite, LE and protein. This test combination warrants further investigation.

An additional dipstick test that provided interesting results was that for urinary glucose, where a negative urinary glucose is regarded as a positive test for UTI. Only four studies of this test were identified, and all were conducted more than 30 years ago. All studies reported excellent specificity for this test. Sensitivity was also very high in three of the studies, but was lower, at 64%, in the fourth. This last study was conducted in children aged less than 1 year, suggesting that the test may be less useful in very young children. This difference in performance of the test with patient age may be explained by its apparent dependence on an overnight, fasting sample; such a sample would be impossible to obtain in children who are not toilet trained. However, given the limited results reported, this test appears to be potentially useful for the diagnosis of UTI in toilet-trained children. Further studies are needed to investigate these apparently promising results.

Although, in practice, microscopy and culture are generally requested in combination, microscopy has the advantage of being quicker to provide a result. Microscopy may also provide additional,

incidental information, such as bacterial type or presence of red cell casts (considered an indicator of renal involvement). It may be that microscopy has some potential as a test that could be performed in the GP's surgery. However, it remains more expensive than a dipstick test and requires some degree of expertise to perform. The studies of microscopy showed considerable heterogeneity, in terms of results (estimates of sensitivity, specificity and likelihood ratios), cut-off points, types of urine sample and population. A urine sample that was positive for both pyuria and bacteriuria on microscopy was found to be very good for ruling in disease. The pooled positive likelihood ratio was higher than that for the combination of dipstick positive for nitrite and LE. Similarly, a urine sample that was negative for both pyuria and bacteriuria on microscopy was found to be very good for ruling out disease. The pooled negative likelihood ratio was lower than that for dipstick negative for both nitrite and LE. It may therefore be inferred that the combination of microscopy for pyuria and bacteriuria represents a more accurate test for UTI than the dipstick. This is balanced by trade-offs in time, skill and cost requirements.

The cut-off points used to define a positive test for bacteriuria and pyuria varied between studies. The definition of 'pyuria' was either at least 5 or at least 10 WBC hpf^{-1}. As would be expected, studies using 5 WBC hpf^{-1} as the cut-off point tended to give higher estimates of sensitivity and lower estimates of specificity than those using 10 WBC hpf^{-1}. The definition of 'bacteriuria', although more subjective, also varied between studies, ranging from 'any bacteria' to 'few bacteria' to 'moderate bacteria'. The majority of studies defined bacteriuria as 'any bacteria'. Analogous to the situation for pyuria, studies using a smaller number of bacteria to define bacteriuria reported higher estimates of sensitivity and lower estimates of specificity. Regression analysis to investigate the observed heterogeneity in studies of microscopy for bacteriuria found that the use of Gram stain improved accuracy. The DOR was over five times greater in studies in which Gram staining was performed than in those where it was not. Regression analysis used to investigate the heterogeneity between studies of pyuria for the diagnosis of UTI found that sample centrifugation decreased accuracy. As these analyses used the DOR as the dependent variable it was impossible to determine effects on sensitivity and specificity independently. Investigation of between-study heterogeneity using regression analysis was not possible for the diagnostic accuracy of combinations of pyuria and bacteriuria as insufficient studies were available. The effects of sample processing factors were therefore not formally investigated in these studies. However, it seems likely that effects would be similar to those observed for bacteriuria and pyuria used as individual tests. Thus, on the basis of the results of this review, microscopy samples should be Gram stained, but not centrifuged, and should be examined for both pyuria and bacteriuria. A cut-off point of less than 5 WBC hpf^{-1} may be more appropriate for ruling out disease: if a child has a urine sample negative for pyuria using <5 WBC hpf^{-1} as the definition of a negative test then they are unlikely to have a UTI. A definition of a positive test for pyuria of greater than 10 WBC hpf^{-1} could be selected for ruling in disease, with the area of uncertainty between 5 and 10 WBC/hpf requiring confirmation by culture. A definition of 'any bacteria' in a Gram-stained sample should be used as the cut-off point to determine the presence of bacteriuria.

One further test for the diagnosis of UTI investigated in a number of studies was dipslide culture. This simplified method of culture was investigated mainly in community screening settings and for home monitoring of high-risk patients. The advantage of this method is that it can be performed outside the laboratory setting and is less labour intensive to perform and interpret. There was considerable heterogeneity in this group of studies. The pooled positive likelihood ratio was lower than those for dipstick and microscopy tests, while the pooled negative likelihood ratio was higher. Given the increased cost of dipslide culture over microscopy or nitrite and LE dipstick, and the longer time taken to give a result, this test appears to be of limited value in the context of general diagnosis of UTI in the GP's surgery. Other tests for the diagnosis of UTI were only investigated in a very limited number of studies and none appeared to offer any improvement over dipstick or microscopy.

Based on the results of this review, dipstick negative for LE and nitrite or microscopic analysis negative for pyuria and bacteriuria of a CVU, bag or nappy/pad specimen may reasonably be used to rule out UTI. These patients can then be excluded from further investigation, without the need for confirmatory culture. Similarly, combinations of positive tests could be used to rule in UTI and trigger further investigation. In the latter case, however, confirmation by culture may be preferred before the initiation of further, possibly invasive, investigations. Additional information on

antibiotic sensitivities, which can be provided by culture, may also be a significant consideration.

Further investigation of UTI

The main aim of further investigation of UTI is to prevent progressive renal damage and its consequences, hypertension, complications of pregnancy, renal insufficiency and end-stage renal failure. Renal parenchymal infection and scarring are established complications of infection of the upper urinary tract, and can lead to symptomatic renal disease. The secondary aim of further investigation of UTI is, therefore, to identify scarring or to identify children who may be at risk of developing further scarring. When considering tests used for the further investigation of UTI it is important to bear in mind the overall aim of investigation. If the information derived from tests cannot be used to prevent renal disease then there is no benefit in performing these tests. Tests should only be carried out if the results of the test will lead to a change in management of the child, and this change is likely to lead to an improved outcome. The ideal study to investigate whether a test leads to a change in outcome for a child would randomise children to different sequences of testing, or to no testing. The outcome, in terms of renal disease, would then be compared between the randomised groups. Unfortunately, only one such study was identified and this was only available as an abstract that reported limited information. All other studies included in this section of the review were diagnostic accuracy studies looking at how good tests were at detecting certain conditions. In the absence of direct evidence, information derived from diagnostic accuracy studies can be combined with information on how the different conditions detected are treated and the effect of these treatments on patient outcome, to infer conclusions about the overall benefits of further investigation of UTI.

A UTI can involve either the lower (cystitis) or the upper (APN) urinary tract. Cystitis does not involve the kidneys and so cannot itself lead to renal scarring. It is unclear whether children presenting with cystitis have an increased risk of future APN. Localisation of infection can be considered the first step in the further investigation of UTI. If APN can be ruled out then the child is not at immediate risk of scarring, and so further investigation is unlikely to be beneficial at this stage. Given that therapeutic delay is known to be associated with renal damage,[271] the possibility that children presenting with a first, lower UTI may benefit from monitoring for recurring infection remains open to question. A test to localise UTI would need to be non-invasive, inexpensive and quick to perform. Further investigation of children with cystitis could therefore be avoided, leading to cost saving and benefit to the child and parent, who would be spared the anxiety and discomfort associated with further testing. A variety of tests was investigated for the localisation of UTI. These included clinical tests, laboratory-based tests, ultrasound, MRI, CT, IVP, cystography and scintigraphy. Studies examining clinical and laboratory tests were diverse, and the techniques used were poorly described. Few useful data were available and, in general, the tests investigated showed poor accuracy for the localisation of UTI. Plasma CRP showed some limited potential as a test for ruling out APN. However, as the thresholds used and outcomes reported varied greatly, substantial further research would be required before routine use of this test could be recommended. Ultrasound showed no potential utility for the localisation of UTI. Available data were very limited for alternative imaging techniques such as MRI and CT. Acute ^{99m}Tc-DMSA remains the reference standard test for the localisation of UTI, with other scintigraphic techniques providing comparable performance. This is an invasive technique involving ionising radiation and so restriction of its use is desirable. The median prevalence of APN in those studies included in this section that were conducted in an appropriate patient spectrum (those with confirmed UTI) was 60%. If this is representative of the prevalence of APN in all children with UTI then localisation of UTI could have an important role in the further investigation of children with UTI: further tests could be avoided in the 40% of children with confirmed UTI who have cystitis. Further research to identify an accurate non-invasive test for the localisation of UTI is therefore justified. The remaining studies of the further evaluation of UTI were divided into four categories: detection of reflux, prediction of scarring, detection of scarring and studies with multiple aims.

The detection of reflux has been considered an important part of the further investigation of UTI as, historically, it has been thought to lead to an increased risk of scarring. This theory is currently the subject of considerable debate. The only study on the effectiveness of imaging compared routine and selective imaging (combination of ultrasound and MCUG) for the detection of reflux.[29] This study found increased rates of reflux detection and prophylaxis with routine imaging, but no reduction in scarring or recurrent UTIs. Four

diagnostic accuracy studies that looked at the association between reflux and the presence of renal scarring were included in this review. These studies found that the presence of reflux correlated poorly with the presence of scarring: some children without reflux were found to have scarring and some with reflux were found not to have scarring. A recent systematic review found that reflux is a weak predictor for renal damage in children hospitalised with UTI.[272] The management of reflux and how this impacts on a patient's future risk of renal disease is also the subject of debate. A systematic review[273] comparing surgical to medical management of reflux found no difference in outcome, deterioration in DMSA findings, between the two treatment groups. Reflux has also been shown to disappear spontaneously. A study of children diagnosed with reflux following a first UTI, which included a 15-year follow-up, found that reflux disappeared or reduced to grade I in 73% of children with dilator reflux.[274] MCUG is an invasive and costly test, involving considerable exposure to ionising radiation. Given the considerable doubts surrounding the link between reflux and renal scarring, and the benefits to be derived from treating reflux, it is difficult to justify the routine use of MCUG (the reference standard test for reflux) in children with UTI. The studies that looked at the diagnostic accuracy of alternative tests for the detection of reflux showed considerable heterogeneity. Standard ultrasound, IVP, indirect voiding radionuclide cystography, NAG/creatinine ratio, scintigraphy and a risk scoring system were all found to be relatively poor tests for the detection of reflux. The only imaging modality to show good accuracy for the detection of reflux was contrast-enhanced ultrasound. This is currently a little-used technique, but the results of the studies included in this section showed that it may have potential accuracy in the diagnosis of reflux. It carries an advantage over MCUG in that it does not involve exposure to ionising radiation and standard ultrasound can be performed at the same time. However, contrast ultrasound remains an invasive procedure requiring catheterisation.

Very few studies looked at tests for the prediction of renal scarring. These studies looked at tests that can be performed during the acute phase of infection and compared them with the presence of scarring on follow-up ^{99m}Tc-DMSA renal scintigraphy several months later. A test that could predict whether a child was at risk of renal scarring would be useful if a treatment were available that could prevent that child from developing scarring. Antimicrobial therapy is often initiated in children with UTI before further investigation, and treatment delay is the only therapeutic factor thought to affect the development of scarring.[3,271] The prediction of the development of renal scarring as a result of a current infection would, therefore, appear to be of academic interest alone. The tests investigated for the prediction of scarring were ultrasound, IVP, and presence of fever and elevated CRP levels. Available data were very limited and none of the tests showed good accuracy for the prediction of scarring.

The presence of renal scarring is considered to be the most important predictor of renal disease where UTI is the causative factor; however, not all children with scarred kidneys will have progressive scarring ending in renal failure.[271,275] There is currently very little that can be done to treat children with renal scarring to prevent complications. If repeat infection is assumed to be the cause of progressive scarring then prophylactic antibiotics may be initiated. However, a systematic review on the effectiveness of long-term antibiotics in preventing UTI found no evidence to support their use, as very few trials met inclusion criteria and those that did were small and of poor quality.[276] The review identified no data specific to children with reflux. Renal scintigraphy, generally using ^{99m}Tc-DMSA, is the accepted reference standard for the detection of renal scarring. It may be used acutely to localise UTI and hence to determine whether there is a risk of scarring from the current infection, and in sequential follow-up to monitor progressive scarring. The diagnostic accuracy of a number of other tests for the detection of scarring has also been investigated. Alternative static and dynamic renal scintigraphic techniques, including ^{99m}Tc-DTPA and ^{99m}Tc-MAG3, have been investigated. These were found to be good tests for the detection of scarring, correlating well with ^{99m}Tc-DMSA. Ultrasound was found to be reasonable for ruling in scarring, but poor for ruling out scarring. This would fit with anecdotal opinion that ultrasound is good at identifying gross scarring, but poor at detecting minor lesions. The studies included in the review did not look at how good ultrasound was at picking up different grades of scarring. It may be that ultrasound only detects more severe scarring, which may be of more clinical importance than scarring of any grade. If this is the case then ultrasound may provide a useful test both for ruling in and ruling out clinically important levels of scarring. Ultrasound carries benefits over ^{99m}Tc-DMSA in that it is non-

invasive, involves no ionising radiation and is easier and cheaper to perform. Further research evaluating the ability of ultrasound to detect different grades of scarring, and scarring progression, may therefore prove useful. Other tests investigated include cystography, IVP, MRI and a combination of ultrasound and MCUG. These were only investigated in a limited number of studies. Cystography and IVP were found to be relatively poor tests for the detection of scarring. MRI and the combination of ultrasound and MCUG showed reasonable diagnostic performance. These were each investigated in only one study and so further research into these tests may be useful if there is a benefit to detecting scarring.

Other studies included in the review looked at a variety of test combinations with multiple aims. These generally showed poor test performance and were too heterogeneous to draw general conclusions.

The further investigation of children with UTI may, in practice, be used to detect other conditions that were not evaluated by any of the studies included in the review. For example, ultrasound may be used to provide a structural overview of renal anatomy, and may also be used to rule out hydronephrosis, abscess or calculus,[6] or to detect malformations such as duplex kidneys.[6] Contrast-enhanced ultrasound and MCUG may be used to detect urethral valves. These can cause severe problems, especially in baby boys, and it is important that they are detected early as surgical treatment is available. However, these are a rare condition and are usually detected on antenatal ultrasound.

Chapter 8
Conclusions

The results of the systematic review of effectiveness (objectives 1–3) were used to inform the consideration of preliminary algorithms for the diagnosis and further investigation of UTI in children under 5 years of age. These are described below and represent the conclusions of the systematic review component of this project in terms of practice. Implications for research are discussed in the final section. The results of economic modelling did not directly inform the diagnostic algorithm.

Diagnosis of UTI

The development of the preliminary algorithm presented was informed by currently available evidence on the accuracy of tests used to diagnose UTI. There are several issues in this algorithm where different options could be considered and further research may be useful. These are discussed below. The algorithm is illustrated in *Figure 33*.

The starting point for the algorithm is a child in whom a GP has a suspicion of UTI based on clinical findings. As the studies included in the review provided very few data on the accuracy of clinical investigations for the diagnosis of UTI, criteria for clinical suspicion of UTI are not further defined. Further research is required to determine which clinical signs and symptoms should inform the decision to test for UTI.

The first step in confirming a suspected UTI is to obtain an uncontaminated urine sample on which further tests can be performed. The majority of studies included in the review found that CVU samples had similar accuracy to SPA samples when cultured. As CVU is a non-invasive collection method that can be used in the GP's surgery, this was the method chosen for the algorithm. Pad, nappy or bag specimens may be appropriate methods for obtaining a urine sample in non-toilet-trained children. There were very few accuracy data using an appropriate reference standard for this technique. It is therefore not included in the algorithm at this stage. Further research on sampling methods is required.

Although the glucose test was reported to have the highest accuracy in terms of both ruling in and ruling out disease, it is not included in the algorithm. This is due to the limitations of the studies of this test, and the practicalities of performing it in very young children, highlighted earlier in the report and in the discussion (Chapter 7).

Dipstick testing is the first diagnostic stage in the algorithm. These tests are easy to perform in the GP's surgery, give an immediate result and are relatively cheap. The systematic review showed that a dipstick for LE and nitrite, where both test results are interpreted in combination, was a good test both for ruling in (both positive) and ruling out a UTI (both negative). A dipstick positive for either LE or nitrite and negative for the other provides inconclusive diagnostic information and further testing is therefore required in these patients. Microscopy was included in the algorithm as the next option. Microscopy is more time consuming and expensive to perform than a dipstick test, but potentially quicker and cheaper than culture. As with dipstick tests, a combination of microscopy for pyuria and bacteriuria can be used accurately to rule in and rule out a UTI. An indeterminate test result is again obtained if microscopy is positive for either pyuria or bacteriuria, and negative for the other. Confirmatory culture is required in these patients. In patients considered to have a UTI, further culture to determine antibiotic sensitivities may be an option to inform treatment decisions.

The economics team (LB, SP) advised that it was not possible appropriately to model an indeterminate test result (as described above) with the available data. The economic analyses separately modelled two alternative strategies of 'LE and nitrite positive' to define a positive result, and 'LE or nitrite positive' to define a positive result. For the 'LE or nitrite positive' strategy indeterminate results were treated as positive. A similar approach was used to model microscopy tests. In practice, an indeterminate result may be treated differently (as proposed in the preliminary algorithm). Since the economic modelling was not able to assess this particular strategy it did not directly inform the development of the algorithm.

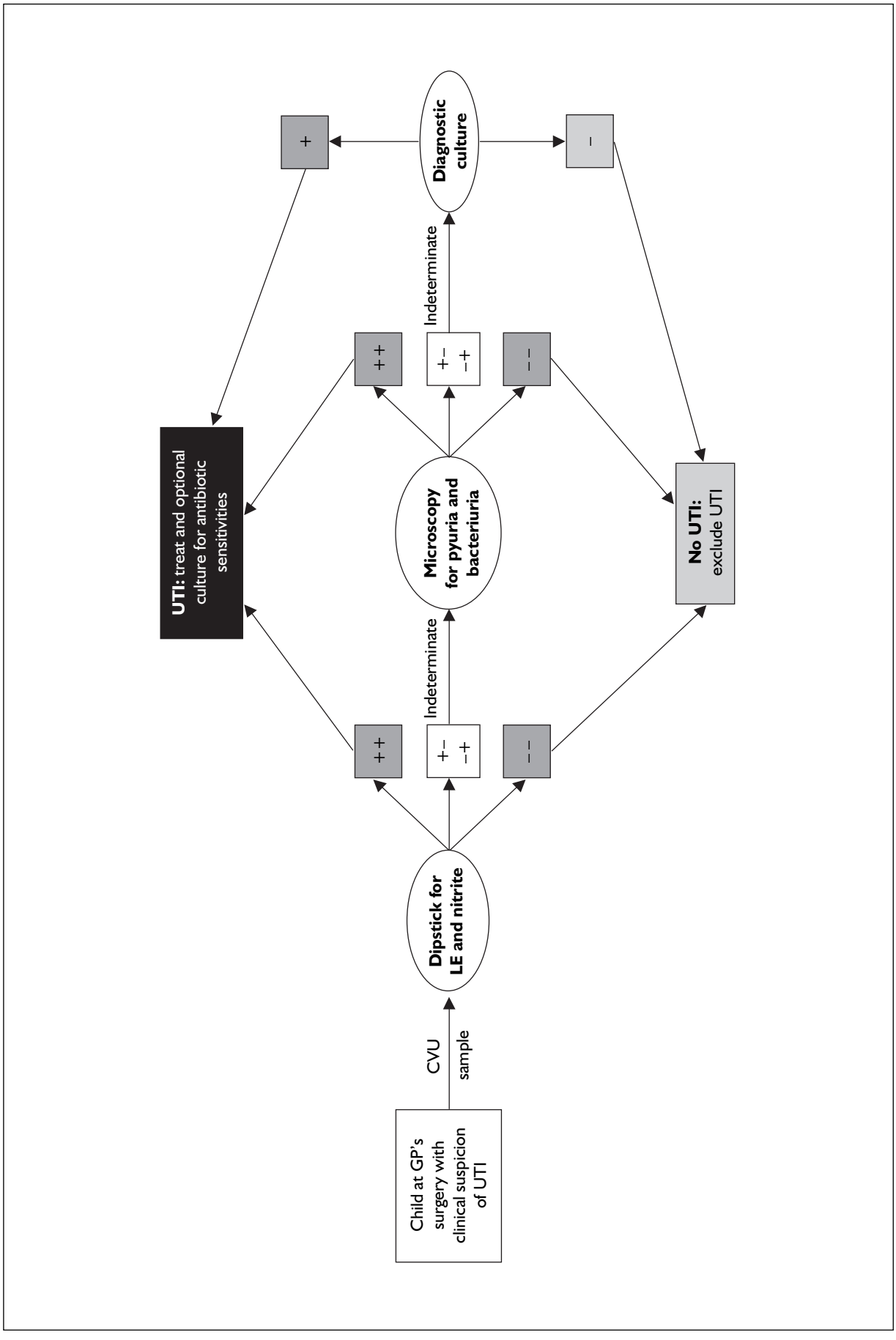

FIGURE 33 *Preliminary algorithm for the diagnosis of UTI*

There are four issues with regard to the diagnostic tests recommended in the algorithm:

- Should patients with an indeterminate dipstick test result be treated differently from those with a positive dipstick test result? Another option would be to recommend culture in all children testing positive on either LE or nitrite dipstick.
- Should microscopy be included as a separate step, or should children with an indeterminate test result receive culture?
- Does the accuracy of microscopy differ in children with an indeterminate dipstick test result? The studies of microscopy that contributed to the algorithm were carried out in children with a suspicion of UTI, not in those who had already been tested with a dipstick.
- Should all children receive confirmatory culture regardless of previous test results?

Future research may inform these areas.

Further investigation of UTI

No algorithm for the further investigation of UTI is presented, as the long-term benefits of imaging, in terms of patient outcome, remain unclear. This review has provided data on the accuracy of various different imaging techniques for the localisation of UTI, detection of reflux, and detection and prediction of scarring. Only one study was found that assessed the effect of imaging evaluations on patient outcome. In the absence of sufficient direct data on the effects of imaging investigations for children with UTI on patient outcome, data are required on the effectiveness of interventions for reflux and the prevention of scarring. The evidence on this is conflicting and a full systematic review of these data was not performed. Therefore, conclusions cannot be drawn on the appropriateness of further investigations in children with UTI. Given the lack of data on the clinical effectiveness of further investigation, data provided by economic modelling remain hypothetical and conditional upon the establishment of clinical effectiveness. The findings of cost-effectiveness modelling did not therefore influence the decision not to present an algorithm for the further investigation of UTI.

The results of the only study on the effectiveness of imaging do not support routine imaging for children with an initial UTI. However, this study only included children aged 2–10 years. No conclusions can therefore be drawn about children under this age; with no firm evidence base, readers will need to consider their own practice in the light of expert opinion. The limited evidence available supports monitoring of all children aged 2–5 years experiencing an initial UTI, with further investigation only if they experience a second UTI.

A test for the localisation of UTI was considered as an initial step in the investigation of these children. This would allow the exclusion of all children with a lower UTI from further investigation. Based on current evidence the only accurate test for this purpose is a DMSA scan. These scans are costly and invasive, and incur a radiation load. Therefore, the use of these scans in all children cannot be justified. If future studies identify a non-invasive test (e.g. a biochemical test) that can accurately localise a UTI, then this may be a valuable first step in the further investigation of UTI.

Ultrasound is currently used as the initial investigation for children with confirmed UTI. This is because it is non-invasive and relatively cheap, and is felt by practitioners to provide information on the presence of anatomical abnormalities that may be the underlying cause of infection. The review did not identify any studies on the accuracy of ultrasound for this purpose. However, this is a potentially important reason for performing ultrasound. Early identification of underlying abnormalities will inform further intervention. Further research is required regarding the accuracy of ultrasound in diagnosing these abnormalities, and its impact on patient outcome.

There is insufficient evidence to recommend any further routine investigation. The main two options for further investigation would be a follow-up DMSA scan or MCUG. The aim of DMSA is to identify the presence of scarring, with the objective of preventing future scarring by preventing further UTI. MCUG aims to detect reflux, which can be treated either surgically or with prophylactic antibiotics. In theory, the presence of reflux is associated with renal scarring and so by identifying and treating reflux further renal scarring can be prevented. However, as highlighted in the discussion there is debate regarding the association between reflux and scarring and the effectiveness of long-term antibiotics in reducing scarring. In the absence of evidence of any effect on patient outcome, the reviewers do not feel that universal DMSA or MCUG can be justified. The decision on whether or not to perform these examinations should

therefore be made on an individual patient basis. Further research regarding the effects of these imaging techniques on long-term patient outcome (development of renal scarring and end-stage renal disease) is urgently required.

Implications for clinical practice

The primary aim of this project was to determine the most effective strategy for the diagnosis and further investigation of UTI in young children such as to reduce repeat infection, renal scarring and its long-term consequences.

It should be particularly noted that, although the original focus of this study was children under 5 years, studies whose populations also include children aged up to 18 years have been included. This approach was adopted owing to the extreme paucity of data specific to children under 5 years. It remains the case that children under 5 years are a population of particular interest. Variation in test performance and the clinical effectiveness of testing is also likely in further subgroups, such as pre-toilet-trained children or neonates. However, at present insufficient data are available that are specific to these groups.

Five main objectives were identified for the project, all of which were only partially realised owing to a lack of relevant data. The majority of studies identified addressed the accuracy of tests for the diagnosis or further investigation of UTI. The main points for practice are as follows.

- Data identified were not adequate to characterise initial signs and symptoms useful in deciding which patients should be tested for UTI.
- Data suggest that a CVU sample is adequate for the diagnosis of UTI.
- The evidence concerning alternative methods of urine collection in pre-toilet-trained children, for example nappy or pad, is limited; clinicians may wish to consider using these methods given the invasive nature of SPA.
- Current data support the usefulness of combined LE and nitrite dipstick testing for rapid diagnosis of UTI.
- Clinicians may wish to consider follow-up microscopy and/or culture in the context of additional information that may be provided, for example antibiotic sensitivities, and its effect or otherwise on treatment decisions/outcomes.

Although a large volume of data was identified on the accuracy of tests used in the further investigation of UTI, the clinical relevance of such studies is questionable. The link between the diagnostic targets of this group of tests (e.g. VUR) and renal scarring and its long-term consequences remains open to question. In addition, there is a lack of evidence concerning the effectiveness of interventions, consequent upon test results, in reducing renal scarring and its consequences. The main points for practice are as follows.

- The review did not identify any studies on the accuracy of ultrasound for determining the presence of anatomical abnormalities that may be the underlying cause of infection. However, as this is a widely used, non-invasive and relatively cheap examination, practitioners may wish to continue its use in this context.
- The decision on whether or not to perform invasive imaging examinations should be made on an individual patient basis, giving careful consideration to the potential for effective use of information obtained.

The review identified only one study evaluating the effectiveness of further investigation of UTI:

- The results of this study do not support routine imaging for children aged 2–10 years with an initial UTI.
- The limited evidence available supports monitoring of all children aged 2–5 years experiencing an initial UTI, with further investigation only if they experience a second UTI.
- No evidence is available concerning the effectiveness of further investigation of UTI in children under 2 years; with no firm evidence base, readers will need to consider their own practice in the light of expert opinion.

Implications for research

General points

The quality assessment highlighted several areas that could be improved upon in future diagnostic accuracy studies. One of the most important features highlighted by the quality assessment was the failure of studies to include an appropriate patient spectrum. Future studies of tests for the diagnosis of UTI should be prospective evaluations of children presenting with possible symptoms of UTI. Studies on tests for the further investigation of UTI should include children with a confirmed UTI. Other areas that could be improved upon relate to reporting. Future studies should follow the Standards for Reporting of Diagnostic

Accuracy (STARD) guidelines for reporting of diagnostic accuracy studies,[277] and in particular should provide clear details on the following:

- how children are selected for inclusion in the study
- whether appropriate clinical information was available to the person interpreting the index test results
- whether the person interpreting the index test was blind to the results of the reference standard, and vice versa
- the flow of patients through the study: this should state how many patients were eligible for inclusion, how many met inclusion criteria but did not take part, and whether any children did not receive any of the index tests or the reference standard
- whether there were any uninterpretable results, and if so how these were handled in the analysis.

Future studies should consider relevant population subgroups such as neonates and pre-toilet-trained children.

Primary research in this field is expanding at a rapid rate, particularly with respect to imaging techniques for the further investigation of UTI. It was with this consideration in mind that update searches were conducted close to completion of the project (May 2004). Although no further studies satisfying the inclusion criteria of this review were identified, 21 publications were retrieved for consideration. On this basis it might be considered that regular update of this review would be necessary to maintain its relevance. However, given that substantial questions remain regarding the clinical effectiveness of further investigation of UTI, further primary or secondary diagnostic accuracy research in this area is unlikely to be immediately informative. New developments in the diagnosis of UTI (e.g. flow-cytometry techniques), for which no accuracy data satisfying the inclusion criteria of this review were identified, may prove a useful subject for future secondary research.

The review highlighted the following specific areas requiring further research.

Diagnosis of UTI

Information is required on the following.

- Are bag and nappy/pad samples appropriate methods of urine sampling? Current evidence on the diagnostic performance of samples collected using these techniques is sparse. Given that they represent the most convenient and practical methods of sample collection in non-toilet-trained children, well-designed diagnostic cohort studies to evaluate their performance should be conducted urgently.
- What is the accuracy of the combination of dipstick positive for nitrite, LE and protein?
- Should an indeterminate nitrite and LE dipstick test result be treated differently from a positive result? It is particularly important to consider the appropriate treatment and investigation pathways in children for whom initial test results are equivocal.
- Should microscopy be included in the diagnosis of UTI?
- Should all children receive confirmatory culture regardless of other test results?
- Should all children diagnosed with a UTI have a culture to determine antibiotic sensitivity?

The final two points should be considered in the context of the effect of additional information derived upon treatment pathways and outcomes (e.g. does early determination of antibiotic sensitivities result in significantly more effective treatment?) The validity of culture as a reference standard of diagnosis may also be an area of interest for future primary research.

Further research in the following areas may provide addition useful information.

- What clinical signs and symptoms should be used to select children to undergo testing for UTI? Although quantifying aspects of clinical experience is a difficult area, clear definitions of the signs and symptoms used in the examination and evaluation of their diagnostic utility in well-designed diagnostic cohort studies make up a potentially informative process. Such studies could increase consistency in the diagnostic process and help to avoid the perpetuation of non-informative diagnostic practice. The effects of practitioner experience on aspects of the clinical examination could also be modelled.
- What is the accuracy of the glucose test, and can it be used in non-toilet-trained children? Although accuracy results for this test were promising, the evidence identified in the review was sparse and out of date, and related to older children. Although this test does not appear a practical option in infants, its performance may warrant further investigation.
- What is the accuracy of microscopy in combination with a dipstick test, in particular in

patients with an indeterminate nitrite and LE dipstick test result? Studies identified in this review have generally determined the accuracy of individual tests. Information on cumulative accuracy, where the application of a second test is conditional on the result of the first, may help to inform the decision on the role of intermediate test (e.g. microscopy) in the diagnosis of UTI.

A single study, which has the potential to answer most of these questions, would be a prevalence function study. Such a study would use logistic regression analysis to model the presence or absence of a UTI as the dependent variable (as determined by culture). A number of other factors could be included as independent variables and their association with the presence or absence of a UTI could be modelled. Such a study would be able to model the added benefit of each test over the other tests already included in the model. To answer the question on the benefits of culturing to determine antibiotic sensitivity, an RCT could be conducted. This would randomise children with confirmed UTI to either all receiving culture to determine antibiotic sensitivities, or only receiving culture if they do not respond to first line antibiotics. For diagnostic cohort studies, priority should be given to evaluating the use of pad and nappy samples and to clarifying further the performance of combined dipstick tests (near patient testing).

Further investigation of UTI

Studies assessing the clinical effectiveness of all stages of the further investigation of UTI, for long-term renal outcomes, are urgently required. These should evaluate:

- imaging for reflux
- imaging for scarring
- routine ultrasound
- imaging for initial versus repeat infection
- differences in the effectiveness of imaging between boys and girls and for different age groups (e.g. <1 year, 1–2 years, 2–5 years).

Such studies could compare imaging with no imaging, single imaging strategies or combination imaging strategies. Given the ethical problems surrounding RCTs of diagnostic tests, the very large sample sizes required and the particular issues around investigating UTI in very young boys, a multivariable prediction modelling approach (as described above) could be used to quantify the cumulative value of tests against the long-term outcome measure of PRS.

Before further diagnostic research is undertaken, priority should be given to clarifying whether any link exists between the presence of reflux and PRS. In addition, it is necessary to establish whether there are any effective interventions to reduce PRS, which would be indicated by the results of testing.

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 APN. Non-invasive methods of localisation require further research addressed at this aim.

Acknowledgements

We would like to thank Sarah King for help in editing the document and Martin Bland for statistical advice. We would also like to thank the advisory panel to the review for commenting on the protocol and draft report. Members of the advisory panel are listed in Appendix 1.

Contribution of authors

Penny Whiting (Research Fellow) and Marie Westwood (Reviews Manager) were responsible for conducting the systematic review and contributed to drafting the report. Laura Bojke (Research Fellow) and Stephen Palmer (Senior Research Fellow) were responsible for conducting the economic modelling and contributed to drafting the economics sections of the report. Gerry Richardson (Research Fellow) contributed preliminary work on the economic modelling. Julie Cooper (Consultant Radiologist) and Ian Watt (Professor of Primary Care) contributed clinical advice and reviewed all draft versions of the report. Julie Glanville (Associate Director) carried out the literature searches and drafted the sections reporting on searches. Mark Sculpher (Professor of Health Economics) managed the economic modelling work. Jos Kleijnen (Director) was the review manager.

References

1. Dark G. *The on-line medical dictionary* [database]. Academic Medical Publishing and CancerWEB; 1998. URL: http://cancerweb.ncl.ac.uk/omd/index.html. Accessed 23 June 2004.
2. *Urinary tract infections. Bacterial infections.* Merck; 2001. URL: http://www.merck.com/pubs/mmanual/section17/chapter227/227a.htm. Accessed 11 January 2002.
3. Larcombe J. Urinary tract infection. In *Clinical evidence,* Issue 7. London: BMJ Publishing; 2002. pp. 377–85.
4. The management of urinary tract infection in children. *Drug Ther Bull* 1997;**35**:65–9.
5. Honkinen O, Lehtonen OP, Ruuskanen O, Huovinen P, Mertsola J. Cohort study of bacterial species causing urinary tract infection and urinary tract abnormalities in children. *BMJ* 1999;**318**: 770–1.
6. Fotter R, editor. *Pediatric uroradiology*. Berlin: Springer; 2001.
7. Vernon SJ, Coulthard MG, Lambert HJ, Keir MJ, Matthews JN. New renal scarring in children who at age 3 and 4 years had had normal scans with dimercaptosuccinic acid: follow up study. *BMJ* 1997;**315**:905–8.
8. Smellie JM, Poulton A, Prescod NP. Retrospective study of children with renal scarring associated with reflux and urinary infection. *BMJ* 1994;**308**: 1193–6.
9. Sacks SH, Verrier Jones K, Roberts R, Asscher AW, Ledingham JG. Effect of symptomless bacteriuria in childhood on subsequent pregnancy. *Lancet* 1987;**2**:991–4.
10. Downs SM. Technical report: urinary tract infections in febrile infants and young children. Urinary Tract Subcommittee of the American Academy of Pediatrics Committee on Quality Improvement. *Pediatrics* 1999;**103**:e54.
11. Children's Hospital Medical Center Cincinnati. *Evidence based clinical practice guideline for patients 6 years of age or less with a first time acute urinary tract infection (UTI).* Cincinnati, OH: Children's Hospital Medical Center, 2000. URL: http://www.guidelines.gov. Accessed 26 July 2006, guideline last updated 18 April 2005.
12. *UTI (lower) – acute.* Department of Health; 2001. URL: http://www.prodigy.nhs.uk/guidance/crd/UTI%20(lower)%20-%20acute.htm. Accessed 11 January 2002.
13. Working Group of the Research Unit of the Royal College of Physicians. Guidelines for the management of acute urinary tract infection in childhood. *J R Coll Physicians* 1991;**25**:36–43.
14. Sherbotie JR, Cornfeld D. Management of urinary tract infections in children. *Med Clin North Am* 1991;**75**:327–38.
15. Dick PT, Feldman W. Routine diagnostic imaging for childhood urinary tract infections: a systematic overview. *J Pediatr* 1996;**128**:15–22.
16. James JM, Testa HJ. Imaging techniques in the diagnosis of urinary tract infection. *Curr Opin Nephrol Hypertens* 1994;**3**:660–4.
17. Society of Nuclear Medicine. *Procedure guideline for radionuclide cystography in children.* Reston, VA: Society of Nuclear Medicine; 1999. URL: http://www.guidelines.gov
18. The diagnosis, treatment and evaluation of the initial urinary tract infection in febrile infants and young children. *Pediatrics* 1999;**103**:843–52.
19. Ahmed SM, Swedlund SK. Evaluation and treatment of urinary tract infections in children. *Am Fam Physician* 1998;**57**:1573–80, 1583–4.
20. Hobbs FD, Delaney BC, Fitzmaurice DA, Wilson S, Hyde CJ, Thorpe GH, *et al*. A review of near patient testing in primary care. *Health Technol Assess* 1997;**1**(5).
21. Upadhyay J, Bolduc S, Braga L, Farhat W, Bagli DJ, McLorie GA, *et al*. Impact of prenatal diagnosis on the morbidity associated with ureterocele management. *J Urol* 2002;**167**:2560–5.
22. Stokland E, Hellstrom M, Hansson S, Jodal O, Oden A, Jacobsson B. Reliability of ultrasonography in identification of reflux nephropathy in children. *BMJ* 1994;**309**:235–9. URL: http://bmj.com/cgi/content/abstract/ 309/6949/235
23. Roebuck DJ, Howard RG, Metreweli C. How sensitive is ultrasound in the detection of renal scars? *Br J Radiol* 1999;**72**:345–8.
24. NHS Centre for Reviews and Dissemination. *Undertaking systematic reviews of research on effectiveness. CRD's guidance for those carrying out or commissioning reviews.* 2nd ed. CRD Report No. 4. York: NHS Centre for Reviews and Dissemination; 2001.
25. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol* 1995;**48**:119–30.

26. Irwig L, Tostesen ANA, Gatsonis C, Lau J, Colditz G, Chalmers TC, *et al*. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med* 1994;**120**:667–76.

27. Whiting P, Rutjes AWS, Dinnes J, Reitsma JB, Bossuyt PMM, Kleijnen J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess* 2004;**8**(25).

28. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of diagnostic accuracy studies included in systematic reviews. *BMC Medical Research Methodology* 2003;**3**:25.

29. Chateil JF, Diard F. Management of urinary tract infection in children. *Feuillets de Radiol* 1988;**28**:229–37.

30. Drummond MF, O'Brien B, Stoddart GL, Torrance GW. Critical assessment of economic evaluation. *Methods for the economic evaluation of health care programmes*. 2nd ed. Oxford: Oxford University Press; 2000. pp. 27–51.

31. National Institute for Clinical Excellence. *Guide to the methods of technology appraisal*. London: National Institute for Clinical Excellence; 2004. URL: http://www.nice.org.uk/pdf/Techappprocessconsultation.pdf. Accessed 26 July 2006.

32. Perri G, Napoli V, Falaschi F, Esposito S, Federighi F. [Digital cystography in the evaluation of vesico-ureteral reflux in childhood]. *Radiol Med* 1989;**77**:373–7.

33. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;**7**:177–88.

34. Fleiss J. The statistical basis of meta-analysis. *Stat Methods Med Res* 1993;**2**:121–45.

35. Galbraith R. A note on graphical representation of estimated odds ratios from several clinical trials. *Stat Med* 1988;**7**:889–94.

36. Moses L, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med* 1993;**12**:1293–316.

37. Lijmer JG, Bossuyt PMM, Heisterkamp SH. Exploring sources of heterogeneity in systematic review of diagnostic tests. *Stat Med* 2002;**21**:1525–37.

38. Ahmad T, Vickers D, Campbell S, Coulthard MG, Pedler S. Urine collection from disposable nappies. *Lancet* 1991;**338**:674–6.

39. Anad FY. A simple method for selecting urine samples that need culturing. *Ann Saudi Med* 2001;**21**:104–5.

40. Armengol CE, Hendley JO, Schlager TA. Urinary tract infection in young children cannot be excluded with urinalysis. *Pediatr Res* 2000;**47**:172A.

41. Aronson AS, Gustafson B, Svenningsen NW. Combined suprapubic aspiration and clean-voided urine examination in infants and children. *Acta Paediatr Scand* 1973;**62**:396–400.

42. Arslan S, Caksen H, Rastgeldi L, Uner A, Oner AF, Odabas D. Use of urinary Gram stain for detection of urinary tract infection in childhood. *Yale J Biol Med* 2002;**75**:73–8.

43. Bachur R, Harper MB. Reliability of the urinalysis for predicting urinary tract infections in young febrile children. *Arch Pediatr Adolesc Med* 2001;**155**:60–5.

44. Baum JD, Darrell JH, Lambert RA. Evaluation of dip inoculum urine culture. *Arch Dis Child* 1972;**47**:977–8.

45. Boreland PC, Stoker M. Dipstick analysis for screening of paediatric urine. *J Clin Pathol* 1986;**39**:1360–2.

46. Braude H, Forfar JO, Gould JC, McLeod JW. Diagnosis of urinary tract infection in childhood based on examination of pared non-catheter and catheter specimens of urine. *BMJ* 1967;**4**:702–5.

47. Bulloch B, Bausher JC, Pomerantz WJ, Connors JM, Mahabee-Gittens M, Dowd MD. Can urine clarity exclude the diagnosis of urinary tract infection? *Pediatrics* 2000;**106**:E60.

48. Rodriguez Caballero AM, Novoa Vazquez P, Perez Ruiz A, Carmona Perez A, Cano Fernandez J, Sanchez Bayle M. Assessment of leukocyturia in the diagnosis of urinary tract infections. *Revista Espanola de Pediatria* 2001;**57**:305–8.

49. Rodriguez Cervilla J, Alonso Alonso C, Fraga Bermudez JM, Perez Munuzuri A, Gil Calvo M, Ariceta Iraola G, *et al*. Urinary tract infection in children: clinical and analytical prospective study for differential diagnosis in children with suspicion of an infectious disease. *Revista Espanola Pediatria* 2001;**57**:144–52.

50. Cid E, Fernandez Seara MJ, Buznego R, Pavon P, Rodrigo E, Castro-Gago M. Comparative study between Uricult and urine culture for the diagnosis of urinary infections in infants. *Revista Espanola Pediatria* 1992;**48**:23–5.

51. Cohen HA, Woloch B, Linder N, Vardi A, Barzilai A. Urine samples from disposable diapers: an accurate method for urine cultures. *J Fam Pract* 1997;**44**:290–2.

52. Craver RD, Abermanis JG. Dipstick only urinalysis screen for the pediatric emergency room. *Pediatr Nephrol* 1997;**11**:331–3.

53. Dayan PS, Bennett J, Best R, Bregstein JS, Levine D, Novick MK, *et al*. Test characteristics of the urine Gram stain in infants less than 60 or 60 days of age with fever. *Pediatr Emerg Care* 2002;**18**:12–14.

54. Dayan PS, Chamberlain JM, Boenning D, Adirim T, Schor JA, Klein BL. A comparison of the initial to the later stream urine in children catheterized to evaluate for a urinary tract infection. *Pediatr Emerg Care* 2000;**16**:88–90.

55. Demi M, Costa L, Zanardo V. [Urinary tract infections in newborns: sensitivity, specificity, and predictive value of urinary screening with the reagent strip test]. *Pediatr Med Chir* 1993;**15**:29–31.

56. Dosa S, Houston IB, Allen LB, Jennison R. Urinary glucose unreliable as test for urinary tract infection in infancy. *Arch Dis Child* 1973;**48**:733–7.

57. Farrell M, Devine K, Lancaster G, Judd B. A method comparison study to assess the reliability of urine collection pads as a means of obtaining urine specimens from non-toilet-trained children for microbiological examination. *J Adv Nurs* 2002;**37**:387–93.

58. Feasey S. Are Newcastle urine collection pads suitable as a means of collecting specimens from infants? *Pediatr Nurs* 1999;**11**:17–21.

59. Fennell RS, Wilson SG, Garin EH, Pryor ND, Sorgen CD, Walker RD, *et al*. The combination of two screening methods in a home culture program for children with recurrent bacteriuria. An evaluation of a culture method plus a nitrite reagent test strip. *Clin Pediatr* 1977;**16**:951–5.

60. Benito Fernandez J, Garcia Ribes A, Trebolazabala Quirante N, Mintegi Raso S, Vazquez Ronco M, Urra Zalbidegoitia E. [Gram stain and dipstick as diagnostic methods for urinary tract infection in febrile infants]. *An Esp Pediatr* 2000;**53**:561–6.

61. Benito Fernandez J, Sanchez Echaniz J, Mintegui Raso S, Montejo F. Urinary tract infection in infants: use of urine specimens obtained by suprapubic bladder aspiration in order to determine the reliability of culture specimen of urine collected in perineal bag. *An Esp Pediatr* 1996;**45**:149–52.

62. Giraldez M, Perozo M, González F, Rodríguez M. Infección urinaria cinta reactiva y sedimento urinario vs. urocultivo para determinación de bacteriuria. *Salus Militiae* 1998;**23**:27–31.

63. Godard C, Frutiger P, Delarue C, Christen JP, Wavre D, Girardet P. Testing for bacteriuria by home culturing in preschool girls. *Helvetica Pediatrica Acta* 1979;**34**:209–12.

64. Gorelick MH, Shaw KN. Clinical decision rule to identify febrile young girls at risk for urinary tract infection. *Arch Pediatr Adolesc Med* 2000;**154**: 386–90.

65. Hardy JD, Furnell PM, Brumfitt W. Comparison of sterile bag, clean catch and suprapubic aspiration in the diagnosis of urinary infection in early childhood. *Br J Urol* 1976;**48**:279–83.

66. Hiraoka M, Hida Y, Hori C, Tsuchida S, Kuroda M, Sudo M. Urine microscopy on a counting chamber for diagnosis of urinary infection. *Acta Paediatrica Japonica* 1995;**37**:27–30.

67. Hoberman A, Wald ER, Reynolds EA, Penchansky L, Charron M. Is urine culture necessary to rule out urinary tract infection in young febrile children? *Pediatr Infect Dis J* 1996;**15**:304–9.

68. Hoberman A, Wald ER, Penchansky L, Reynolds EA, Young S. Enhanced urinalysis as a screening test for urinary tract infection. *Pediatrics* 1993;**91**:1196–9.

69. Hoberman A, Wald ER, Reynolds EA, Penchansky L, Charron M. Pyuria and bacteriuria in urine specimens obtained by catheter from young children with fever. *J Pediatr* 1994;**124**: 513–19.

70. Holland PD, Doyle CT, English L. An evaluation of chemical tests for significant bacteriuria. *Journal of the Irish Medical Association* 1968;**61**:128–30.

71. Kohler L, Fritz H, Schersten B. Screening for bacteriuria with Uriglox in children. *Acta Paediatr Scand Suppl* 1970;**206**:76–8.

72. Kunin CM, DeGroot JE. Sensitivity of a nitrite indicator strip method in detecting bacteriuria in preschool girls. *Pediatrics* 1977;**60**:244–5.

73. Labbe J. [Usefulness of testing for nitrites in the diagnosis of urinary infections in children]. *Union Medicale du Canada* 1982;**111**:261–5.

74. Lagos Zuccone R, Carter S J, Herrera Labarca P. Utilidad de una tira reactiva y del aspecto macroscópico de la orina para descartar la sospecha clínica de infección del tracto urinario en niños ambulatorios. *Revista Chilena de Pediatria* 1994;**65**:88–94.

75. Lejeune B, Baron R, Guillois B, Mayeux D. Evaluation of a screening test for detecting urinary tract infection in newborns and infants. *J Clin Pathol* 1991;**44**:1029–30.

76. Lin DS, Huang SH, Lin CC, Tung YC, Huang TT, Chiu NC, *et al*. Urinary tract infection in febrile infants younger than eight weeks of age. *Pediatrics* 2000;**105**:E20.

77. Lin DS, Huang FY, Chiu NC, Koa HA, Hung HY, Hsu CH, *et al*. Comparison of hemocytometer leukocyte counts and standard urinalyses for predicting urinary tract infections in febrile infants. *Pediatr Infect Dis J* 2000;**19**:223–7.

78. Liptak GS, Campbell J, Stewart R, Hulbert WC. Screening for urinary tract infection in children with neurogenic bladders. *Am J Phys Med Rehabil* 1993;**72**:122–6.

79. Littlewood JM, Jacobs SI, Ramsden CH. Comparison between microscopical examination of unstained deposits of urine and quantitative culture. *Arch Dis Child* 1977;**52**:894–6.

80. Lockhart GR, Lewander WJ, Cimini DM, Josephson SL, Linakis JG. Use of urinary gram stain for detection of urinary tract infection in infants. *Ann Emerg Med* 1995;**25**:31–5.

81. Lohr JA, Portilla MG, Geuder TG, Dunn ML, Dudley SM. Making a presumptive diagnosis of urinary tract infection by using a urinalysis performed in an on-site laboratory. *J Pediatr* 1993;**122**:22–5.

82. Manson R, Scholefield J, Johnston RJ, Scott R. The screening of more than 2000 schoolgirls for bacteriuria using an automated fluorescence microscopy system. *Urol Res* 1985;**13**:143–8.

83. Marret M, Tay S, Yap HK, Murugasu B. Comparison of two rapid screening tests for urinary tract infection in children. *Ann Acad Med Singapore* 1995;**24**:299.

84. Marsik FJ, Owens D, Lewandowski J. Use of the leukocyte esterase and nitrite tests to determine the need for culturing urine specimens from a pediatric and adolescent population. *Diagn Microbiol Infect Dis* 1986;**4**:181–3.

85. Matthai J, Ramaswamy M. Urinalysis in urinary tract infection. *Indian J Pediatr* 1995;**62**:713–16.

86. Mongeau JG, Robillard JE, Brousseau Y. Screening for bacteriuria in children: comparison of two dip-tests. *CMAJ* 1972;**107**:227–9.

87. Morton RE, Lawande R. I: The diagnosis of urinary tract infection: comparison of urine culture from suprapubic aspiration and midstream collection in a children's out-patient department in Nigeria. *Ann Trop Paediatr* 1982;**2**:109–12.

88. Dominguez Navarrete N. [Evaluation of triphenyl tetrazolium chloride as a diagnostic test in urinary infections]. *Anales de la Facultad de Medicina, Universidad Nacional Mayor de San Marcos de Lima* 1966;**49**:294–307.

89. Villanustre Ordonez C, Buznego Sanchez R, Rodicio Garcia M, Rodrigo Saez E, Fernandez Seara MJ, Pavon Belinchon P, *et al*. Comparative study of semiquantitative methods (leukocytes, nitrite test and uricult) with urine culture for the diagnosis of urinary tract infection during infancy. *An Esp Pediatr* 1994;**41**:325–8.

90. Palmer LS, Richards I, Kaplan WE. Clinical evaluation of a rapid diagnostic screen (URISCREEN) for bacteriuria in children. *J Urol* 1997;**157**:654–7.

91. Parmington J, Kornberg A. Nitrite screening for urinary tract infection in a pediatric emergency department. *Pediatr Emerg Care* 1989;**5**:285–6.

92. Pryles CV, Eliot CR. Pyuria and bacteriuria in infants and children. The value of pyuria as a diagnostic criterion of urinary tract infections. *Am J Dis Child* 1965;**110**:628–35.

93. Purwar VN, Agrawal SP, Dikshit SK. Gram stained urine slides in the diagnosis of urinary tract infections in children. *J Indian Med Assoc* 1972;**59**:387–8.

94. Pylkkanen J, Vilska J, Koskimies O. Diagnostic value of symptoms and clean-voided urine specimen in childhood urinary tract infection. *Acta Paediatr Scand* 1979;**68**:341–4.

95. Ramage IJ, Chapman JP, Hollman AS, Elabassi M, McColl JH, Beattie TJ. Accuracy of clean-catch urine collection in infancy. *J Pediatr* 1999;**135**: 765–7.

96. Rich G, Glass NJ, Selkon JB. Cost-effectiveness of two methods of screening for asymptomatic bacteriuria. *British Journal of Preventive and Social Medicine* 1976;**30**:54–9.

97. Santos MA, Mos EdN, Schmidt BJ, Piva S. Comparacion entre el estudio bacterioscopico cuantitativo y el urocultivo para el diagnostico de infeccion urinaria en pediatria. *Boletin Médico del Hospital Infantil de México* 1982;**39**:526–30.

98. Saxena H, Ajwani KD, Mehrotra D. Quantitative pyuria in the diagnosis of urinary infections in children. *Indian J Pediatr* 1975;**42**:35–8.

99. Schersten B, Dahlqvist A, Fritz H, Kohler L, Westlund L. Screening for bacteriuria with a test paper for glucose. *JAMA* 1968;**204**:205–8.

100. Schreiter G, Buhtz P. [Diagnostic value of the cytologic and bacteriologic urine examinations in pediatrics. II. Comparison of leukocyturia and bacteriuria]. *Deutsche Gesundheitswesen* 1971;**26**:1318–23.

101. Sharief N, Hameed M, Petts D. Use of rapid dipstick tests to exclude urinary tract infection in children. *Br J Biomed Sci* 1998;**55**:242–6.

102. Shaw KN, McGowan KL, Gorelick MH, Schwartz JS. Screening for urinary tract infection in infants in the emergency department: which test is best? *Pediatrics* 1998;**101**:E1–5.

103. Shaw KN, Hexter D, McGowan KL, Schwartz JS. Clinical evaluation of a rapid screening test for urinary tract infections in children. *J Pediatr* 1991;**118**:733–6.

104. Struthers S, Scanlon J, Parker K, Goddard J, Hallett R. Parental reporting of smelly urine and urinary tract infection. *Arch Dis Child* 2003;**88**:250–2.

105. Tahirovic H, Pasic M. A modified nitrite test as a screening test for significant bacteriuria. *Eur J Pediatr* 1988;**147**:632–3.

106. Todd J, McLain L, Duncan B, Brown M. A nonculture method for home follow-up of urinary tract infections in childhood. *J Pediatr* 1974;**85**:514–16.

107. Vangone G, Russo G. [Bacteria and leukocyte count in the urine in the diagnosis of urinary tract infections]. *Pediatr Med Chir* 1985;**7**:125–9.

108. Vickers D, Ahmad T, Coulthard MG. Diagnosis of urinary tract infection in children: fresh urine microscopy or culture? *Lancet* 1991;**338**:767–70.

109. Waisman Y, Zerem E, Amir L, Mimouni M. The validity of the uriscreen test for early detection of urinary tract infection in children. *Pediatrics* 1999;**104**:e41.

110. Wammanda RD, Aikhionbare HA, Ogala WN. Use of nitrite dipstick test in the screening for urinary tract infection in children. *West Afr J Med* 2000;**19**:206–8.

111. Weinberg AG, Gan VN. Urine screen for bacteriuria in symptomatic pediatric outpatients. *Pediatr Infect Dis J* 1991;**10**:651–4.

112. Wiggelinkhuizen J, Maytham D, Hanslo DH. Dipstick screening for urinary tract infection. *S Afr Med J* 1988;**74**:224–8.

113. Woodward MN, Griffiths DM. Use of dipsticks for routine analysis of urine from children with acute abdominal pain. *BMJ* 1993;**306**:1512.

114. Alon U, Pery M, Davidai G, Berant M. Ultrasonography in the radiologic evaluation of children with urinary tract infection. *Pediatrics* 1986;**78**:58–64.

115. Alzen G, Wildberger JE, Muller-Leisse C, Deutz FJ. [Ultrasound screening of vesico-uretero-renal reflux]. *Klin Padiatr* 1994;**206**:178–80.

116. Andrich MP, Majd M. Evaluation of febrile urinary tract infections in children utilizing technetium-99m DMSA scanning. *J Nucl Med* 1992;**33**:976.

117. Bagni B, Orsolon P, Fattori A, Guerra UP. Renal SPECT with Tc-99m DMSA in children with upper urinary tract infections using a triple-headed gamma camera. *Clin Nucl Med* 1997;**22**:838–43.

118. Barnett GR, Abbott GD. Localization of gram negative urinary tract infection by immunofluorescence in infants and children. *Aust Paediatr J* 1978;**14**:143–6.

119. Baronciani D, Bonora G, Andreoli A, Cambie M, Nedbal M, Dellagnola CA. The value of ultrasound for diagnosing the uropathy in children with urinary-tract infections. *Rivista Italiana di Pediatria* 1986;**12**:214–20.

120. Barry BP, Hall N, Cornford E, Broderick NJ, Somers JM, Rose DH. Improved ultrasound detection of renal scarring in children following urinary tract infection. *Clin Radiol* 1998;**53**: 747–51.

121. Benador D, Benador N, Slosman DO, Nussle D, Mermillod B, Girardin E. Cortical scintigraphy in the evaluation of renal parenchymal changes in children with pyelonephritis. *J Pediatr* 1994;**124**: 17–20.

122. Benigno V, Di Peri S, Distefano F, Como G, Boncori R. [Laboratory parameters as a guide for the radiologic study of infections of the urinary tract in childhood]. *Pediatr Med Chir* 1986;**8**:91–3.

123. Bergius AR, Niskanen K, Kekomaki M. Detection of significant vesico-ureteric reflux by ultrasound in infants and children. *Zeitschift für Kinderchirurgie* 1990;**45**:144–5.

124. Berrocal T, Gaya F, Arjonilla A, Lonergan GJ. Vesicoureteral reflux: diagnosis and grading with echo-enhanced cystosonography versus voiding cystourethrography. *Radiology* 2001;**221**:359–65.

125. Biggi A, Dardanelli L, Pomero G, Cussino P, Noello C, Sernia O, *et al*. Acute renal cortical scintigraphy in children with a first urinary tract infection. *Pediatr Nephrol* 2001;**16**:733–8.

126. Bircan ZE, Buyan N, Hasanoglu E, Ozturk E, Bayhan H, Isik S. Radiologic evaluation of urinary tract infection. *Int Urol Nephrol* 1995;**27**:27–32.

127. Boudailliez B, Berro Y, Hosri JA, Esper IE, Grumbach Y. A prospective study of imaging modalities in acute pyelonephritis (49 cases): DMSA renal scintigraphy versus power Doppler (PD) renal sonography. *Pediatr Nephrol* 1998;**12**:C132.

128. Bower G, Lovegrove FT, Geijsel H, Van der Schaff A, Guelfi G. Comparison of 'direct' and 'indirect' radionuclide cystography. *J Nucl Med* 1985;**26**:465–8.

129. Buyan N, Bircan ZE, Hasanoglu E, Ozturk E, Bayhan H, Rota S. The importance of ^{99m}Tc DMSA scanning in the localization of childhood urinary tract infections. *Int Urol Nephrol* 1993;**25**:11–17.

130. Capa Kaya G, Taskiran Y, Bak M, Aydin A, Toksavul O, Uslu Y, *et al*. Urinary N-acetyl-beta-glucosaminidase in children with upper urinary tract infection, in relation to Tc-99m DMSA scintigraphy. *Eur J Nucl Med* 2001;**28**:1156.

131. Cavanagh PM, Sherwood T. Too many cystograms in the investigation of urinary tract infection in children? *Br J Urol* 1983;**55**:217–19.

132. Chan YL, Chan KW, Yeung CK, Roebuck DJ, Chu WC, Lee KH, *et al*. Potential utility of MRI in the evaluation of children at risk of renal scarring. *Pediatr Radiol* 1999;**29**:856–62.

133. Clarke SEM, Risheq F, Mistry R, Maisey MN. An evaluation of technetium-99m dimercaptosuccinic acid DMSA tomographic imaging in children with urinary tract infection. *Eur J Nucl Med* 1990;**16**:415.

134. Dacher JN, Pfister C, Monroc M, Eurin D, Le Dosseur P. Power Doppler sonographic pattern

of acute pyelonephritis in children: comparison with CT. *AJR Am J Roentgenol* 1996;**166**:1451–5.

135. De Sadeleer C, De Boe V, Keuppens F, Desprechins B, Verboven M, Piepsz A. How good is technetium-99m mercaptoacetyltriglycine indirect cystography? *Eur J Nucl Med* 1994;**21**:223–7.

136. Ditchfield MR, De Campo JF, Cook DJ, Nolan TM, Powell HR, Sloane R, *et al*. Vesicoureteral reflux: an accurate predictor of acute pyelonephritis in childhood urinary tract infection? *Radiology* 1994;**190**:413–15.

137. Drachman R, Valevici M, Vardy PA. Excretory urography and cystourethrography in the evaluation of children with urinary tract infection. *Clin Pediatr* 1984;**23**:265–7.

138. Elison BS, Taylor D, Van der Wall H, Pereira JK, Cahill S, Rosenberg AR, *et al*. Comparison of DMSA scintigraphy with intravenous urography for the detection of renal scarring and its correlation with vesicoureteric reflux. *Br J Urol* 1992;**69**:294–302.

139. Evans ED, Meyer JS, Harty MP, Bellah RD. Assessment of increase in renal pelvic size on post-void sonography as a predictor of vesicoureteral reflux. *Pediatr Radiol* 1999;**29**:291–4.

140. Everaert K, Raes A, Hoebeke P, Samijn W, Delanghe J, Vande Wiele C, *et al*. Combined use of urinary alpha$_1$-microglobulin and ^{99m}Tc DMSA scintigraphy in the diagnosis and follow-up of acute pyelonephritis and cystitis in children. *Eur Urol* 1998;**34**:486–91.

141. Farnsworth RH, Rossleigh MA, Leighton DM, Bass SJ, Rosenberg AR. The detection of reflux nephropathy in infants by ^{99m}technetium dimercaptosuccinic acid studies. *J Urol* 1991;**145**:542–6.

142. Foresman WH, Hulbert WC, Rabinowitz R. Does urinary tract ultrasonography at hospitalization for acute pyelonephritis predict vesicoureteral reflux? *J Urol* 2001;**165**:2232–4.

143. Fretzayas A, Moustaki M, Gourgiotis D, Bossios A, Koukoutsakis P, Stavrinadis C. Polymorphonuclear elastase as a diagnostic marker of acute pyelonephritis in children. *Pediatrics* 2000;**105**:E28.

144. Berrocal Frutos T, Gaya Moreno F, Gomez Leon N, Jaureguizar Monereo E. [Cystosonography with echoenhancer. A new imaging technique for the diagnosis of vesicoureteral reflux]. *An Esp Pediatr* 2000;**53**:422–30.

145. Gervaix A, Galetto-Lacour A, Gueron T, Vadas L, Zamora S, Suter S, *et al*. Usefulness of procalcitonin and C-reactive protein rapid tests for the management of children with urinary tract infection. *Pediatr Infect Dis J* 2001;**20**:507–11.

146. Castello Girona F, Vilaplana Canto E, Yeste Fernandez D, Roca Bielsa I, Enriquez Civico G. ^{99m}Tc dimercaptosuccinic scan in the study of the first urinary tract infection in infants. *An Esp Pediatr* 1995;**42**:118–22.

147. Gordon I, Anderson PJ, Lythgoe MF, Orton M. Can technetium-99m-mercaptoacetyltriglycine replace technetium-99m-dimercaptosuccinic acid in the exclusion of a focal renal defect? *J Nucl Med* 1992;**33**:2090–3.

148. Guermazi F, Lenoir P, Verboven M, Smets A, Braeckman J, Jonckheer MH, *et al*. [Technetium 99m labeled dimercaptosuccinic acid (^{99m}Tc-DMSA) scintigraphy in the diagnosis and follow-up of urinary infections in children]. *Archives Françaises de Pediatrie* 1993;**50**:391–8.

149. Haberlik A. Detection of low-grade vesicoureteral reflux in children by color Doppler imaging mode. *Pediatr Surg Int* 1997;**12**:38–43.

150. el Hajjar M, Launay S, Hossein-Foucher C, Foulard M, Robert Y. [Power Doppler sonography and acute pyelonephritis in children: comparison with Tc-DMSA scintigraphy]. *Arch Pediatr* 2002;**9**:21–5.

151. Hanbury DC, Whitaker RH, Sherwood T, Farman P. Ultrasound and plain X-ray screening in childhood urinary tract infection. *Br J Urol* 1989;**64**:638–40.

152. Hedman PJ, Kempi V, Voss H. Measurement of vesicoureteral reflux with intravenous ^{99m}Tc-DTPA compared to radiographic cystography. *Radiology* 1978;**126**:205–8.

153. Hellerstein S, Kennedy E, Nussbaum L, Rice K. Localization of the site of urinary tract infections by means of antibody-coated bacteria in the urinary sediments. *J Pediatr* 1978;**92**:188–93.

154. Hellstrom M, Jacobsson B, Marild S, Jodal U. Voiding cystourethrography as a predictor of reflux nephropathy in children with urinary-tract infection. *AJR Am J Roentgenol* 1989;**152**:801–4.

155. Hitzel A, Liard A, Vera P, Manrique A, Menard JF, Dacher JN. Color and power Doppler sonography versus DMSA scintigraphy in acute pyelonephritis and in prediction of renal scarring. *J Nucl Med* 2002;**43**:27–32.

156. Ilyas M, Mastin ST, Richard GA. Age-related radiological imaging in children with acute pyelonephritis. *Pediatr Nephrol* 2002;**17**:30–4.

157. Jakobsson B, Nolstedt L, Svensson L, Soderlundh S, Berg U. ^{99m}Technetium-dimercaptosuccinic acid scan in the diagnosis of acute pyelonephritis in children: relation to clinical and radiological findings. *Pediatr Nephrol* 1992;**6**:328–34.

158. Jantausch BA, Rifai N, Getson P, Akram S, Majd M, Wiedermann BL. Urinary *N*-acetyl-beta-

glucosaminidase and beta-2-microglobulin in the diagnosis of urinary tract infection in febrile infants. *Pediatr Infect Dis J* 1994;**13**:294–9.

159. Jequier S, Forbes PA, Nogrady MB. The value of ultrasonography as a screening procedure in a first-documented urinary tract infection in children. *J Ultrasound Med* 1985;**4**:393–400.

160. Jequier S, Jequier JC, Hanquinet S. Acute childhood pyelonephritis: predictive value of positive sonographic findings in regard to later parenchymal scarring. *Acad Radiol* 1998;**5**:344–53.

161. Johnson CE, Shurin PA, Marchant CD, Strieter CM, Murdell-Panek D, Debaz BP, *et al*. Identification of children requiring radiologic evaluation for urinary infection. *Pediatr Infect Dis* 1985;**4**:656–63.

162. Johnson CE, Vacca CV, Fattlar D, Fulton DJ, Hall PW. Urinary *N* acetyl-beta-glucosaminidase and the selection of children for radiologic evaluation after urinary tract infection. *Pediatrics* 1990;**86**:211–16.

163. Kenda R, Kenig T, Silc M, Zupancic Z. Renal ultrasound and excretory urography in infants and young children with urinary tract infection. *Pediatr Radiol* 1989;**19**:299–301.

164. Kenda RB, Novljan G, Kenig A, Hojker S, Fettich JJ. Echo-enhanced ultrasound voiding cystography in children: a new approach. *Pediatr Nephrol* 2000;**14**:297–300.

165. Kessler RM, Altman DH. Real-time sonographic detection of vesicoureteral reflux in children. *AJR Am J Roentgenol* 1982;**138**:1033–6.

166. Krzemien G, Roszkowska-Blaim M, Brzewski M, Kostro I, Szmigielska A, Karpinska M, *et al*. Comparison of power Doppler ultrasonography with ^{99m}Tc-DMSA renal scintigraphy in the diagnosis of acute pyelonephritis. *Pol Merkuriusz Lek* 2002;**12**:405–7.

167. Landau D, Brennan J, Turner ME, Majd M. A negative urinalysis predicts the absence of acute pyelonephritis in febrile infants. *Pediatr Res* 1994;**35**:185.

168. Landau D, Turner ME, Brennan J, Majd M. The value of urinalysis in differentiating acute pyelonephritis from lower urinary tract infection in febrile infants. *Pediatr Infect Dis J* 1994;**13**:777–81.

169. Lavocat MP, Granjon D, Allard D, Gay C, Freycon MT, Dubois F. Imaging of pyelonephritis. *Pediatr Radiol* 1997;**27**:159–65.

170. Leonidas JC, McCauley RG, Klauber GC, Fretzayas AM. Sonography as a substitute for excretory urography in children with urinary tract infection. *AJR Am J Roentgenol* 1985;**144**:815–19.

171. LeQuesne GW, Davies R. Ultrasonic assessment of reflux nephropathy. *Pediatr Nephrol* 1986;**16**:335. URL: http://www.pediatrics.org/cgi/content/full/105/3/e42

172. Lindsell D, Moncrieff M. Comparison of ultrasound examination and intravenous urography after a urinary tract infection. *Arch Dis Child* 1986;**61**:81–2.

173. Lonergan GJ, Pennington DJ, Morrison JC, Haws RM, Grimley MS, Kao TC. Childhood pyelonephritis: comparison of gadolinium-enhanced MR imaging and renal cortical scintigraphy for diagnosis. *Radiology* 1998;**207**:377–84.

174. MacKenzie JR, Fowler K, Hollman AS, Tappin D, Murphy AV, Beattie TJ, *et al*. The value of ultrasound in the child with an acute urinary tract infection. *Br J Urol* 1994;**74**:240–4.

175. Mage K, Zoppardo P, Cohen R, Reinert P, Ponet M. Imagerie et premiere infection urinaire de l'enfant: place respective de chaque examen lors du bilan initial a propos de 122 observations. *J Radiol* 1989;**70**:279–83.

176. Mahant S, Friedman J, MacArthur C. Renal ultrasound findings and vesicoureteral reflux in children hospitalised with urinary tract infection. *Arch Dis Child* 2002;**86**:419–20.

177. Hitzel A, LiardZmuda A, Manrique A, Dacher JN, Vera P. Comparative study of DMSA scintigraphy (DMSA) and Doppler sonography (DS) in the diagnosis of acute pyelonephritis and scarring in children. *J Nucl Med* 2000;**41**:209.

178. McLorie GA, Aliabadi H, Churchill BM, Ash JM, Gilday DL. Technetium-99m-dimercapto-succinic acid renal scanning and excretory urography in diagnosis of renal scars in children. *J Urol* 1989;**142**:790–2.

179. Mentzel HJ, Vogt S, John U, Kaiser WA. Voiding urosonography with ultrasonography contrast medium in children. *Pediatr Nephrol* 2002;**17**:272–6.

180. Merrick MV, Uttley WS, Wild SR. The detection of pyelonephritic scarring in children by radioisotope imaging. *Br J Radiol* 1980;**53**:544–56.

181. Misselwitz J, Stoll W, Vogt S. [Value of radioisotope renography compared to urography, phenol test and concentration test in the diagnosis of pyelonephritis in childhood]. *Kinderarztliche Praxis* 1971;**44**:69–76.

182. Montplaisir S, Courteau C, Martineau B, Pelletier M. Limitations of the direct immunofluorescence test for antibody-coated bacteria in determining the site of urinary tract infections in children. *CMAJ* 1981;**125**:993–6.

183. Morin D, Veyrac C, Kotzki PO, Lopez C, Dalla Vale F, Durand MF, *et al*. Comparison of ultrasound and dimercaptosuccinic acid scintigraphy changes in acute pyelonephritis. *Pediatr Nephrol* 1999;**13**:219–22.

184. Mucci B, Maguire B. Does routine ultrasound have a role in the investigation of children with urinary tract infection? *Clin Radiol* 1994;**49**:324–5.

185. Muensterer OJ. Comprehensive ultrasound versus voiding cystourethrography in the diagnosis of vesicoureteral reflux. *Eur J Pediatr* 2002;**161**:435–7.

186. Oostenbrink R, van der Heijden AJ, Moons KG, Moll HA. Prediction of vesico-ureteric reflux in childhood urinary tract infection: a multivariate approach. *Acta Paediatr* 2000;**89**:806–10.

187. Piaggio G, Degli'Innocenti ML, Toma P, Calevo MG, Perfumo F. Cystosonography and voiding cystourethrography in the diagnosis of vesicoureteral reflux. *Pediatr Nephrol* 2003;**18**:18–22.

188. Pickworth FE, Vivian GC, Franklin K, Brown EF. ^{99m}Tc-mercapto acetyl triglycine in paediatric renal tract disease. *Br J Radiol* 1992;**65**:21–9.

189. Piepsz A, Pintelon H, Verboven M, Keuppens F, Jacobs A. Replacing ^{99m}Tc-DMSA for renal imaging? *Nucl Med Commun* 1992;**13**:494–6.

190. Pylkkanen J. Antibody-coated bacteria in the urine of infants and children with their first two urinary tract infections. *Acta Paediatr Scand* 1978;**67**:275–9.

191. Radmayr C, Klauser A, Pallwein L, Zurnedden D, Bartsch G, Frauscher F. Contrast enhanced reflux sonography in children: a comparison to standard radiological imaging. *J Urol* 2002;**167**:1428–30.

192. Redman JF, Seibert JJ. The role of excretory urography in the evaluation of girls with urinary tract infection. *J Urol* 1984;**132**:953–5.

193. Rehling M, Jensen JJ, Scherling B, Egeblad M, Lonborg-Jensen H, Kanstrup I, *et al*. Evaluation of renal function and morphology in children by ^{99m}Tc-DTPA gamma camera renography. *Acta Paediatr Scand* 1989;**78**:601–7.

194. Rickwood AM, Carty HM, McKendrick T, Williams MP, Jackson M, Pilling DW, *et al*. Current imaging of childhood urinary infections: prospective survey. *BMJ* 1992;**304**:663–5.

195. Von Rohden L, Bosse U, Wiemann D. [Reflux sonography in children with an ultrasound contrast medium in comparison to radiologic voiding cystourethrography]. *Paediatria Praxis* 1995;**49**:49–58.

196. Rossleigh MA, Wilson MJ, Rosenberg AR, Elison BS, Cahill S, Farnsworth RH. DMSA studies in infants under one year of age. *Contrib Nephrol* 1990;**79**:166–9.

197. Salih M, Baltaci S, Kilic S, Anafarta K, Beduk Y. Color flow Doppler sonography in the diagnosis of vesicoureteric reflux. *Eur Urol* 1994;**26**:93–7.

198. Scherz HC, Downs TM, Caesar R. The selective use of dimercaptosuccinic acid renal scans in children with vesicoureteral reflux. *J Urol* 1994;**152**:628–31.

199. Schneider K, Jablonski C, Weissner M, Kohn M, Fendel H. Screening for vesicoureteral reflux in children using real-time sonography. *Pediatr Radiol* 1984;**14**:400–3.

200. La Cava G, Sciagra R, Materassi M, Ienuso R, Meldolesi U. Accuracy of renal sequential scintigraphy for the recognition of renal involvement in pediatric patients affected by urinary tract infection. *Eur J Nucl Med* 1990;**16**:415.

201. Sfakianakis GN, Mylonakis T, Zilleruelo G, Leon M, Flores F, Ganz W, *et al*. The importance of technetium-99m GH scintigraphy in infants with first UTI. *J Nucl Med* 1989;**30**:915–16.

202. Siamplis D, Vasiou K, Giarmenitis S, Frimas K, Zavras G, Fezoulidis I. Sonographic detection of vesicoureteral reflux with fluid and air cystography. Comparison with VCUG. *Rofo Fortschr Geb Rontgenstr Neuen Bildgeb Verfahr* 1996;**165**:166–9.

203. Smellie JM, Rigden SP, Prescod NP. Urinary tract infection: a comparison of four methods of investigation. *Arch Dis Child* 1995;**72**:247–50.

204. Smolkin V, Koren A, Raz R, Colodner R, Sakran W, Halevy R. Procalcitonin as a marker of acute pyelonephritis in infants and children. *Pediatr Nephrol* 2002;**17**:409–12.

205. Sreenarasimhaiah V, Alon US. Uroradiologic evaluation of children with urinary tract infection: are both ultrasonography and renal cortical scintigraphy necessary? *J Pediatr* 1995;**127**:373–7.

206. Stokland E, Hellstrom M, Hansson S, Jodal U, Oden A, Jacobsson B. Reliability of ultrasonography in identification of reflux nephropathy in children. *BMJ* 1994;**309**:235–9.

207. Stokland E, Hellstrom M, Jacobsson B, Jodal U, Sixt R. Renal damage one year after first urinary tract infection: role of dimercaptosuccinic acid scintigraphy. *J Pediatr* 1996;**129**:815–20.

208. Stokland E, Hellstrom M, Jacobsson B, Jodal U, Lundgren P, Sixt R. Early ^{99m}Tc dimercaptosuccinic acid (DMSA) scintigraphy in symptomatic first-time urinary tract infection. *Acta Paediatr* 1996;**85**: 430–6.

209. Stokland E, Hellstrom M, Jacobsson B, Jodal U, Sixt R. Evaluation of DMSA scintigraphy and urography in assessing both acute and permanent renal damage in children. *Acta Radiol* 1998;**39**:447–52.

210. Tan SM, Chee T, Tan KP, Cheng HK, Ooi BC. Role of renal ultrasonography (RUS) and micturating cystourethrogram (MCU) in the assessment of vesico-ureteric reflux (VUR) in children and infants with urinary tract infection (UTI). *Singapore Med J* 1988;**29**:150–2.

211. Traisman ES, Conway JJ, Traisman HS, Yogev R, Firlit C, Shkolnik A, *et al*. The localization of urinary tract infection with ^{99}Tc glucoheptonate scintigraphy. *Pediatr Radiol* 1986;**16**:403–6.

212. Dura Trave T, Gonzalez Montero R, Juste Ruiz M, Gonzalez de Dios J, Carratala Marco F, Moya Benavent M, *et al*. [Usefulness of renal scintigraphy in the assessment of the first febrile urinary infection in children]. *An Esp Pediatr* 1997;**47**:378–82.

213. Valentini AL, Salvaggio E, Manzoni C, Rendeli C, Destito C, Summaria V, *et al*. Contrast-enhanced gray-scale and color Doppler voiding urosonography versus voiding cystourethrography in the diagnosis and grading of vesicoureteral reflux. *J Clin Ultrasound* 2001;**29**:65–71.

214. Verber IG, Strudley MR, Meller ST. ^{99m}Tc dimercaptosuccinic acid (DMSA) scan as first investigation of urinary tract infection. *Arch Dis Child* 1988;**63**:1320–5.

215. Verboven M, Ingels M, Delree M, Piepsz A. ^{99m}Tc-DMSA scintigraphy in acute urinary tract infection in children. *Pediatr Radiol* 1990;**20**:540–2.

216. Li Volti S, Di Bella D, Garozzo R, Di Fede GF, Mollica F. Imaging of urinary tract malformations: intravenous urography and/or kidney ultrasonography? *Child Nephrol Urol* 1991;**11**:96–9.

217. Whitear P, Shaw P, Gordon I. Comparison of ^{99m}Tc dimercaptosuccinic acid scans and intravenous urography in children. *Br J Radiol* 1990;**63**:438–43.

218. Wujanto R, Testa HJ, Shields RA, Prescott MC, Lawson RS, Cohen SJ. Assessment of renal function and scarring: is a DMSA scan always necessary? *Contrib Nephrol* 1987;**56**:250–5.

219. Dick PT, Childhood Urinary Tract Imaging Evaluation Study Investigators. Routine imaging does not benefit children with a first uncomplicated UTI. *Pediatr Res* 2002;**51**:213A. URL: http://www.pedresearch.org/

220. Stokland E. Renal imaging in children with urinary tract infection (bacteriuria) [Med.Dr Thesis]. Goteborg: Goteborgs Universitet; 1996.

221. Higueras W, Hall M, Janssen F, Verhas M. Rol de la scintigrafía renal ^{99m}Tc-DMSA en el diagnóstico de localización y evolución de la infección urinaria en niños. *Rev Méd Inst Peru Segur Soc* 1992;**1**:10–14.

222. Huicho L, Dávila M, Zerpa R. Infección urinaria: estudio prospectivo del valor diagnóstico de pruebas de tamizaje en niños preescolares hospitalizados. *Anales Facultad de Medicina (Peru)* 1995;**56**:24–9.

223. Van De Walle J-M. *Etude comparee de l'urographie intraveineuse et de l'echotomographie renale dans les pyelonephrites chroniques de l'enfant* [Th. Med]. Lille: Lille University; 1978.

224. Kuzmanovska D, Tasic V, Sahpazova E. [Detection of vesicoureteral reflux with radionuclide cystography]. *Srp Arh Celok Lek* 1996;**124**:78–81.

225. Varga F, Sasinka M, Kanik K. [Possibility of applying the TTC-test in non-catheterized urine for the diagnosis of urinary tract infections in girls]. *Ceskoslovenska Pediatrie* 1965;**20**:1080–4.

226. Tahirovic H, Pasic M. The value of nitrate test in detection of significant bacteriuria. *Jug Pediatr* 1984;**27**:39–41.

227. Nielsen EL, Petersen KE, Vejlsgaard R. [Screening for urinary tract infections in hospitalized children. Investigation of 458 children employing quantitative culture, dip-slides and microscopic examination]. *Ugeskr Laeger* 1977;**139**:2256–8.

228. Rantala H, Vorne M, Anttolainen I, Kurri J. [Detection of vesico-ureteral reflux with radionuclide or radiographic cystography in children]. *Duodecim* 1981;**97**:78–83.

229. Thabet L, Siala E, Kaabachi O, Ben Hamida A, Kechrid A. [The role of rapid tests in the screening of childhood urinary tract infection]. *Tunis Med* 2002;**80**:338–40.

230. Armengol CE, Hendley JO, Schlager TA. Should we abandon standard microscopy when screening for urinary tract infections in young children? *Pediatr Infect Dis J* 2001;**20**:1176–7.

231. Mutlib AE, Gerson RJ, Meunier PC, Haley PJ, Chen H, Gan LS, *et al*. The species-dependent metabolism of efavirenz produces a nephrotoxic glutathione conjugate in rats. *Toxicol Appl Pharmacol* 2000;**169**:102–13.

232. Urine collection from disposable nappies. *Lancet* 1991;**338**:1077–8.

233. Fujii R, Abe T, Tajima T, Terashima I, Meguro H, Sunakawa K, *et al*. Pediatric Study Group of Azithromycin. [Pharmacokinetic and clinical studies with azithromycin (capsule) in the pediatric field]. *Jpn J Antibiot* 1995;**48**:1074–92.

234. MacKenzie JR. DMSA – the new 'gold standard'. *Nucl Med Commun* 1990;**11**:725–6.

235. Larcombe J. Urinary tract infection in children. *BMJ* 1999;**319**:1173–5.

236. Smellie JM, Ransley PG, Normand IC, Prescod N, Edwards D. Development of new renal scars: a collaborative study. *BMJ* 1985;**290**:1957–60.

237. Coulthard MG, Lambert HJ, Keir MJ. Occurrence of renal scars in children after their first referral for urinary tract infection. *BMJ* 1997;**315**:918–19.

238. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using

Monte Carlo simulation. A practical approach. *Med Decis Making* 1985;**5**:157–77.

239. Critchfield GC, Willard KE. Probabilistic analysis of decision trees using Monte Carlo simulation. *Med Decis Making* 1986;**6**:85–92.

240. Briggs AH. Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics* 2000;**17**:479–500.

241. Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Econ* 1996;**5**:513–24.

242. Claxton K, Sculpher M, Drummond MF. A rational framework for decision making by the National Institute for Clinical Excellence. *Lancet* 2002;**360**:711–5.

243. Department of Health. *Reference costs 2002*. Leeds: Department of Health; 2002.

244. Claxton K, Ginnelly L, Sculpher M, Philips Z, Palmer S. A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment Programme. *Health Technol Assess* 2004;**8**(31).

245. Barry HC, Ebell MH, Hickner J. Evaluation of suspected urinary tract infection in ambulatory women. *J Fam Pract* 1997;**44**:49–60.

246. Fenwick EAL, Briggs AH, Hawke CI. Management of urinary tract infection in general practice: a cost-effectiveness analysis. *Br J Gen Pract* 2000;**50**:635–9.

247. Smellie JM, Barratt TM, Chantler C, Gordon I, Prescod NP, Ransley PG, *et al*. Medical versus surgical treatment in children with severe bilateral vesicoureteric reflux and bilateral nephropathy: a randomised trial. *Lancet* 2001;**357**:1329–33.

248. Panaretto K, Craig J, Knight J, Howman-Giles R, Sureshkumar P, Roy L. Risk factors for recurrent urinary tract infection in preschool children. *Journal of Paediatrics and Child Health* 1999;**35**:454–9.

249. Brown MJ, Sutton AJ, Bell PRF, Sayers RD. A meta-analysis of 50-years of ruptured abdominal aortic aneurysm repair. *Br J Surg* 2002;**89**:1–18.

250. Gill PE, Murray W, Wright MH. *Numerical linear algebra and optimization*. Redwood City, CA: Addision-Wesley; 1991.

251. British Medical Association, Royal Pharmaceutical Society of Great Britain. *British National Formulary*. 43rd ed. [CD-ROM]. London: British Medical Association and Royal Pharmaceutical Society of Great Britain; 2002.

252. Molyneux EM, Robson WJ. A dipstick test for urinary tract infections. *Journal of Accident and Emergency Medicine* 1995;**12**:191–3.

253. Netten A, Curtis L. *The unit costs of health and social care*. Canterbury: University of Kent, Personal Social Services Research Unit; 2003.

254. Smellie JM, Katz G, Gruneberg RN. Controlled trial of prophylactic treatment in childhood urinary-tract infection. *Lancet* 1978;**ii**:175–8.

255. Stansfeld JM. Duration of treatment for urinary tract infections in children. *BMJ* 1975;**3**:65–6.

256. Jodal U. The natural history of bacteriuria in childhood. *Infect Dis Clin North Am* 1987;**1**:713–29.

257. Alexander SR, Arbus GS, Butt KMH, Conley S, Fine RN, Greifer I, *et al*. The 1989 Report of the North American Pediatric Renal Transplant Cooperative Study. *Pediatr Nephrol* 1990;**4**:542–53.

258. Arant BS Jr. Vesicoureteric reflux and renal injury. *Am J Kidney Dis*. 1991;**17**:491–511.

259. Jacobson SH, Hansson S, Jakobsson B. Vesico-ureteric reflux: occurrence and long-term risks. *Acta Paediatr Suppl* 1999;**88**(431):22–30.

260. Mowatt G, Vale L, Perez P, Wyness L, Fraser C, MacLeod A, *et al*. *Systematic review of the effectiveness and cost-effectiveness of home versus hospital or satellite unit haemodialysis for people with end stage renal failure*. Aberdeen: Health Services Research Unit; 2002. URL: http://www.nice.org.uk/pdf/HvH_Assessment_report.pdf. Accessed 26 July 2006.

261. Chartered Institute of Public Finance and Accountancy. *Health Service Financial Database 2001*. London: CIPFA; 2001.

262. Office for National Statistics. *Census 2001: National report for England and Wales*. London: ONS; 2003.

263. Moyer VA, Elliott EJ, Davis RL, Davis RL, Gilbert R, Klassen T, *et al*. *Evidence based pediatrics and child health*. London: *BMJ* Books; 2000.

264. Williams GJ, Lee A, Craig JC. Long-term antibiotics for preventing recurrent urinary tract infection in children. *Cochrane Database Syst Rev* 2001;CD001534.

265. Stinnett A, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making*, 1998;**18** (Suppl):S68–80.

266. Deville WLJM, Bezemer PD, Bouter LM. Publications on diagnostic test evaluation in family medicine journals: an optimal search strategy. *J Clin Epidemiol* 2000;**53**:65–9. URL: http://www.sciencedirect.com/science/article/B6T84-3YKK9K4-9/2/9447c572c15251fe9734d10d7c9228ca

267. Dickersin K, Chan S, Chalmers T, Sacks H, Smith H Jr. Publication bias and clinical trials. *Control Clin Trials* 1987;**8**:343–53.

268. Begg C, Berlin J. Publication bias and dissemination of clinical research. *J Natl Cancer Inst* 1989;**81**:107–15.

269. Steurer J, Fischer J, Bachmann L, Koller M, ter Riet G. Communicating test accuracy terms to

practising physicians: a controlled study. *BMJ* 2002;**324**:824–6.

270. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PMM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;**58**(10):982–90.

271. Winberg J, Andersen HJ, Bergstrom T, Jacobsson B, Larson H, Lincoln K. Epidemiology of symptomatic urinary tract infection in childhood. *Acta Paediatr Scand Suppl* 1974;**252**:1–20.

272. Gordon I, Barkovics M, Pindoria S, Cole TJ, Woolf AS. Primary vesicoureteric reflux as a predictor of renal damage in children hospitalized with urinary tract infection: a systematic review and meta-analysis. *J Am Soc Nephrol* 2003;**14**:739–44.

273. Piepsz A, Tamminen-Mobius T, Reiners C, Heikkila J, Kivisaari A, Nilsson NJ, *et al*. International Reflux Study Group in Europe. Five-year study of medical or surgical treatment in children with severe vesico-ureteral reflux dimercaptosuccinic acid findings. *Eur J Pediatr* 1998;**157**:753–8.

274. Wennerstrom M, Hansson S, Jodal U, Stockland E. Disappearance of vesicoureteral reflux in children. *Arch Pediatr Adolesc Med* 1998;**152**:879–83.

275. Merrick MV, Notghi A, Chalmers N, Wilkinson AG, Uttley WS. Long-term follow up to determine the prognostic value of imaging after urinary tract infections. Part 2: Scarring. *Arch Dis Child* 1995;**72**:393–6.

276. Le Saux N, Pham B, Moher D. Evaluating the benefits of antimicrobial prophylaxis to prevent urinary tract infections in children: a systematic review. *CMAJ* 2000;**163**:523–9. URL: http://www.cmaj.ca/cgi/content/abstract/163/5/523

277. Bossuyt PMM, Reitsma JB, Bruns D, Gatsonis C, Glasziou P, Irwig L, *et al*. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003;**138**:W1–12.

278. McCormick A, Fleming D, Charlton J. *Morbidity statistics from general practice. Fourth national study 1991–1992.* London: HMSO; 1995.

279. Leigh, DA, Williams JD. Method for the detection of bacteriuria in large groups of patients. *J Clin Pathol* 1964;**174**:498.

280. Pryles CV. Percutaneous bladder aspiration and other methods of urine collection for bacteriologic study. *Pediatrics* 1965;**36**:128–31.

281. Barry AL, Smith PB, Turk M. Laboratory diagnosis of urinary tract infections. In Gavan TL, Editor. *Cumitech 2*. Washington DC: American Society Microbiology; 1975.

282 Nelson JD, Peters PC. Suprapubic aspiration of urine in premature and term infants. *Pediatrics* 1965;**36**:132–3.

283. Schersten B, Fritz H, Kohler L. Subnormal concentration of urinary glucose as a sign of urinary tract infection in children. *Acta Paediatr Scand* 1967;Suppl 177:54–6.

284. Lebowitz R, Olbling H, Parkkulainen K, Smellie J, Tamminen-Mobius T. International system of radiographic grading of vesicoureteric reflux. *Pediatr Radiol* 1985;**15**:105–9.

285. Dwoskin JY, Perlmutter AD. Vesicoureteral reflux in children: a computerized review. *J Urol* 1973;**109**:888–90.

286. Hofmann V, Beyer HJ. [Diagnostic value of sonography in vesico-ureteral reflux in infancy and childhood]. *Monatsschrift für Kinderheilkunde* 1985;**133**:834–9.

287. Thomas V, Shelokov A, Forland M. Antibody-coated bacteria in the urine and the site of urinary-tract infection. *N Engl J Med* 1974;**290**:588–90.

288. Shopfner CE. Cystourethrography: an evaluation of method. *AJR Am J Roentgenol* 1965;**95**:468–74.

289. Hellstrom M, Hjalmas K, Jacobsson B, Jodal U, Oden A. Normal ureteral diameter in infancy and childhood. *Acta Radiol Diagn* 1985;**26**:433–9.

290. Forbes PA, Drummond KN, Nogrady MB. Initial urinary tract infections. Observations in children without major radiologic abnormalities. *J Pediatr* 1969;**75**:187–92.

291. Monnen L, Smulders Y, van Lier H, *et al*. DDAVP test for assessment of renal concentrating capacity in infants and children. *Nephron* 1981;**29**:151.

292. Johnson CE, Debaz BP, Shurin PA, Debartolomeo R. Renal ultrasound evaluation of urinary-tract infections in children. *Pediatrics* 1986;**78**:871–8.

293. Toyonaga Y, Yamori K, Sakaguchi N, Sugita M, Kawamura K, Seo K. [Pharmacokinetic and clinical evaluation of cefpirome in the pediatric field]. *Jpn J Antibiot* 1991;**44**:184–206.

294. Lockwood TD, Bosmann HB. The use of urinary *N*-acetyl-beta-glucosaminidase in human renal toxicology. I. Partial biochemical characterization and excretion in humans and release from the isolated perfused rat kidney. *Toxicol Appl Pharmacol* 1979;**49**:323–36.

295. Horak E, Hopfer SM, Sunderman FW. Spectrophotometric assay for urinary *N*-acetyl-beta-*D*-glucosaminidase activity. *Clin Chem* 1981;**27**:1180–5.

296. Jodal U, Lindberg U, Lincoln K. Level diagnosis of symptomatic urinary tract infections in childhood. *Acta Paediatr Scand* 1975;**64**:201–8.

297. Zum Winkel K. *Nierendiagnostik mit Radioisotopen*. Stuttgart: Georg Thieme; 1964.

298. Montplaisir S, Cote PP, Martineau B, Roche AJ, Mongeau JG, Robitaille P. [Localization of the site of urinary infection in children by an investigation of antibody-covered bacteria]. *CMAJ* 1976;**115**: 1096–9.

299. Fernbach S, Maizels M, Conway J. Ultrasound grading of hydronephrosis: introduction to the system used by the Society for Fetal Urology. *Pediatr Radiol* 1993;**23**:478–80.

300. Heikel PE, Parkkulainen KV. Vesico-ureteric reflux in children: a classification and results of conservative treatment. *Ann Radiol* 1966;**9**:37–40.

301. International Reflux Study Committee. Medical versus surgical treatment of primary vesicoureteral reflux: a prospective international reflux study in children. *J Urol* 1981;**125**:277–83.

302. Alon U, Berant M, Pery M. Intravenous pyelography in children with urinary tract infection and vesicoureteral reflux. *Pediatrics* 1989;**83**:332–6.

303. Andrich MP, Majd M. Diagnostic imaging in the evaluation of the first urinary tract infection in infants and young children. *Pediatrics* 1992;**90**:436–41.

304. Kass EJ, Fink-Bennett D, Cacciarelli AA, Balon H, Pavlock S. The sensitivity of renal scintigraphy and sonography in detecting nonobstructive acute pyelonephritis. *J Urol* 1992;**148**:606–8.

305. McAfee JG. Radionuclide imaging in the assessment of primary chronic pyelonephritis. *Radiology* 1979;**133**:203–6.

306. Eggli DF, Tulchinsky M. Scintigraphic evaluation of pediatric urinary tract infection. *Semin Nucl Med* 1993;**23**:199–218.

307. Conway JJ. The role of scintigraphy in urinary tract infection. *Semin Nucl Med* 1988;**18**:308–19.

308. Wallin L, Bajc M. Typical technetium dimercaptosuccinic acid distribution patterns in acute pyelonephritis. *Acta Paediatr* 1993;**82**:1061–5.

309. Hodson CJ. The radiological contribution toward the diagnosis of chronic pyelonephritis. *Radiology* 1967;**88**:857–71.

310. Pylkkanen J, Vilska J, Koskimies O. The value of level diagnosis of childhood urinary tract infection in predicting renal injury. *Acta Paediatr Scand* 1981;**70**:879–83.

This version of HTA monograph volume 10, number 36 does not include the 209 pages of appendices. This is to save download time from the HTA website.

The printed version of this monograph also excludes the appendices.

View/download the appendices (1561 kbytes).

Health Technology Assessment Programme

Prioritisation Strategy Group

Members

HTA Commissioning Board

Members

Current and past membership details of all HTA 'committees' are available from the HTA website (www.hta.ac.uk)

Diagnostic Technologies & Screening Panel

Members

Chair,
Dr Ron Zimmern, Director of the Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

Ms Norma Armston, Lay Member, Bolton

Professor Max Bachmann Professor of Health Care Interfaces, Department of Health Policy and Practice, University of East Anglia

Professor Rudy Bilous Professor of Clinical Medicine & Consultant Physician, The Academic Centre, South Tees Hospitals NHS Trust

Dr Paul Cockcroft, Consultant Medical Microbiologist and Clinical Director of Pathology, Department of Clinical Microbiology, St Mary's Hospital, Portsmouth

Professor Adrian K Dixon, Professor of Radiology, University Department of Radiology, University of Cambridge Clinical School

Dr David Elliman, Consultant Paediatrician/ Hon. Senior Lecturer, Population Health Unit, Great Ormond St. Hospital, London

Professor Glyn Elwyn, Primary Medical Care Research Group, Swansea Clinical School, University of Wales Swansea

Mr Tam Fry, Honorary Chairman, Child Growth Foundation, London

Dr Jennifer J Kurinczuk, Consultant Clinical Epidemiologist, National Perinatal Epidemiology Unit, Oxford

Dr Susanne M Ludgate, Medical Director, Medicines & Healthcare Products Regulatory Agency, London

Professor William Rosenberg, Professor of Hepatology, Liver Research Group, University of Southampton

Dr Susan Schonfield, Consultant in Public Health, Specialised Services Commissioning North West London, Hillingdon Primary Care Trust

Dr Phil Shackley, Senior Lecturer in Health Economics, School of Population and Health Sciences, University of Newcastle upon Tyne

Dr Margaret Somerville, PMS Public Health Lead, Peninsula Medical School, University of Plymouth

Dr Graham Taylor, Scientific Director & Senior Lecturer, Regional DNA Laboratory, The Leeds Teaching Hospitals

Professor Lindsay Wilson Turnbull, Scientific Director, Centre for MR Investigations & YCR Professor of Radiology, University of Hull

Professor Martin J Whittle, Associate Dean for Education, Head of Department of Obstetrics and Gynaecology, University of Birmingham

Dr Dennis Wright, Consultant Biochemist & Clinical Director, Pathology & The Kennedy Galton Centre, Northwick Park & St Mark's Hospitals, Harrow

Pharmaceuticals Panel

Members

Chair,
Dr John Reynolds, Chair Division A, The John Radcliffe Hospital, Oxford Radcliffe Hospitals NHS Trust

Professor Tony Avery, Head of Division of Primary Care, School of Community Health Services, Division of General Practice, University of Nottingham

Ms Anne Baileff, Consultant Nurse in First Contact Care, Southampton City Primary Care Trust, University of Southampton

Professor Stirling Bryan, Professor of Health Economics, Health Services Management Centre, University of Birmingham

Mr Peter Cardy, Chief Executive, Macmillan Cancer Relief, London

Professor Imti Choonara, Professor in Child Health, Academic Division of Child Health, University of Nottingham

Dr Robin Ferner, Consultant Physician and Director, West Midlands Centre for Adverse Drug Reactions, City Hospital NHS Trust, Birmingham

Dr Karen A Fitzgerald, Consultant in Pharmaceutical Public Health, National Public Health Service for Wales, Cardiff

Mrs Sharon Hart, Head of DTB Publications, *Drug & Therapeutics Bulletin*, London

Dr Christine Hine, Consultant in Public Health Medicine, South Gloucestershire Primary Care Trust

Professor Stan Kaye, Cancer Research UK Professor of Medical Oncology, Section of Medicine, The Royal Marsden Hospital, Sutton

Ms Barbara Meredith, Lay Member, Epsom

Dr Andrew Prentice, Senior Lecturer and Consultant Obstetrician & Gynaecologist, Department of Obstetrics & Gynaecology, University of Cambridge

Dr Frances Rotblat, CPMP Delegate, Medicines & Healthcare Products Regulatory Agency, London

Professor Jan Scott, Professor of Psychological Treatments, Institute of Psychiatry, University of London

Mrs Katrina Simister, Assistant Director New Medicines, National Prescribing Centre, Liverpool

Dr Richard Tiner, Medical Director, Medical Department, Association of the British Pharmaceutical Industry, London

Dr Helen Williams, Consultant Microbiologist, Norfolk & Norwich University Hospital NHS Trust

Therapeutic Procedures Panel

Members

Chair,
Professor Bruce Campbell,
Consultant Vascular and General Surgeon, Department of Surgery, Royal Devon & Exeter Hospital

Dr Aileen Clarke,
Reader in Health Services Research, Public Health & Policy Research Unit, Barts & the London School of Medicine & Dentistry, London

Dr Matthew Cooke, Reader in A&E/Department of Health Advisor in A&E, Warwick Emergency Care and Rehabilitation, University of Warwick

Dr Carl E Counsell, Clinical Senior Lecturer in Neurology, Department of Medicine and Therapeutics, University of Aberdeen

Ms Amelia Curwen, Executive Director of Policy, Services and Research, Asthma UK, London

Professor Gene Feder, Professor of Primary Care R&D, Department of General Practice and Primary Care, Barts & the London, Queen Mary's School of Medicine and Dentistry, London

Professor Paul Gregg, Professor of Orthopaedic Surgical Science, Department of General Practice and Primary Care, South Tees Hospital NHS Trust, Middlesbrough

Ms Bec Hanley, Co-Director, TwoCan Associates, Hurstpierpoint

Ms Maryann L Hardy, Lecturer, Division of Radiography, University of Bradford

Professor Alan Horwich, Director of Clinical R&D, Academic Department of Radiology, The Institute of Cancer Research, London

Dr Simon de Lusignan, Senior Lecturer, Primary Care Informatics, Department of Community Health Sciences, St George's Hospital Medical School, London

Professor Neil McIntosh, Edward Clark Professor of Child Life & Health, Department of Child Life & Health, University of Edinburgh

Professor James Neilson, Professor of Obstetrics and Gynaecology, Department of Obstetrics and Gynaecology, University of Liverpool

Dr John C Pounsford, Consultant Physician, Directorate of Medical Services, North Bristol NHS Trust

Karen Roberts, Nurse Consultant, Queen Elizabeth Hospital, Gateshead

Dr Vimal Sharma, Consultant Psychiatrist/Hon. Senior Lecturer, Mental Health Resource Centre, Cheshire and Wirral Partnership NHS Trust, Wallasey

Dr L David Smith, Consultant Cardiologist, Royal Devon & Exeter Hospital

Professor Norman Waugh, Professor of Public Health, Department of Public Health, University of Aberdeen

Expert Advisory Network

Members

Professor Douglas Altman,
Director of CSM & Cancer Research UK Med Stat Gp, Centre for Statistics in Medicine, University of Oxford, Institute of Health Sciences, Headington, Oxford

Professor John Bond,
Director, Centre for Health Services Research, University of Newcastle upon Tyne, School of Population & Health Sciences, Newcastle upon Tyne

Mr Shaun Brogan,
Chief Executive, Ridgeway Primary Care Group, Aylesbury

Mrs Stella Burnside OBE,
Chief Executive, Office of the Chief Executive. Trust Headquarters, Altnagelvin Hospitals Health & Social Services Trust, Altnagelvin Area Hospital, Londonderry

Ms Tracy Bury,
Project Manager, World Confederation for Physical Therapy, London

Professor Iain T Cameron,
Professor of Obstetrics and Gynaecology and Head of the School of Medicine, University of Southampton

Dr Christine Clark,
Medical Writer & Consultant Pharmacist, Rossendale

Professor Collette Clifford,
Professor of Nursing & Head of Research, School of Health Sciences, University of Birmingham, Edgbaston, Birmingham

Professor Barry Cookson,
Director, Laboratory of Healthcare Associated Infection, Health Protection Agency, London

Professor Howard Cuckle,
Professor of Reproductive Epidemiology, Department of Paediatrics, Obstetrics & Gynaecology, University of Leeds

Dr Katherine Darton,
Information Unit, MIND – The Mental Health Charity, London

Professor Carol Dezateux,
Professor of Paediatric Epidemiology, London

Mr John Dunning,
Consultant Cardiothoracic Surgeon, Cardiothoracic Surgical Unit, Papworth Hospital NHS Trust, Cambridge

Mr Jonothan 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,
Professor of Community Rehabilitation, Institute of General Practice and Primary Care, University of Sheffield

Mr Leonard R Fenwick,
Chief Executive, Newcastle upon Tyne Hospitals NHS Trust

Professor David Field,
Professor of Neonatal Medicine, Child Health, The Leicester Royal Infirmary NHS Trust

Mrs Gillian Fletcher,
Antenatal Teacher & Tutor and President, National Childbirth Trust, Henfield

Professor Jayne Franklyn,
Professor of Medicine, Department of Medicine, University of Birmingham, Queen Elizabeth Hospital, Edgbaston, Birmingham

Ms Grace Gibbs,
Deputy Chief Executive, Director for Nursing, Midwifery & Clinical Support Services, West Middlesex University Hospital, Isleworth

Dr Neville Goodman,
Consultant Anaesthetist, Southmead Hospital, Bristol

Professor Alastair Gray,
Professor of Health Economics, Department of Public Health, University of Oxford

Professor Robert E Hawkins,
CRC Professor and Director of Medical Oncology, Christie CRC Research Centre, Christie Hospital NHS Trust, Manchester

Professor Allen Hutchinson,
Director of Public Health & Deputy Dean of ScHARR, Department of Public Health, University of Sheffield

Dr Duncan Keeley,
General Practitioner (Dr Burch & Ptnrs), The Health Centre, Thame

Dr Donna Lamping,
Research Degrees Programme Director & Reader in Psychology, Health Services Research Unit, London School of Hygiene and Tropical Medicine, London

Mr George Levvy,
Chief Executive, Motor Neurone Disease Association, Northampton

Professor James Lindesay,
Professor of Psychiatry for the Elderly, University of Leicester, Leicester General Hospital

Professor Julian Little,
Professor of Human Genome Epidemiology, Department of Epidemiology & Community Medicine, University of Ottawa

Professor Rajan Madhok,
Medical Director & Director of Public Health, Directorate of Clinical Strategy & Public Health, North & East Yorkshire & Northern Lincolnshire Health Authority, York

Professor David Mant,
Professor of General Practice, Department of Primary Care, University of Oxford

Professor Alexander Markham,
Director, Molecular Medicine Unit, St James's University Hospital, Leeds

Dr Chris McCall,
General Practitioner, The Hadleigh Practice, Castle Mullen

Professor Alistair McGuire,
Professor of Health Economics, London School of Economics

Dr Peter Moore,
Freelance Science Writer, Ashtead

Dr Sue Moss, Associate Director,
Cancer Screening Evaluation Unit, Institute of Cancer Research, Sutton

Mrs Julietta Patnick,
Director, NHS Cancer Screening Programmes, Sheffield

Professor Tim Peters,
Professor of Primary Care Health Services Research, Academic Unit of Primary Health Care, University of Bristol

Professor Chris Price,
Visiting Chair – Oxford, Clinical Research, Bayer Diagnostics Europe, Cirencester

Professor Peter Sandercock,
Professor of Medical Neurology, Department of Clinical Neurosciences, University of Edinburgh

Dr Eamonn Sheridan,
Consultant in Clinical Genetics, Genetics Department, St James's University Hospital, Leeds

Dr Ken Stein,
Senior Clinical Lecturer in Public Health, Director, Peninsula Technology Assessment Group, University of Exeter

Professor Sarah Stewart-Brown,
Professor of Public Health, University of Warwick, Division of Health in the Community Warwick Medical School, LWMS, Coventry

Professor Ala Szczepura,
Professor of Health Service Research, Centre for Health Services Studies, University of Warwick

Dr Ross Taylor,
Senior Lecturer, Department of General Practice and Primary Care, University of Aberdeen

Mrs Joan Webster,
Consumer member, HTA – Expert Advisory Network

Feedback

The HTA Programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (http://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.

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 23 8059 5639 Email: hta@hta.ac.uk
http://www.hta.ac.uk

ISSN 1366-5278