The Diagnosis of Urinary Tract infection in Young children (DUTY): a diagnostic prospective observational study to derive and validate a clinical algorithm for the diagnosis of urinary tract infection in children presenting to primary care with an acute illness

Alastair D Hay,[1]* Kate Birnie,[2] John Busby,[2] Brendan Delaney,[3] Harriet Downing,[1] Jan Dudley,[4] Stevo Durbaba,[5] Margaret Fletcher,[6,7] Kim Harman,[1] William Hollingworth,[2] Kerenza Hood,[8] Robin Howe,[9] Michael Lawton,[2] Catherine Lisles,[8] Paul Little,[10] Alasdair MacGowan,[11] Kathryn O'Brien,[12] Timothy Pickles,[8] Kate Rumsby,[10] Jonathan AC Sterne,[2] Emma Thomas-Jones,[8] Judith van der Voort,[13] Cherry-Ann Waldron,[8] Penny Whiting,[2] Mandy Wootton[9] and Christopher C Butler[12,14] on behalf of the DUTY team

[1]Centre for Academic Primary Care, National Institute for Health Research (NIHR) School of Primary Care Research, School of Social and Community Medicine, University of Bristol, Bristol, UK
[2]School of Social and Community Medicine, University of Bristol, Bristol, UK
[3]Department of Primary Care and Public Health Sciences, National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London, London, UK
[4]Bristol Royal Hospital for Children, University Hospitals Bristol NHS Foundation Trust, Bristol, UK
[5]Department of Primary Care and Public Health Sciences, Division of Health and Social Care Research, King's College London, London, UK
[6]Centre for Health and Clinical Research, University of the West of England, Bristol, UK
[7]South West Medicines for Children Local Research Network, University Hospitals Bristol NHS Foundation Trust, Bristol, UK
[8]South East Wales Trials Unit (SEWTU), Institute for Translation, Innovation, Methodology and Engagement, School of Medicine, Cardiff University, Cardiff, UK
[9]Specialist Antimicrobial Chemotherapy Unit, Public Health Wales Microbiology Cardiff, University Hospital Wales, Cardiff, UK
[10]Primary Care and Population Sciences Division, University of Southampton, Southampton, UK
[11]Southmead Hospital, North Bristol NHS Trust, Bristol, UK
[12]Cochrane Institute of Primary Care & Public Health, School of Medicine, Cardiff University, Cardiff, UK
[13]Department of Paediatrics and Child Health, University Hospital of Wales, Cardiff, UK
[14]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

*Corresponding author

Declared competing interests of authors: Paul Little is a member of the NIHR Journals Library Board, although he was not involved in the editorial processes for this report, and has provided consultancy work to Bayer Pharmaceuticals.

Published July 2016
DOI: 10.3310/hta20510

Scientific summary

The Diagnosis of UTI in Young Children (DUTY) study

Health Technology Assessment 2016; Vol. 20: No. 51
DOI: 10.3310/hta20510

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

It is not clear which young children with acute illness presenting to primary care should be investigated for urinary tract infection (UTI) and whether or not dipstick testing should be used to inform antibiotic treatment.

Objectives

The **D**iagnosis of **U**rinary **T**ract infection in **Y**oung children (DUTY) study objectives were to (1) develop algorithms, based on symptoms and signs to accurately identify children in whom a urine sample should be obtained; (2) assess whether or not dipstick urinalysis provides additional diagnostic information; (3) model algorithm cost-effectiveness; and (4) compare contamination rates between the clean-catch and 'Newcastle' nappy pad sampling methods.

Design

The DUTY study was a multicentre, prospective diagnostic cohort study that included: a comparison of reliability and accuracy between NHS laboratories and a single research laboratory to establish the optimum reference standard; the derivation and validation of algorithms to identify children warranting urine collection and establish the added value of urine dipstick testing for antibiotic treatment decisions; a health economic evaluation of validated algorithms; and a comparison of urinary contamination rates from clean-catch and nappy pad samples.

Setting

NHS 'first-point-of-contact' primary care sites, including 225 general practitioner (GP) surgeries, four children's emergency departments (CEDs) and four walk-in centres across England and Wales.

Participants

Children < 5 years of age and presenting with any acute (up to 28 days) illness and/or new urinary symptoms. Children were excluded if they were not constitutionally unwell; if they were known to have a neurogenic or surgically reconstructed bladder; if they were using a permanent or intermittent urinary catheter; if the main presenting problem was trauma; or if antibiotics had been taken in the previous 7 days. Clinicians were asked to recruit consecutive eligible children and, where this was not possible, to collect non-recruited children's age and sex.

Index tests and urine collection methods

Following consent, and blind to the reference standard, index tests (symptoms, signs and dipstick results) were recorded on a case report form. Symptoms included the child's medical history and parent-reported symptoms (graded as absent, mild, moderate or severe, when at their worst during the illness). Clinically qualified NHS staff (GPs, nurse practitioners and emergency department doctors/nurses) performed and

recorded examination findings, which included 'clinicians' global impression of the child's illness severity' and full respiratory and abdominal assessments. In total, 107 symptoms and signs were recorded and, preceding urine dipstick testing, clinicians recorded their opinion of UTI likelihood ('clinical diagnosis'), and their urine sampling and UTI treatment intentions ('clinical judgement'). Urine was collected by 'clean catch' (preferred) or nappy pad.

Methods to compare culture results from NHS and research laboratories

Microbiology reports from the NHS and research laboratories were classified based on extent and purity of growth and whether or not the species grown was a uropathogen (defined as a member of the *Enterobacteriaceae* group). For NHS laboratories, pure/predominant growths of uropathogens at $\geq 10^5$ colony-forming units (CFU) per ml were considered positive and research laboratory samples were considered positive if $\geq 10^5$ CFU/ml of a single uropathogen ('pure growth') or $\geq 10^5$ CFU/ml of a uropathogen with ≥ 3 $\log_{10}$ difference between the growth of this and the next species ('predominant growth') was present. This analysis included only those children with index test and both NHS and research laboratory results available. Agreement between laboratories was assessed using kappa statistics, with analyses additionally stratified by urine collection method (clean catch or nappy pad) and by age (0 to < 2, 2 to < 3 and 3 to < 5 years). Laboratory accuracy was investigated by comparing the strength of association between NHS and research laboratory UTI positivity and a small number of symptoms, signs and dipstick test results selected because they had previously been reported in the literature to be clearly related to UTI, albeit largely in emergency care settings. Those with the strongest associations and thought suitable for all ages and collection methods were urinary symptoms (pain/crying when passing urine, passing urine more often, changes in urine appearance); temperature ≥ 39 °C; and nitrite- or leucocyte-positive results from urine dipstick tests. We used logistic regression models to quantify associations of selected variables with laboratory UTI positivity and we plotted receiver operating characteristic (ROC) curves and used the area under ROC curve (AUROC) to quantify diagnostic accuracy.

Algorithm development methods

We sequentially evaluated selected index tests in two groups: parent-reported symptoms and clinician-reported signs (from the physical examination); and urine dipstick results. First, we selected those variables with either trend or heterogeneity univariable *p*-value < 0.01 for either collection method or when all samples were analysed together. Second, we derived models from among all the selected symptoms and signs, separately for nappy pad and clean-catch samples, using backwards stepwise selection and an exclusion criterion of heterogeneity *p*-value > 0.1. Third, we used backwards stepwise selection with the same exclusion *p*-value for models in which dipstick results were added to the previously selected symptoms and signs, to give models including symptoms, signs and dipstick results. For each model, we quantified diagnostic accuracy as the AUROC and compared this with the 'clinician-diagnosis' AUROC. We internally validated the models using the bootstrap procedure. As these coefficient-based models require relatively complex computation to estimate UTI probabilities, we also developed points-based models, the results of which are presented in the main report.

Health economic methods

We developed decision-analytic models using decision trees and Markov models to identify the optimal urine sampling strategy. We developed a 'clean-catch' model and a 'nappy pad' model to reflect the different symptoms and signs predictive of UTI in older and younger pre-school children and the different diagnostic accuracy of the two urine collection methods. The models synthesised data from the DUTY study and the wider literature to estimate the lifetime costs and health outcomes. We compared six urine sampling risk stratification strategies: three derived from the DUTY risk score reflecting high specificity (DUTY5%), intermediate (DUTY10%) and high sensitivity (DUTY20%) thresholds, one based on 'clinical judgement' and two boundary strategies (sample none, sample all). The model comprised three parts: short term (diagnosis and acute illness; up to 21 days), medium term (recurrent UTI; up to 3 years) and

long term (long-term sequelae; lifetime). Costs were estimated from a NHS perspective and included diagnostic costs and short- and long-term treatment costs. Health outcomes were expressed using quality-adjusted life-years (QALYs) and quality-adjusted life-days (QALDs). Net benefit statistics were used to compare the cost-effectiveness of urine sampling and testing strategies.

Contamination comparison methods

We selected a research laboratory contamination definition for its specificity (and previously termed 'frankly contaminated'), that is > 2 organisms all grown at $\geq 10^5$ CFU/ml. Univariable associations with contamination were estimated using logistic regression, grouped by clinical and dipstick variables and stratified by urine collection method. We selected variables with p-values < 0.01 for multivariable modelling.

Results: general

Between April 2010 and April 2012, 516 clinicians from 233 primary care sites screened 14,724 children for eligibility. Of these, 4390 were ineligible; 1276 declined participation; 1684 could not be recruited; 196 were subsequently excluded; and 15 withdrew. This left a recruited sample of 7163 children, of whom 50% were female; 49% were < 2 years old; 82% were white; and 26% had parents who reported education to degree level. A total of 6390 (89%) children provided a urine sample, 6241 (87%) by clean catch or nappy pad. Culture results were available from the NHS, research and both laboratories, respectively, for 5945 (83%), 5017 (70%) and 4828 (67%).

Results: comparison of NHS and research laboratory reliability and accuracy

This was conducted in the 4808 children with culture results from both laboratories originating from clean-catch or nappy pad samples and index test data. NHS laboratories reported UTI positivity in 6.6% (< 3 years) and 3.2% (≥ 3 years). The research laboratory reported positivity in 1.8% and 1.9% for the same age groups. Overall agreement [95% confidence interval (CI)] between the NHS and research laboratories was moderate (kappa = 0.36; 95% CI 0.29 to 0.43). Agreement was better for clean-catch samples (0.54; 95% CI 0.45 to 0.63) than for nappy pads (0.20; 95% CI 0.12 to 0.28). For clean-catch samples, agreement was similar in children aged ≥ 3 years (0.55; 95% CI 0.43 to 0.67) and < 3 years (0.52; 95% CI 0.37 to 0.67), which was better than for nappy pad samples in children aged < 3 years (0.20; 95% CI 0.12 to 0.28). Similar patterns were seen when comparisons were further stratified into age groups < 2 and ≥ 2 to < 3 years, suggesting that the lower reliability was attributable to nappy pad samples rather than the child's age. The AUROC (95% CI) for the six pre-specified symptoms, signs and dipstick test findings in clean-catch samples for the research and NHS laboratories were 0.86 (95% CI 0.79 to 0.92) and 0.75 (95% CI 0.69 to 0.80), respectively. The corresponding AUROCs for nappy pad samples were 0.79 (95% CI 0.70 to 0.88) and 0.65 (95% CI 0.61 to 0.70). As a result, we used the research laboratory culture result for the algorithm development reference standard and stratified algorithm development and health economic analyses by urine collection method.

Results: clean-catch diagnostic algorithm

Of the 2740 children providing clean-catch urines, 2.2% met the laboratory definition of UTI, 94% were aged ≥ 2 years and 54% were female. 'Clinical diagnosis' correctly identified 46.6% of the culture-positive children, with 94.7% specificity and an AUROC 0.77 (95% CI 0.71 to 0.83). Four symptoms (pain/crying while passing urine, smelly urine, history of UTI and absence of severe cough) and three signs (clinician-reported global impression of illness severity, abdominal tenderness and the absence of acute ear abnormality) were independently associated with UTI. The AUROC (95% CI; bootstrap-validated AUROC) for the symptoms and signs model was 0.89 (95% CI 0.85 to 0.95; validated 0.88), increasing to 0.93 (95% CI 0.90 to 0.97; validated 0.90) with leucocytes, nitrites and blood on dipstick testing.

Results: nappy pad diagnostic algorithm

Of the 2277 children providing nappy pad samples, 1.3% met the laboratory definition of UTI, of whom 82% were < 2 years old and 48% were female. 'Clinical diagnosis' correctly identified 13.3% of the culture-positive children, with 98.5% specificity and AUROC 0.63 (95% CI 0.53 to 0.72). Four symptoms (parent-reported smelly urine, darker urine, female sex and the absence of a nappy rash) and two dipstick results (leucocytes and nitrites) were independently associated with UTI. The AUROC (95% CI; bootstrap-validated AUROC) for the symptom model was 0.81 (95% CI 0.72 to 0.90, validated 0.78), increasing to 0.87 (95% CI 0.80 to 0.94, validated 0.82) with the dipstick findings.

Results: health economic analyses (clean-catch samples)

In clean-catch samples, the 'DUTY5%' (high specificity) threshold urine sampling strategy resulted in fewer urine samples being collected than risk stratification based on clinical judgement (4.8% vs. 9.2%), and slightly higher sensitivity (58.6% vs. 56.7%) and specificity (96.1% vs. 91.4%). The high specificity threshold of the clean-catch model was both cheaper and no less effective than clinical judgement in terms of QALDs in the short term and QALYs in the long term. The absolute difference in short-term net benefits among the three DUTY risk score thresholds evaluated was very small (ranging from £1088 in 'DUTY20%' to £1090 for 'DUTY5%'). The relatively low cost of urine sampling and an antibiotic prescription, the high rate of serendipitous antibiotic prescriptions, and the low prevalence of UTI within the DUTY population all contributed to the narrow range of estimated net benefits. Our results slightly favoured conservative (i.e. high specificity) urine sampling strategies, particularly for GPs concerned about the societal impact of bacterial resistance to antibiotics.

A greater percentage of children would be treated with immediate appropriate antibiotics according to urinary bacterial susceptibility (presumptive treatment 45.2%; dipstick testing 41.8%; laboratory testing 31.2%) and fewer would have delayed antibiotics (presumptive treatment 0.9%; dipstick testing 4.5%; laboratory testing 16.6%) if treatment was presumptive or dipstick-test guided. However, average sampling, testing and treatment costs were higher for the presumptive treatment and dipstick testing strategies (£1.18) than for the laboratory testing strategy (£1.10). Short-term net benefits were similar across all three testing and treatment strategies.

Results: health economic analyses (nappy pad samples)

In younger children, if urine was collected using a nappy pad, the distinction in cost-effectiveness between the DUTY risk score and clinical judgement was not clear-cut. This is due to the lower diagnostic value of the DUTY risk score in younger children, the higher contamination rates necessitating repeat urine sampling and the lower accuracy of NHS laboratory results in urine collected using nappy pads raising the possibility that a correct clinical diagnosis is overturned by an incorrect laboratory test result.

Using dipstick tests to determine treatment in children at intermediate risk of UTI slightly increased initial sampling, testing and treatment costs compared with a laboratory test-based treatment strategy (£1.18 vs. £1.10 per patient); however, it increased the proportion of children with UTI treated immediately with antibiotics (41.8% vs. 31.2%). There was no difference in short-term net benefits between dipstick- (£1090) and laboratory test- (£1090) based treatment strategies.

Results: comparison of contamination rates

'Frankly contaminated' urine was found in 1.8% and 12.2% of clean-catch and nappy pad samples, respectively, giving a risk ratio (95% CI) of 6.66 (95% CI 4.95 to 8.96; $p < 0.001$). Contamination and UTI was reported more often by the NHS than by research laboratories, especially for nappy pad samples. Probability of contamination was not increased by increasing the time taken for samples to arrive at laboratories or the presence of a nappy rash, but was increased by being female, home sampling, and increased frequency of nappy use.

Conclusions

Agreement of microbiological UTI diagnosis in routine NHS laboratories and a research laboratory was lower than expected and worse for nappy pads than for clean-catch samples. Accuracy was lower for NHS laboratory than research laboratories and for nappy pad than for clean-catch samples. Algorithms provided better diagnostic accuracy than 'clinical diagnosis' in identifying the children in whom urine collection was warranted and diagnostic accuracy was greater for clean-catch than nappy pad samples. Diagnosis and treatment based on a clean-catch coefficient model was more cost-effective than clinical judgement and although dipstick testing provided additional diagnostic utility, its benefit was offset by dipstick test costs.

Implications for health care

Primary care clinicians should prioritise the use of clean-catch sampling wherever possible. Parent-reported symptoms and clinical examination signs can be efficiently used to identify children who warrant clean-catch urine sampling. Dipstick testing has additional diagnostic value in deciding which children should receive antibiotic treatment, albeit at a higher cost than awaiting the laboratory result. NHS laboratories may wish to adopt the processing and reporting methods used by the research laboratory for paediatric urine samples and update national procedures accordingly.

Future research

Further research is needed to distinguish pathogens from contaminants when multiple urinary bacteria are found in significant concentrations. The impact of using the DUTY clean-catch coefficient algorithm on clinical behaviour and patient outcome in routine clinical practice, and the cost-effectiveness of presumptive versus dipstick versus laboratory-guided antibiotic treatment, should be assessed in randomised trials.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.058

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index.

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

Editorial contact: nihredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers 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.

HTA programme

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

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 08/66/01. The contractual start date was in January 2010. The draft report began editorial review in July 2013 and was accepted for publication in June 2014. 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 reviewers 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.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).