Clinical prediction rules for childhood UTIs: a cross-sectional study in ambulatory care

Boon, Hanne Ann; Verbakel, Jan Y; De Burghgraeve, Tine; Bruel, Ann Van den

DOI: <https://doi.org/10.3399/BJGPO.2021.0171>

To access the most recent version of this article, please click the DOI URL in the line above.

Received 17 September 2021

Revised 14 December 2021

Accepted 04 January 2022

© 2022 The Author(s). This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>). Published by BJGP Open. For editorial process and policies, see: <https://bjgpopen.org/authors/bjgp-open-editorial-process-and-policies>

When citing this article please include the DOI provided above.

Author Accepted Manuscript

This is an 'author accepted manuscript': a manuscript that has been accepted for publication in BJGP Open, but which has not yet undergone subediting, typesetting, or correction. Errors discovered and corrected during this process may materially alter the content of this manuscript, and the latest published version (the Version of Record) should be used in preference to any preceding versions

Figure 1

Figure 1 Flowchart of recruited children and urine samples obtained / %= percent, h= hours, n= total number of samples included in the analyses

1011x711mm (138 x 138 DPI)

Figure 2 ROC plot of the DUTY coefficient-based model for UTI in children below 5 years of age / Receiver operating characteristic (ROC) plot showing sensitivity versus 1-specificity. AUC= Area Under the ROC curve, 95%CI= 95% confidence intervals. The re-calibrated intercept and slope were -1.5731 and 0.1473 for the signs and symptoms coefficient-based model and -0.9252 and 0.2526 for the dipstick coefficient-based model, respectively. The calibration was weak for both scores (Supplementary Figure 1)

229x229mm (600 x 600 DPI)

Figure 4

Figure 4: Simulation for 1000 acutely ill children based on sampling and treatment strategies / Simulation for 1000 acutely ill children based on ERNIE4 study results (assuming UTI prevalence of 6%)(22)

a positive DUTY dipstick model ≥ 6 points (e.g. clinical model positive and ≥ 1 variable on the dipstick test positive)

b positive UTicalc dipstick model (probability of UTI $\geq 5\%$ in children with positive clinical model)

UTI= urinary tract infection, TP= true positives, FP= false positives, FN= false negatives, TN= true negatives

1011x711mm (138 x 138 DPI)

TITLE PAGE

Title: Clinical prediction rules for childhood UTIs: a cross-sectional study in ambulatory care

Authors: Hanne Ann Boon (MD)*¹, Jan Y Verbakel (MD, PhD)*^{1,2}, Tine De Burghgraeve (MSc, PhD)¹, Ann Van den Bruel (MD, PhD)¹

Affiliations:

¹EPI-Centre, Department of Public Health and Primary Care, KU Leuven, Kapucijnenvoer 7, 3000 Leuven, Belgium;

²Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe Observatory Quarter, Woodstock Rd, Oxford OX2 6GG, UK, ***joint first authorship**

Corresponding author: Ann Van den Bruel, EPI-Centre, Academic Centre for Primary Care, KU Leuven, Kapucijnenvoer 7, 3000 Leuven, Belgium, <ann.vandenbruel@kuleuven.be>, +32(0)16 37 66 21

Main text word count: 2658 **Abstract word count:** 249

Tables:2 **Figures:** 4 **Supplementary data:** 5 **References:** 24

Short Title: External validation of clinical prediction rules for childhood UTIs

Accepted Manuscript - BJGP Open - BJGPO.2021.0171

ABSTRACT

Background

Diagnosing childhood urinary tract infections (UTI) is challenging.

Aim

Validate clinical prediction rules (UTicalc, DUTY, Gorelick) for paediatric UTIs in primary care.

Design & setting

Post-hoc analysis of a cross-sectional study in 39 general practices and 2 emergency departments (Belgium, March 2019 to March 2020).

Method

Physicians recruited acutely ill children ≤ 18 years and sampled urine systematically for culture. Per rule, we performed an apparent validation; calculated sensitivities and specificities with 95%CI per threshold in the target group. For the DUTY coefficient-based algorithm, we performed a logistic calibration and calculated the Area Under the Curve with 95%CI.

Results

Of 834 children ≤ 18 years recruited, there were 297 children < 5 years. The UTicalc and Gorelick score had high to moderate sensitivity and low specificity (UTicalc $\geq 2\%$) 75%; and 16% respectively; Gorelick (≥ 2 variables) 91%; and 8%. In contrast, the DUTY score ≥ 5 points had low sensitivity (8%), but high specificity (99%). Urine samples would be obtained in 72% versus 38% (UTicalc), 92% versus 38% (Gorelick) or 1% versus 32% (DUTY) of children, compared to routine care. The

number of missed infections per score was 1/4 (UTlcalc), 2/23 (Gorelick) and 24/26 (DUTY). The UTlcalc+dipstick model had high sensitivity and specificity (100%; and 91%); resulting in no missed cases and 59% (95%CI 49-68%) of antibiotics prescribed inappropriately.

Conclusion

In this study, the UTlcalc and Gorelick score were useful for ruling out UTI but resulted in high urine sampling rates. The DUTY score had low sensitivity, meaning that 92% of UTIs would be missed.

KEYWORDS (MESH) Ambulatory Care, Child, Clinical Decision Rules, Primary Health Care, Urinary Tract Infections, Validation Studies [Publication Type]

HOW THIS FITS IN

- Three clinical prediction rules for urinary tract infections in children were previously developed to identify children requiring urine sampling.
- These prediction rules have not yet been validated in another setting and therefore, the robustness of these scores has not yet been established.
- In this dataset, the UTlcalc and Gorelick score had low specificity, leading to a high number of children in whom a urine sample would have to be obtained (72%, 92%) compared to standard clinical care (38%).
- The DUTY score had a low sensitivity, meaning that 92% of urinary tract infections would be missed.

INTRODUCTION

Urinary tract infections (UTI) occur in 3% to 14% of acutely ill children in ambulatory care.(1) Diagnosis is important because early antibiotic treatment can prevent progression to a severe illness and might prevent renal scarring.(2)

Ruling out UTIs is challenging because children with UTI have non-specific clinical features.(3) Additionally, guidelines are not very specific in describing which children require sampling.(4-6)

Urine sampling in all children is undesirable, as urine collection is difficult to combine with routine care, particularly in young children, and not cost-effective.(7-9)

Prediction rules may help in identifying children requiring urine sampling. Since missing a UTI is more problematic than oversampling, any clinical prediction rule should have high sensitivity because this minimizes the number of false negatives.

The target population for a prediction rule should ideally reflect the population seen in daily practice, a broad spectrum of acutely ill children.

In a recent systematic review,(10) we identified three prediction rules for UTI, all based on clinical features.(11-13)

(1) The DUTY score is a points-based algorithm, derived from a large cohort study (n=7163) in general practices in the UK, including acutely ill children <5 years. The score was validated internally through bootstrapping (Area Under the Receiver Operating Characteristic Curve analysis (AUC) 0.89, 95% confidence intervals (95%CI) 0.85-0.95).(11)

(2) The UTI calculator (UTIcalc) was derived using a nested case-control study in the USA with internal validation using a separate sample (AUC 0.81, 95%CI 0.72-

0.89) of febrile children <2 years evaluated for UTI at the emergency department (ED) (n=2070). The calculator is available online: <https://uticalc.pitt.edu/>.(12)

(3) Gorelick *et al.*(13) derived a prediction rule using a prospective cohort study in febrile girls <2 years (n=1469) at the ED (2000, AUC 0.76), with internal validation using a case-control study (2003, AUC 0.72).(14) This score is implemented in the American Academy of Pediatric (AAP) guidelines.(15)

To our knowledge, these prediction rules have not yet been validated externally and therefore the robustness of these scores has not yet been established. The population of the UTIcalc and Gorelick score consisted of children at higher risk of UTI, and therefore might be less applicable to general practice.

The aim of this study was to validate clinical prediction rules for UTI in primary care, in order to determine the accuracy of these scores.

METHOD

Study registration

This was a post-hoc analysis of the ERNIE4 study, of which the methods and results are reported elsewhere.(16) The ERNIE4 study was pre-registered (clinicaltrials.gov, NCT03835104) and is reported following the STARD 2015 guidelines.(16)

Study design

The ERNIE4 study was a multicenter, prospective cross-sectional study in 39 general practices and 2 EDs in Belgium (March 2019 to March 2020). Urine was sampled systematically and sent for analyses to one of four laboratories (A.M.L.,

C.M.A., Antwerp; AZ Maria Middelaers, Ghent; and Jessa ziekenhuis, Hasselt). For toilet-trained children, samples were obtained by midstream voiding at the time of study inclusion. For non-toilet trained children, physicians were asked to perform the Quick-Wee method, i.e. a direct catch of a first-stream sample (17); if unsuccessful, urine was collected using adhesive bags. In such a case, parents were asked to provide the sample within 24 hours after inclusion.

Participants

Children between 3 months to 18 years with an acute illness ≤ 10 days duration were eligible. Patients were not included if they presented with a traumatic injury, had a urinary catheter, were critically unstable, were referred to the hospital, or had been on immunosuppressive medication (≤ 30 days) or antibiotics (≤ 7 days).

Clinical prediction rules

We selected prediction rules for UTI in children based on clinical features, to determine which children require urine sampling. For each rule, we selected those patients from the ERNIE4 study based on the inclusion criteria of the rules' derivation studies and adapted the urine culture threshold accordingly, for optimal between-study comparison. The prediction rule variables are presented in Table 1 and a comparison with the ERNIE4 variables are provided in Supplementary Table 1.

- (1) For the DUTY models, we selected all children < 5 years. The authors derived a coefficient-based algorithm and a points-based algorithm (Supplementary Table 1). Urine sampling is recommended for children scoring ≥ 5 points on the points-

based model. Dipstick test results can be added to decide upon antibiotic treatment but the optimal threshold is unclear as none of the thresholds were cost-effective in the original study. In this study, we used ≥ 6 points as threshold for the DUTY score+dipstick model meaning that treatment would be initiated when the clinical model is positive and either blood, nitrite or leukocyte esterase (LE) are positive, in order to obtain a high sensitivity.

- (2) For the UTIcalc (version 3.0), we selected all febrile children ($\geq 38^{\circ}\text{C}$) < 2 years without urinary tract abnormalities. After urine has been obtained (UTI probability $\geq 2\%$), dipstick or microscopy results (=‘hematocytometer model’) can be added to the score to guide initiation of treatment (probability $\geq 5\%$).
- (3) For the score by Gorelick *et al.*, we selected all febrile children < 2 years. In contrary to the original study, we included both girls and boys, because UTIs occur frequently in boys < 1 year and implementation of a score for girls only did not seem practical.

Reference standard

In our study, UTI was defined as a single pathogen $\geq 10^5$ colony-forming units per milliliter (cfu/mL) on urine culture. (6) Contamination was defined as multiple pathogens or one pathogen $< 10^5$ cfu/mL. Samples were excluded if there was no result for culture or if the sample was received > 72 hours after inclusion in the laboratory.

For the DUTY models, the reference standard was one pathogen $\geq 10^5$ cfu/mL; for the Gorelick score, a pathogen $\geq 5 \times 10^4$ cfu/mL; and for the UTIcalc, a pathogen

$\geq 5 \times 10^4$ cfu/mL with pyuria, e.g. LE $\geq$ trace or white blood cells (WBC) (≥ 5 /high-power-field (hpf) or ≥ 10 /microliter(μ l)).

Data collection

At inclusion, clinical features were recorded for each child by the treating physician. We additionally collected 30-day follow-up information including laboratory or imaging results and hospital records, which were all conducted as part of routine care and not study-specific. We asked the treating physician to formulate a working hypothesis at the end of the initial consultation. In the analyses, we defined suspicion of UTI as a working hypothesis: 'UTI', 'cystitis' or 'pyelonephritis'.

All children underwent study-specific urine sampling. For each child, a study-specific urine culture was performed by lab technicians that were blinded to the index tests. Additionally, physicians were blinded for all study-specific test results, and therefore they were instructed to obtain an additional urine sample for clinical management if they deemed it necessary.

Statistical analyses

All statistical analyses were performed using R software version 4.0.4. We calculated sensitivities, specificities, positive and negative likelihood ratios with 95%CI for clinical features ('epiR' package).(18) When values for clinical features were missing, we considered them as normal.

For the DUTY models, an apparent validation (points-based model) and a logistic regression (coefficient-based model) were performed.(19) We calculated the AUC

with 95%CI ('pROC' package).(20) A calibration plot was made using the 'val.prob.ci.2' function ('CalibrationCurves' package).(21) For the UTIcalc and Gorelick score, the original regression coefficients were not available and therefore an apparent validation was performed, i.e. the model performance was assessed as is, without modifications.

As sensitivity analyses, we adapted urine culture thresholds to: one pathogen of $\geq 10^5$ cfu/mL following the European Association of Urology (EAU) guidelines (6) and lowered the threshold to 5×10^4 cfu/mL with pyuria (LE $\geq$ trace or ≥ 10 WBC/ μ l), following the AAP guidelines(4).

RESULTS

Study recruitment

There were 834 children recruited, of whom 643 children provided a urine sample. After exclusion of 68 samples because of arrival >72 hours after inclusion or no results for culture, 575 urine samples were available for analysis, of which 297 samples were from children <5 years (Figure 1). The median number of recruited children per practice was 13 (range 1 to 87).

Patient characteristics

Patient characteristics are listed in Table 2; per subgroup. The median age was 6 years (IQR 4-10) and 48% were girls. There were 51 children (9%) with a previous history of UTI of whom nine children had vesicoureteral reflux. Most children

presented with respiratory (81%) or abdominal symptoms (34%), while 8% presented with either frequency, dysuria or malodorous urine.

In addition to the study-specific urine sample that was obtained in all children, treating physicians requested a urine sample for clinical management in 151/575 children (26%). The sensitivity and specificity of a UTI working hypothesis was 7% (95%CI 2%-20%) and 95% (95%CI 93%-97%) respectively.

Samples and UTI prevalence

For the DUTY models, there were 297 children <5 years, of which 26 (9%) had a UTI (1 pathogen $\geq 10^5$ cfu/mL). For the UTIcalc, there were 96 febrile children <2 years and 4 of them (4%) had a UTI (pathogen $\geq 5 \times 10^4$ cfu/mL with pyuria). For the Gorelick, there were 100 febrile children <2 years of which 23 (23%) had a UTI (pathogen $\geq 5 \times 10^4$ cfu/mL).

Of all children with UTI, two children were hospitalized with pyelonephritis, which was confirmed on renal ultrasound in combination with elevated CRP-levels.

Obtaining urine samples

The diagnostic accuracies of the prediction rules are presented in Table 1, Figure 2 and Figure 3.

The UTIcalc ($\geq 2\%$) and Gorelick score (≥ 2 variables) had high to moderate sensitivity and low specificity (Table1). In contrast, the DUTY score (≥ 5 points) had low sensitivity but high specificity. Assuming a urine sample would be requested in children testing positive on the prediction rule, the urine sampling rate would be 92%

(Gorelick), 72% (UTlcalc) and 1% (DUTY), compared to 38%, 38% and 32% for standard care.

A flowchart for the management of 1000 simulated acutely ill children, assuming a UTI prevalence of 6%, (22), 2% (23) or 8% (1) are shown as Figure 4, Supplementary Figure 2 and 3. The number of missed infections for each of the scores is 15/60 (UTlcalc), 5/60 (Gorelick) and 55/60 (DUTY).

Initiation of empirical treatment

When urine has been obtained based on the clinical model (n=69), the UTlcalc+dipstick model had a sensitivity of 100% (95%CI 29%-100%) and specificity of 91% (95%CI 79%-97%), while the UTlcalc 'hemocytometer model' including WBC/ μ L had a sensitivity of 100% (95%CI 29%-100%) and specificity of 88% (95%CI 75%-95%). For the DUTY dipstick score ≥ 6 points, the sensitivity was 12% (95%CI 2%-30%) and specificity was 96% (95%CI 93%-98%).

Using the UTlcalc dipstick model, no UTIs would be missed and 59% (71/121, 95%CI 49-68%) of antibiotics would be given incorrectly, while using the dipstick test per standard care, one of three (n=6/23) UTIs would be missed and 72% (46/64, 95%CI 59-82%) of antibiotics would be given incorrectly (p=0.1073). Using the DUTY dipstick score, few children would be tested, as the sensitivity of the clinical model was low. Therefore, all UTIs (n=1/1) would be missed, and all prescriptions (n=1/1) would be given incorrectly (Figure 4).

Sensitivity analyses

Because the urine culture threshold is debatable in children,(24) and to allow optimal between-rule comparison in this study, we performed sensitivity analyses, comparing the diagnostic accuracies using identical criteria per rule:

At 1 pathogen $\geq 10^5$ cfu/mL,(6) sensitivities and specificities were 71% (95%CI 42%-92%) and 28% (95%CI 19%-39%) for the UTIcalc; 93% (95%CI 68%-100%) and 10% (95%CI 5%-17%) for the Gorelick score and 8% (95%CI 1%-25%) and 99% (95%CI 96%-100%) for the DUTY score.

Using a lower threshold, e.g. $\geq 5 \times 10^4$ cfu/mL with pyuria,(15) resulted in sensitivities and specificities of 75% (95%CI 19%-99%) and 16% (95%CI 9%-25%) for the UTIcalc; 100% (95%CI 40%-100%) and 8% (95%CI 4%-16%) for the Gorelick score and 12% (95%CI 0%-53%) and 98% (95%CI 96%-99%) for the DUTY score.

Because the Gorelick score was derived in girls, we calculated the sensitivity and specificity for girls separately (n=38), which was 100% (95%CI 48%-100%) and 6% (95%CI 1%-20%).

DISCUSSION

Summary

In our data, the sensitivities of the UTIcalc (75%) and Gorelick score (91%) were moderate to high, at low specificities (16% and 8%) leading to a very high number of children in whom a urine sample would have to be obtained (72% and 92%). This is much higher than the urine sampling rate per routine care measured in this study (38%). For the DUTY score, the sensitivity was low (8%), and the specificity was high (99%). The AUC showed little discriminatory value (0.55), meaning that urine

sampling would only be done in 1% of children, but at the expense of missing the majority of UTIs.

When urine has been obtained, the UTIcalc dipstick model appeared to be more sensitive than using the dipstick test as per routine care, based on very few cases.

Strengths and limitations

This was a cross-sectional study in primary care with systematic urine sampling. Because of the pragmatic nature of this study, the results will likely reflect real-life clinical practice.

Caution is needed in the interpretation, because the sample size was low, and included only 26 children with UTI. The study was terminated early, at the start of the SARS-CoV-2 pandemic. Face-to-face clinical care was heavily restricted and general practitioners indicated that recruiting was no longer possible. Because the population was no longer representative of a normal spectrum of children seen in daily practice, we decided to end study recruitment to obtain applicable results, but with much less precision. The calibration of the DUTY models was weak, most likely because there were too few UTI cases. Additionally, the original regression coefficients for the UTIcalc and Gorelick score were not available, meaning that the models were assessed without recalibration.

It is possible that some children in whom UTI was suspected; or urine collection was difficult were not included, due to the need of an additional urine sample for clinical management. This may have caused selection bias and an underestimation of the sensitivity.

In a minority of cases (32%), urine samples were obtained using adhesive bags. Although adhesive bags result in a high amount of contamination, we chose to avoid using invasive procedures in order to obtain real-world results that are applicable to clinical practice. Misallocation bias due to contaminated samples could have caused an underestimation of sensitivity and specificity (more false negatives and false positives).

Because the prediction rules were validated in primary care in Belgium; it is possible that our findings are not applicable to low-resource countries or ambulatory care settings where the population of children is different than our case-mix. Because we selected the target population that the original prediction rule was meant for, this limits the generalizability of our findings.

Comparison with existing literature

These results differ substantially from the derivation studies where sensitivities of 95%, 95% and 52% were found for the UTIcalc, Gorelick and DUTY score, respectively.

The population of the DUTY study was most comparable to our target population, however the score's sensitivity in our study was lower. Possible reasons are the low sample size of our study, over-fitting in the original study, selection bias in the current data, or random error.

The specificities of the UTIcalc and Gorelick score were substantially lower than in the original studies. This might have been caused by differences in study design (retrospective derivation),(12) or selection bias in our data. Additionally, the case-mix of children in these studies included a more severe spectrum in whom UTI was

suspected;(12, 13) the population was more diverse (25% Afro-American)(12) and there was a higher circumcision rate (19%).(12)

The urine sampling rate per normal care in this study (32-38%) was higher than in other studies performed in general practices,(23) which could have been caused by selection bias, a Hawthorne effect due to the nature of the study or the availability of an study-specific urine sample, per protocol.

Implications for practice

If future external validation and impact analysis confirm our findings, the UTIcalc or Gorelick score could be useful to decrease the number of missed UTIs in children in general practice. These simple metrics could be easily implemented to determine which children require urine sampling. One advantage of the UTIcalc is its integration of urine dipstick test results and therefore this score might be useful to avoid excessive use of urine culture, if proven sensitive.

Novel point-of-care tests for UTI should be assessed in combination with a clinical prediction rule prior to implementation because clinical suspicion of UTI is not sensitive enough to identify children for testing.

ADDITIONAL INFORMATION

Funding: The study was funded by the Research Foundation Flanders (Fonds Wetenschappelijk Onderzoek (FWO), Odysseus Program, grant number G0H8518N) and by a KU Leuven starting grant (grant number ERX-D5331-STG/18/008 [to HB]) The financial sponsor played no role in the design, execution, analysis and

interpretation of data, nor in the writing of the study or the decision to submit the manuscript.

Ethics approval: The protocol and study documents were approved by the Ethical Research Committee of UZ/ KU Leuven. The study was performed in accordance with the principals of the declaration of Helsinki.

Conflicts of interest: The authors declare they have no financial interests.

Availability of data and material: The data underlying this article will be shared on reasonable request to the corresponding author.

Consent to participate Formal written informed consent was obtained from the parent or guardian of each child.

Acknowledgments We thank the children, the general practitioners and pediatricians that participated in this study, with a special thanks to Dr Claudia Coolen en Dr. Jelle Degaeuwe for organizing the study at their department. Additionally, we are thankful for the expert team that evaluated unclear cases (Dr. Liselore De Rop, Dr. Steve Vandenbulck and Dr Axelle Vuylsteke). Finally, we would like to thank the Research Foundation Flanders (FWO) and KU Leuven for providing financial support for this study.

REFERENCES

1. Shaikh N, Morone NE, Bost JE, et al. Prevalence of urinary tract infection in childhood - A meta-analysis. *Pediatr Infect Dis J*. 2008;27(4):302-8.
2. Shaikh N, Mattoo TK, Keren R, Ivanova A, Cui G, Moxey-Mims M, et al. Early Antibiotic Treatment for Pediatric Febrile Urinary Tract Infection and Renal Scarring. *JAMA Pediatr*. 2016;170(9):848-54.
3. Boon HA, Van den Bruel A, Struyf T, Gillemot A, Bullens D, Verbakel JY. Clinical Features for the Diagnosis of Pediatric Urinary Tract Infections: Systematic Review and Meta-Analysis. *The Annals of Family Medicine*. 2021;19(5):437-46.
4. Roberts KB. Revised AAP guideline on UTI in febrile infants and young children. *Am Fam Physician*. 2012;86(10):940-6.
5. NICE guideline CG54: Urinary tract infection in under 16s Evidence reviews for UTI diagnosis in under 3 years 2017.
6. Stein R, Dogan HS, Hoebeke P, Kocvara R, Nijman RJ, Radmayr C, et al. Urinary tract infections in children: EAU/ESPU guidelines. *Eur Urol*. 2015;67(3):546-58.
7. Liaw LC, Nayar DM, Pedler SJ, Coulthard MG. Home collection of urine for culture from infants by three methods: survey of parents' preferences and bacterial contamination rates. *BMJ*. 2000;320(7245):1312-3.
8. Kaufman J, Knight AJ, Bryant PA, Babl FE, Dalziel K. Liquid gold: The cost-effectiveness of urine sample collection methods for young precontinent children. *Arch Dis Child*. 2019.
9. Hollingworth W, Busby J, Butler CC, O'Brien K, Sterne JA, Hood K, et al. The Diagnosis of Urinary Tract Infection in Young Children (DUTY) Study Clinical Rule: Economic Evaluation. *Value Health*. 2017;20(4):556-66.
10. Boon H. A., Van den Bruel A, Struyf T., Gillemot A., Bullens D, Verbakel JY. Clinical features to diagnose paediatric urinary tract infections in children. Accepted for publication in *Annals of Family Medicine*. 2021.
11. Hay AD, Birnie K, Busby J, Delaney B, Downing H, Dudley J, et al. 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. *Health Technol Assess*. 2016;20(51):1-294.
12. Shaikh N, Hoberman A, Hum SW, Alberty A, Muniz G, Kurs-Lasky M, et al. Development and Validation of a Calculator for Estimating the Probability of Urinary Tract Infection in Young Febrile Children. *JAMA Pediatr*. 2018;172(6):550-6.
13. Gorelick MH, Shaw KN. Clinical decision rule to identify febrile young girls at risk for urinary tract infection. *Arch Pediatr Adolesc Med*. 2000;154(4):386-90.
14. Gorelick MH, Hoberman A, Kearney D, Wald E, Shaw KN. Validation of a decision rule identifying febrile young girls at high risk for urinary tract infection. *Pediatr Emerg Care*. 2003;19(3):162-4.
15. Subcommittee On Urinary Tract Infection. Reaffirmation of AAP Clinical Practice Guideline: The Diagnosis and Management of the Initial Urinary Tract Infection in Febrile Infants and Young Children 2-24 Months of Age. *Pediatrics*. 2016;138(6).
16. Boon HA, De Burghgraeve T, Verbakel JY, Van den Bruel A. Point-of-care tests for pediatric urinary tract infections in general practice: a diagnostic accuracy study. *Fam Pract*. 2021.
17. Kaufman J, Fitzpatrick P, Tosif S, Hopper SM, Donath SM, Bryant PA, et al. Faster clean catch urine collection (Quick-Wee method) from infants: randomised controlled trial. *Bmj-British Medical Journal*. 2017;357.
18. Stevenson M. Package 'epiR' version 1.0-15 2020 [Available from: <https://cran.r-project.org/web/packages/epiR/epiR.pdf>].
19. Steyerberg EW. Clinical prediction models: A Practical Approach to Development, Validation, and Updating Switzerland: Springer; 2019. 558 p.

20. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.
21. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol. 2016;74:167-76.
22. O'Brien K, Edwards A, Hood K, Butler CC. Prevalence of urinary tract infection in acutely unwell children in general practice: a prospective study with systematic urine sampling. Br J Gen Pract. 2013;63(607):e156-64.
23. Hay AD, Birnie K, Busby J, Delaney B, Downing H, Dudley J, et al. 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. Health Technol Assess. 2016;20(51):1-197.
24. Tullus K. Low urinary bacterial counts: do they count? Pediatr Nephrol. 2016;31(2):171-4.

TABLES

Table 1 Diagnostic accuracies of clinical prediction rules for childhood urinary tract infections

	UTicalc (≥2%)	Gorelick score (≥2 variables)	DUTY (≥5 points) points-based algorithm
Number of UTIs /number of children included	4/96	23/100	26/297
Population	3m-2y with fever and no urinary tract abnormalities	3m-2y with fever	3m-5y with acute illness
Variables	Age <12 months; Fever ≥39°C; Non-African-American ethnicity; Female gender; Uncircumcised male; Fever without source*	Age <12 months, Caucasian, Fever ≥39°C; Fever ≥2 days; Fever without source**	Dysuria (2 points); Malodorous urine (2 points); History of UTI (1 point); Absence of severe cough (2 points); Severity of illness (2 points when >6 on a scale of 0-10).
Reference standard	1 or 2 pathogens >5x10 ⁴ cfu/mL and pyuria	1 or 2 pathogens >5x10 ⁴ cfu/mL	1 pathogen >10 ⁵ cfu/mL
Sensitivity/Specificity (%) (95%CI) Derivation study	95 (NR)/ 35 (NR)	95 (85-99) / 31 (28-34)	52 (39-64) / 95 (94-95)
Sensitivity/Specificity (%) (95%CI) ERNIE4 study	75 (19-99) / 16 (9-25)	91 (72-99) / 8 (3-16)	8 (1-25) / 99 (96-100)
Urine sampling rate following Prediction rule / Routine care (%) (95%CI)	72 (62-81) / 38 (28-48)	92 (85-97) / 38 (29-48)	1 (0.8-4) / 32 (26-37)
Number of missed infections Prediction rule / Routine care (%) (95%CI)	1/4 versus 2/4	2/23 versus 17/23	24/26 versus 16/26

UTI= urinary tract infection, UTicalc= UTI calculator, DUTY score= Diagnosis of Urinary Tract Infections in Children points-based model, 95%CI= 95% confidence intervals, m= months old, y= years old, cfu/mL= colony-forming units per milliliter, °C= Degrees Celsius

*was defined in the original study as no upper respiratory tract infection, no bronchiolitis, no pneumonia, no acute otitis media, no gastroenteritis, no meningitis and no viral syndrome, **was defined in the original study as discharge diagnoses: 'fever', 'fever without source' or 'viral infection'.

Table 2 Characteristics of acutely ill children

Characteristic	n=575 (≤18 years)	n=297 (DUTY sample)	n= 100 (Gorelick sample)	n=96 (UTlcalc sample)
Age (Median, IQR)	6.38 (3.97-10.25)	2.60 (1.12-3.75)	0.94 (0.56-1.49)	0.94 (0.56-1.49)
Girl	276 (48%)	134 (45%)	38 (38%)	37 (39%)
Boy	298 (52%)	163 (55%)	62 (62%)	59 (62%)
NA	1 (0%)	0 (0%)	0 (0%)	0 (0%)
circumcised boys	31 (10%)	12 (7%)	5 (8%)	5 (9%)
Fever: "yes"	415 (72%)	249 (84%)	100 (100%)	100 (100%)
Duration of illness (Median, IQR)	3.00 (1.00-4.00)	2.00 (1.00-3.00)	2.00 (1.00-3.00)	2.00 (1.00-3.00)
Dysuria	31 (5%)	11 (4%)	1 (1%)	1 (1%)
No dysuria	535 (93%)	282 (95%)	99 (99%)	95 (99%)
NA	9 (2%)	4 (1%)	0 (0%)	0 (0%)
Frequency	26 (4%)	9 (3%)	1 (1%)	1 (1%)
No frequency	539 (94%)	283 (95%)	99 (99%)	95 (99%)
NA	10 (2%)	5 (2%)	0 (0%)	0 (0%)
Malodorous urine	16 (3%)	9 (3%)	3 (3%)	3 (3%)
No malodorous urine	547 (95%)	284 (96%)	97 (97%)	93 (97%)
NA	12 (2%)	4 (1%)	0 (0%)	0 (0%)
Abdominal pain	121 (21%)	41 (14%)	5 (5%)	5 (5%)
No abdominal pain	450 (78%)	253 (85%)	95 (95%)	91 (95%)
NA	4 (1%)	3 (1%)	0 (0%)	0 (0%)
History of UTI	51 (9%)	23 (8%)	10 (10%)	9 (9%)
No history of UTI	514 (89%)	270 (91%)	90 (90%)	87 (91%)
NA	10 (2%)	4 (1%)	0 (0%)	0 (0%)
VUR	10 (2%)	5 (2%)	2 (2%)	2 (2%)
No VUR	555 (96%)	288 (97%)	98 (98%)	94 (98%)
NA	10 (2%)	4 (1%)	0 (0%)	0 (0%)

N= sample size, NA= missing value, IQR= Inter-Quartile Range, VUR= vesicoureteral reflux, UTI= urinary tract infection

FIGURE LEGENDS

Figure 1 Flowchart of recruited children and urine samples obtained

%= percent, h= hours, n= total number of samples included in the analyses

Figure 2 ROC plot of the DUTY coefficient-based model for UTI in children below 5 years of age

Receiver operating characteristic (ROC) plot showing sensitivity versus 1-specificity. AUC= Area Under the ROC curve, 95%CI= 95% confidence intervals. The re-calibrated intercept and slope were -1.5731 and 0.1473 for the signs and symptoms coefficient-based model and -0.9252 and 0.2526 for the dipstick coefficient-based model, respectively. The calibration was weak for both scores (Supplementary Figure 1)

Figure 3: ROC plot of the Gorelick score for UTI in children below 2 years of age

Receiver operating characteristic (ROC) plot showing sensitivity versus 1-specificity. var= variable, 95%CI= 95% confidence intervals

Figure 4: Simulation for 1000 acutely ill children based on sampling and treatment strategies

Simulation for 1000 acutely ill children based on ERNIE4 study results (assuming UTI prevalence of 6%)(22)

^a positive DUTY dipstick model ≥ 6 points (e.g. clinical model positive and ≥ 1 variable on the dipstick test positive)

^b positive UTIcalc dipstick model (probability of UTI $\geq 5\%$ in children with positive clinical model)

UTI= urinary tract infection, TP= true positives, FP= false positives, FN= false negatives, TN= true negatives