

# Clinical prediction rules for childhood urinary tract infections: a cross-sectional study in ambulatory care

Hanne Ann Boon<sup>1†</sup>, Jan Y Verbakel<sup>1,2†</sup>, Tine De Burghgraeve<sup>1</sup>, Ann Van den Bruel<sup>1\*</sup>

<sup>1</sup>EPI-Centre, Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium; <sup>2</sup>Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

## Abstract

**Background:** Diagnosing childhood urinary tract infections (UTIs) is challenging. Clinical prediction rules may help to identify children that require urine sampling. However, there is a lack of research to determine the accuracy of the scores in general practice.

**Aim:** To validate clinical prediction rules (UTI Calculator [UTICalc], A Diagnosis of Urinary Tract Infection in Young Children [DUTY], and Gorelick score) for paediatric UTIs in primary care.

**Design & setting:** Post-hoc analysis of a cross-sectional study in 39 general practices and two emergency departments (EDs). The study took place in Belgium from March 2019–March 2020.

**Method:** Physicians recruited acutely ill children aged  $\leq 18$  years and sampled urine systematically for culture. Per rule, an apparent validation was performed, and sensitivities and specificities were calculated with 95% confidence intervals (CIs) per threshold in the target group. For the DUTY coefficient-based algorithm, a logistic calibration was performed and the area under the receiver operating characteristic curve (AUC) was calculated with 95% CI.

**Results:** Of 834 children aged  $\leq 18$  years recruited, there were 297 children aged  $< 5$  years. The UTICalc and Gorelick score had high-to-moderate sensitivity and low specificity: UTICalc ( $\geq 2$ ) 75% and 16%, respectively; Gorelick ( $\geq 2$  variables) 91% and 8%, respectively. In contrast, the DUTY score  $\geq 5$  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 with routine care. The number of missed infections per score was 1/4 (UTICalc), 2/23 (Gorelick), and 24/26 (DUTY). The UTICalc + dipstick model had high sensitivity and specificity (100% and 91%), resulting in no missed cases and 59% (95% CI = 49% to 68%) of antibiotics prescribed inappropriately.

**Conclusion:** In this study, the UTICalc 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.

\*For correspondence: ann.vandenbruel@kuleuven.be

†These authors contributed equally to this work

**Competing interest:** The authors declare that no competing interests exist.

**Received:** 17 September 2021

**Accepted:** 04 January 2022

**Published:** 06 April 2022

©This article is Open Access: CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

**Author Keywords:** ambulatory care, child, clinical decision rules, primary health care, urinary tract infections, validation studies (publication type)

Copyright © 2022, The Authors; DOI:10.3399/BJGPO.2021.0171

## How this fits in

Three clinical prediction rules for UTIs 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 UTICalc and Gorelick score had low specificity, leading to a high number of children in whom a urine sample would have to be obtained (72% and 92%) compared with standard clinical care (38%). The DUTY score had a low sensitivity, meaning that 92% of UTIs would be missed.

## Introduction

UTIs occur in 3%–14% of acutely ill children in ambulatory care.<sup>1</sup> Diagnosis is important because early antibiotic treatment can prevent progression to a severe illness and might prevent renal scarring.<sup>2</sup>

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

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.<sup>7–9</sup> Prediction rules may help to identify children who require urine sampling. Since missing a UTI is more problematic than oversampling, any clinical prediction rule should have high sensitivity because this minimises the number of false negatives. The target population for a prediction rule should ideally reflect the population seen in daily practice, which is a broad spectrum of acutely ill children.

In a recent systematic review,<sup>10</sup> three prediction rules for UTI were identified, all based on clinical features:<sup>11–13</sup>

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 aged <5 years. The score was validated internally through bootstrapping (area under the receiver operating characteristic curve analysis [AUC] 0.89 [95% CI = 0.85 to 0.95]).<sup>11</sup>
2. UTICalc was derived using a nested case-control study in the US with internal validation using a separate sample (AUC 0.81 [95% CI = 0.72 to 0.89]) of febrile children aged <2 years evaluated for UTI at the ED ( $n = 2070$ ). The calculator is available online: <https://uticalc.pitt.edu>.<sup>12</sup>
3. Gorelick *et al*<sup>13</sup> derived a prediction rule using a prospective cohort study in febrile girls aged <2 years ( $n = 1469$ ) at the ED (AUC 0.76), with internal validation using a case-control study (AUC 0.72).<sup>14</sup> This score is implemented in the American Academy of Pediatrics (AAP) guidelines.<sup>15</sup>

To the authors' 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.<sup>16</sup> The ERNIE4 study is reported following the Standards for Reporting of Diagnostic Accuracy Studies (STARD) 2015 guidelines.<sup>16</sup>

### Study design

The ERNIE4 study was a multicentre, prospective cross-sectional study in 39 general practices and two EDs in Belgium (March 2019–March 2020). Urine was sampled systematically and sent for analyses to one of four laboratories (AML, CMA, Antwerp; AZ Maria Middelaes, 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, that is, a direct catch of a first-stream sample;<sup>17</sup> 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 aged between 3 months and 18 years with an acute illness  $\leq 10$  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 ( $\leq 30$  days) or antibiotics ( $\leq 7$  days).

## Clinical prediction rules

Prediction rules for UTI in children were selected based on clinical features, to determine which children required urine sampling. For each rule, those patients from the ERNIE4 study were selected based on the inclusion criteria of the rules' derivation studies and the urine culture threshold was adapted 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 S1.

1. For the DUTY models, all children aged <5 years were selected. The authors derived a coefficient-based algorithm and a points-based algorithm (see Supplementary Table S1). Urine sampling is recommended for children scoring  $\geq 5$  points on the points-based model. Dipstick test results can be added to decide on initiation of antibiotic treatment but the optimal threshold is unclear as none of the thresholds were cost-effective in the original study. In this study,  $\geq 6$  points was used as threshold for the DUTY score + dipstick model, meaning that treatment would be initiated when the clinical model was positive and either blood, nitrite, or leukocyte esterase (LE) were positive, in order to obtain a high sensitivity.
2. For UTICalc (version 3.0), all febrile children ( $\geq 38^\circ\text{C}$ ) aged <2 years without urinary tract abnormalities were selected. After urine had been obtained (UTI probability  $\geq 2\%$ ), dipstick or microscopy results (hemocytometer model) were added to the score to guide initiation of treatment (probability  $\geq 5\%$ ).
3. For the score by Gorelick *et al*, all febrile children aged <2 years were selected. In contrary to the original study, both girls and boys were included, because UTIs occur frequently in boys aged <1 year, and implementation of a score for only girls did not seem practical.

## Reference standard

In the study, UTI was defined as a single pathogen  $\geq 10^5$  colony-forming units per millilitre (CFU/ml) on urine culture.<sup>6</sup> 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; for example, LE  $\geq$ trace or white blood cells (WBC) ( $\geq 5$ /high-power field or  $\geq 10$ /microliter [ $\mu\text{l}$ ]).

## Data collection

At inclusion, clinical features were recorded for each child by the treating physician. Additionally, 30-day follow-up information was collected including laboratory or imaging results and hospital records, which were all conducted as part of routine care and not study-specific. The treating physician was asked to formulate a working hypothesis at the end of the initial consultation. In the analyses, suspicion of UTI was defined 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 laboratory 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 (version 4.0.4). Sensitivities, specificities, and positive and negative likelihood ratios were calculated with 95% CI for clinical features ('epiR' package).<sup>18</sup> When values for clinical features were missing, they were considered as normal.

For the DUTY models, an apparent validation (points-based model) and a logistic regression (coefficient-based model) were performed.<sup>19</sup> The AUC was calculated with 95% CI ('pROC' package).<sup>20</sup> A calibration plot was made using the 'val.prob.ci.2' function ('CalibrationCurves' package).<sup>21</sup> For the UTICalc and Gorelick score, the original regression coefficients were not available and, therefore, an apparent validation was performed, that is, the model performance was assessed as is, without modifications.

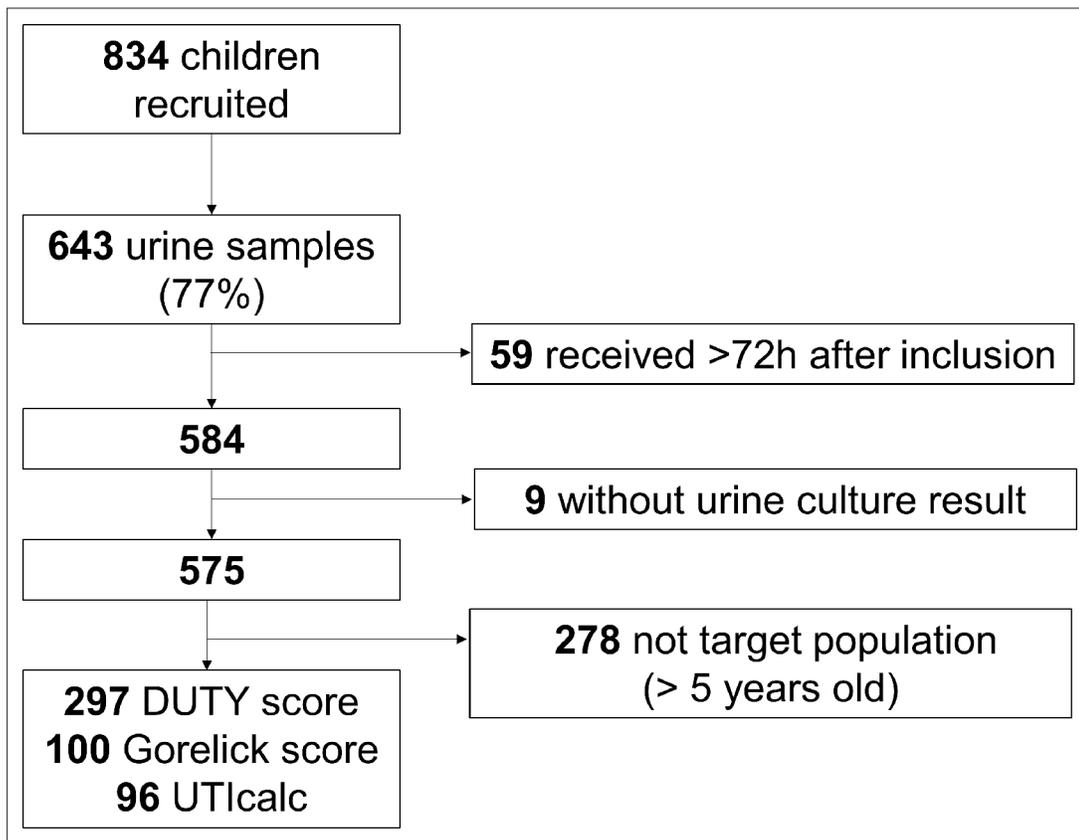
As sensitivity analyses, urine culture thresholds were adapted to: one pathogen of  $\geq 10^5$  CFU/ml following the European Association of Urology guidelines<sup>6</sup> and lowered the threshold to  $5 \times 10^4$  CFU/ml with pyuria (LE  $\geq$ trace or  $\geq 10$ WBC/ $\mu\text{l}$ ), following the AAP guidelines.<sup>4</sup>

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

Category	UTICalc ( $\geq 2\%$ )	Gorelick score ( $\geq 2$ variables)	DUTY ( $\geq 5$ points) points-based algorithm
Number of UTIs / number of children included, n	4 / 96	23 / 100	26 / 297
Population	3 months–2 years with fever and no urinary tract abnormalities	3 months–2 years with fever	3 months–5 years with acute illness
Variables	Age <12 months; fever $\geq 39^\circ\text{C}$ ; non-Black ethnicity; female sex; uncircumcised male; and fever without source <sup>a</sup>	Aged <12 months; White ethnicity; fever $\geq 39^\circ\text{C}$ ; fever $\geq 2$ days; fever without source <sup>b</sup>	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 $>5 \times 10^4$ CFU/ml and pyuria	1 or 2 pathogens $>5 \times 10^4$ CFU/ml	1 pathogen $>10^5$ CFU/ml
Sensitivity / specificity, derivation study, % (95% CI)	95 (NR) / 35 (NR)	95 (85 to 99) / 31 (28 to 34)	52 (39 to 64) / 95 (94 to 95)
Sensitivity / specificity, ERNIE4 study, % (95% CI)	75 (19 to 99) / 16 (9 to 25)	91 (72 to 99) / 8 (3 to 16)	8 (1 to 25) / 99 (96 to 100)
Urine sampling rate following prediction rule / routine care, % (95% CI)	72 (62 to 81) / 38 (28 to 48)	92 (85 to 97) / 38 (29 to 48)	1 (0.8 to 4) / 32 (26 to 37)
Number of missed infections, prediction rule versus routine care, n	1/4 versus 2/4	2/23 versus 17/23	24/26 versus 16/26

<sup>a</sup>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.

<sup>b</sup>Defined in the original study as discharge diagnoses: 'fever', 'fever without source', or 'viral infection'. CFU = colony-forming unit. DUTY = Diagnosis of Urinary Tract Infections in Children. NR = not reported. UTI = urinary tract infection. UTICalc = UTI Calculator.



**Figure 1** Flowchart of recruited children and urine samples obtained. DUTY = Diagnosis of Urinary Tract Infections in Children. h = hours. UTIcalc = UTI Calculator.

## 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 aged <5 years (**Figure 1**). The median number of recruited children per practice was 13 (range 1–87).

### 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 had vesicoureteral reflux. Most children presented with respiratory (81%) or abdominal features (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% to 20%) and 95% (95% CI = 93% to 97%), respectively.

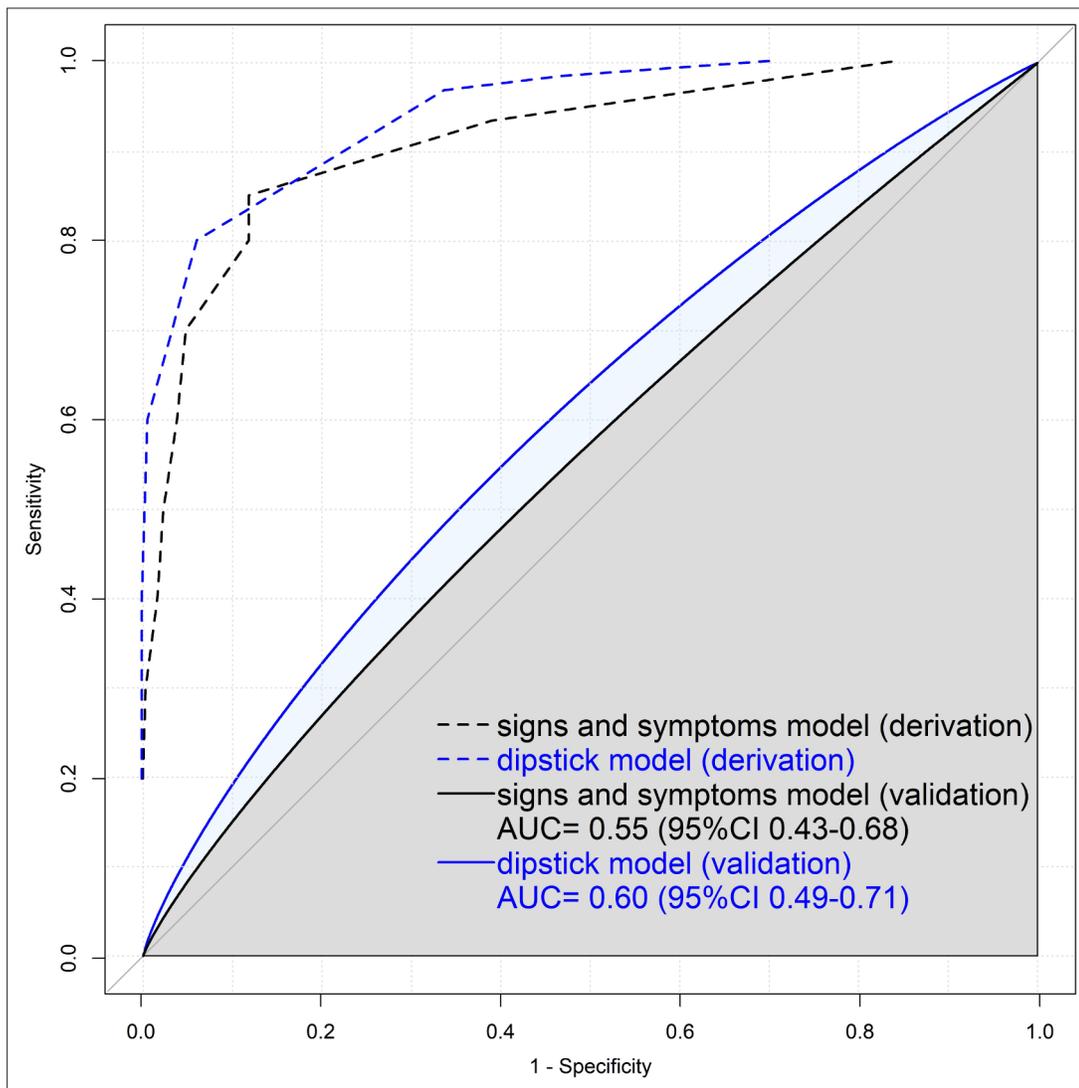
### Samples and UTI prevalence

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

**Table 2** Characteristics of acutely ill children

Characteristic	Aged ≤18 years, n = 575	DUTY sample, n = 297	Gorelick sample, n = 100	UTICalc sample, n = 96
Age, years, 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)
<b>Sex, n (%)</b>				
Girl	276 (48)	134 (45)	38 (38)	37 (39)
Boy	298 (52)	163 (55)	62 (62)	59 (61)
Missing	1 (0.2)	0 (0)	0 (0)	0 (0)
Circumcised boys, n (% of boys)	31 (10)	12 (7)	5 (8)	5 (8)
Fever: 'yes', n (%)	415 (72)	249 (84)	100 (100)	96 (100)
Duration of illness, days, 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)
<b>Dysuria, n (%)</b>				
Yes	31 (5)	11 (4)	1 (1)	1 (1)
No	535 (93)	282 (95)	99 (99)	95 (99)
Missing	9 (2)	4 (1)	0 (0)	0 (0)
<b>Frequency, n (%)</b>				
Yes	26 (5)	9 (3)	1 (1)	1 (1)
No	539 (94)	283 (95)	99 (99)	95 (99)
Missing	10 (2)	5 (2)	0 (0)	0 (0)
<b>Malodorous urine, n (%)</b>				
Yes	16 (3)	9 (3)	3 (3)	3 (3)
No	547 (95)	284 (96)	97 (97)	93 (97)
Missing	12 (2)	4 (1)	0 (0)	0 (0)
<b>Abdominal pain, n (%)</b>				
Yes	121 (21)	41 (14)	5 (5)	5 (5)
No	450 (78)	253 (85)	95 (95)	91 (95)
Missing	4 (1)	3 (1)	0 (0)	0 (0)
<b>History of UTI, n (%)</b>				
Yes	51 (9)	23 (8)	10 (10)	9 (9)
No	514 (89)	270 (91)	90 (90)	87 (91)
Missing	10 (2)	4 (1)	0 (0)	0 (0)
<b>VUR, n (%)</b>				
Yes	10 (2)	5 (2)	2 (2)	2 (2)
No	555 (97)	288 (97)	98 (98)	94 (98)
Missing	10 (2)	4 (1)	0 (0)	0 (0)

DUTY = Diagnosis of Urinary Tract Infections in Children. IQR = interquartile range. UTI = urinary tract infection. UTICalc = UTI Calculator. VUR = vesicoureteral reflux.



**Figure 2** Receiver operating characteristic (ROC) plot of the DUTY coefficient-based model for urinary tract infection in children aged <5 years. ROC plot showing sensitivity versus 1-specificity. The recalibrated 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 (see Supplementary Figure S1). AUC = area under the ROC curve.

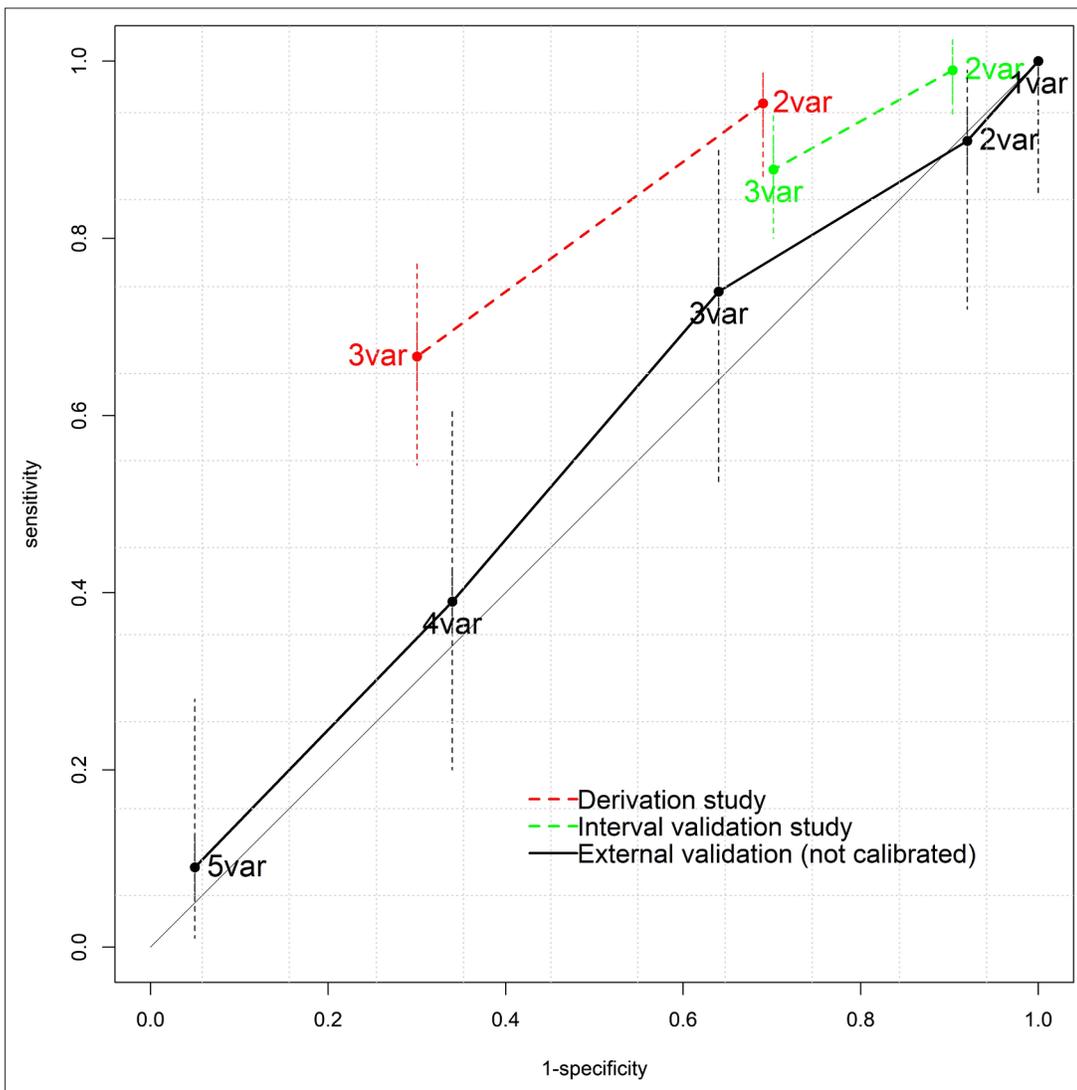
Of all children with UTI recruited in the ERNIE4 study, two children were hospitalised with pyelonephritis. Both children had elevated C-reactive protein levels (171 and 170 mg/L) at the general practice.

### Obtaining urine samples

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

The UTICalc ( $\geq 2\%$ ) and Gorelick score ( $\geq 2$  variables) had high-to-moderate sensitivity and low specificity (**Table 1**). In contrast, the DUTY score ( $\geq 5$  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% (UTICalc), and 1% (DUTY), compared with 38%, 38%, and 32% for standard care.

A flowchart for the management of 1000 simulated acutely ill children, assuming a UTI prevalence of 6%,<sup>22</sup> 2%,<sup>11</sup> or 8%<sup>1</sup> are shown as **Figure 4**, and Supplementary Figures S2 and S3. The number of missed infections for each of the scores is 15/60 (UTICalc), 5/60 (Gorelick), and 55/60 (DUTY).

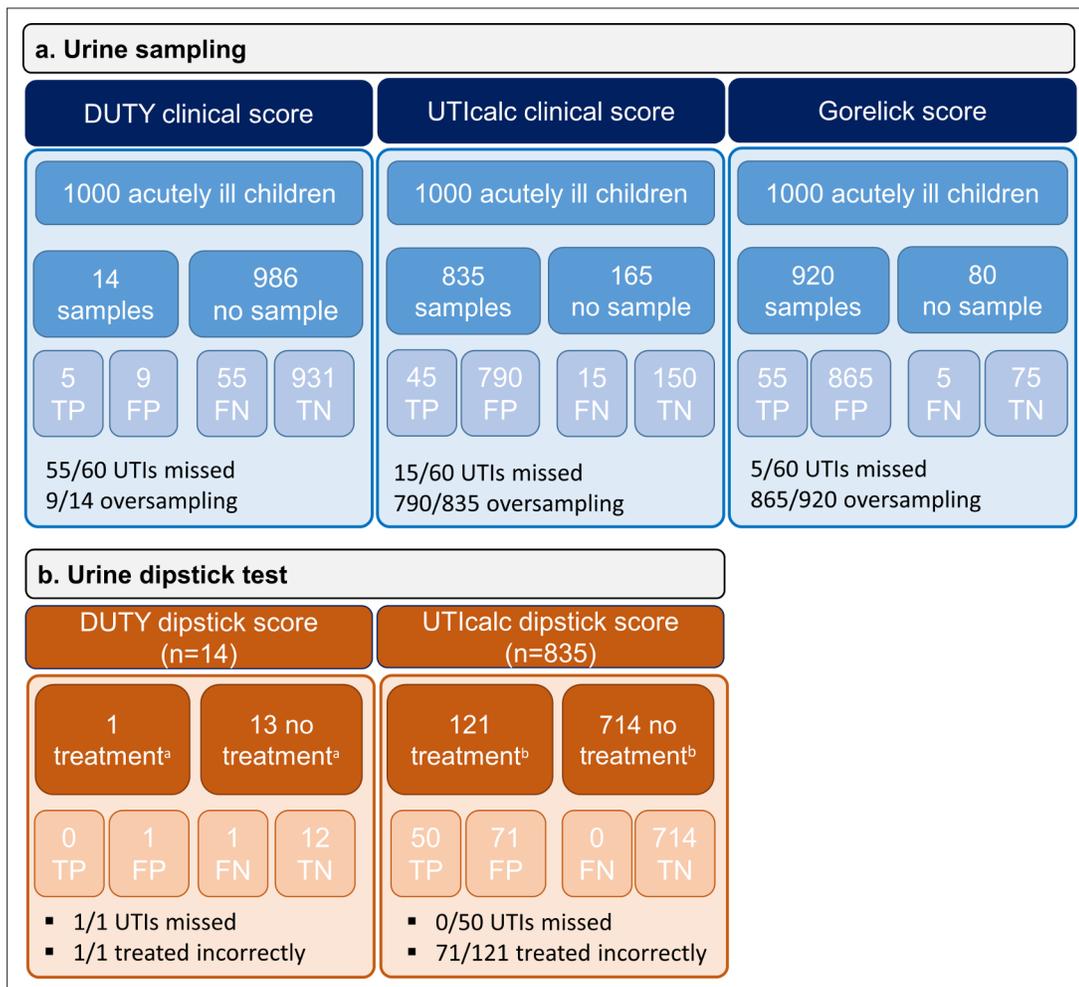


**Figure 3** Receiver operating characteristic (ROC) plot of the Gorelick score for urinary tract infection in children aged <2 years. ROC plot showing sensitivity versus 1-specificity. var = variable.

### Initiation of empirical treatment

When urine has been obtained based on the clinical model ( $n = 69$ ), the UTICalc dipstick model had a sensitivity of 100% (95% CI = 29% to 100%) and specificity of 91% (95% CI = 79% to 97%), while the UTICalc 'hemocytometer model', including WBC/ $\mu\text{l}$ , had a sensitivity of 100% (95% CI = 29% to 100%) and specificity of 88% (95% CI = 75% to 95%). For the DUTY dipstick score  $\geq 6$  points, the sensitivity was 12% (95% CI = 2% to 30%) and specificity was 96% (95% CI = 93% to 98%).

Using the UTICalc dipstick model, no UTIs would be missed and 59% ( $n = 71/121$  [95% CI = 49% to 68%]) of antibiotics would be given incorrectly (**Figure 4**), while using the dipstick test per standard care, one of three ( $n = 6/23$ ) UTIs would be missed and 72% ( $n = 46/64$  [95% CI = 59% to 82%]) of antibiotics would be given incorrectly ( $P = 0.1073$ ). Using the DUTY dipstick score  $\geq 6$  points, few children would be tested, as the sensitivity of the clinical model was low. Therefore, the only child with a UTI would have been missed, and the only prescription given for UTI would have been incorrect. (**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%).<sup>22</sup> <sup>a</sup>Positive DUTY dipstick model  $\geq 6$  points (for example, clinical model positive and  $\geq 1$  variable on the dipstick test positive). <sup>b</sup>Positive UTICalc dipstick model (probability of UTI  $\geq 5\%$  in children with positive clinical model). DUTY = Diagnosis of Urinary Tract Infections in Children. FN = false negatives. FP = false positives. TN = true negatives. TP = true positives. UTICalc = UTI Calculator. UTI = urinary tract infection.

### Sensitivity analyses

Because the urine culture threshold is debatable in children,<sup>23</sup> and to allow optimal between-rule comparison in this study, sensitivity analyses were performed comparing the diagnostic accuracies using identical criteria per rule.

At one pathogen  $\geq 10^5$  CFU/ml,<sup>6</sup> sensitivities and specificities were 71% (95% CI = 42% to 92%) and 28% (95% CI = 19% to 39%) for the UTICalc; 93% (95% CI = 68% to 100%) and 10% (95% CI = 5% to 17%) for the Gorelick score; and 8% (95% CI = 1% to 25%) and 99% (95% CI = 96% to 100%) for the DUTY score.

Using a lower threshold, for example,  $\geq 5 \times 10^4$  CFU/ml with pyuria,<sup>15</sup> resulted in sensitivities and specificities of 75% (95% CI = 19% to 99%) and 16% (95% CI = 9% to 25%) for the UTICalc; 100% (95% CI = 40% to 100%) and 8% (95% CI = 4% to 16%) for the Gorelick score; and 12% (95% CI = 0% to 53%) and 98% (95% CI = 96% to 99%) for the DUTY score.

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

## Discussion

### Summary

In the 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 GPs indicated that recruiting was no longer possible. Because the population was no longer representative of a normal spectrum of children seen in daily practice, it was 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, owing 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, it was chosen to avoid using invasive procedures in order to obtain real-world results that are applicable to clinical practice. Misallocation bias owing 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 the findings are not applicable to low-resource countries or ambulatory care settings where the population of children is different than the study's case-mix. Because the target population that the original prediction rule was meant for was selected, this limits the generalisability of the 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,<sup>12</sup> Gorelick,<sup>13</sup> and DUTY score,<sup>11</sup> respectively.

The population of the DUTY study was most comparable with the target population; however, the score's sensitivity in the study was lower. Possible reasons are the low sample size of the study, overfitting 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),<sup>12</sup> or selection bias in the data. Additionally, the case-mix of children in these studies included a more severe spectrum in whom UTI was suspected,<sup>12,13</sup> the population was more diverse (25% Black),<sup>12</sup> and there was a higher circumcision rate (19%).<sup>12</sup>

The urine sampling rate per normal care in this study (32%–38%) was higher than in other studies performed in general practices,<sup>11</sup> which could have been caused by selection bias, a Hawthorne effect owing to the nature of the study, or the availability of a study-specific urine sample, per protocol.

### Implications for practice

If future external validation and impact analysis confirm the 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 before implementation because clinical suspicion of UTI is not sensitive enough to identify children for testing.

### Funding

The study was funded by the FWO, Odysseus Programme (grant number: G0H8518N), and by a KU Leuven starting grant for Hanne Ann Boon (grant number: ERX-D5331-STG/18/008) 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.

### Ethical approval

The protocol and study documents were approved by the Ethical Research Committee of UZ/ KU Leuven (reference: S61991). The study was performed in accordance with the principals of the declaration of Helsinki.

### Trial registration number

The ERNIE4 study was pre-registered (clinicaltrials.gov: NCT03835104).

### Provenance

Freely submitted; externally peer reviewed.

### Patient consent

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

### Acknowledgements

The authors thank the children, GPs, and paediatricians who participated in this study, with a special thanks to Claudia Coolen and Jelle Degaeuwe for organising the study at their department. Additionally, the authors are thankful for the expert team that evaluated unclear cases (Liselore De Rop, Steve Vandenbulck, and Axelle Vuylsteke). Finally, the authors would like to thank the Research Foundation Flanders (Fonds Wetenschappelijk Onderzoek [FWO]) and KU Leuven for providing financial support for this study.

## References

1. Shaikh N, Morone NE, Bost JE, Farrell MH. Prevalence of urinary tract infection in childhood: a meta-analysis. *Pediatr Infect Dis J* 2008; **27**(4): 302–308. DOI: <https://doi.org/10.1097/INF.0b013e31815e4122>
2. Shaikh N, Mattoo TK, Keren R, et al. Early antibiotic treatment for pediatric febrile urinary tract infection and renal scarring. *JAMA Pediatr* 2016; **170**(9): 848–854. DOI: <https://doi.org/10.1001/jamapediatrics.2016.1181>
3. Boon HA, Van den Bruel A, Struyf T, et al. Clinical features for the diagnosis of pediatric urinary tract infections: systematic review and meta-analysis. *Ann Fam Med* 2021; **19**(5): 437–446. DOI: <https://doi.org/10.1370/afm.2684>
4. Roberts KB. Revised AAP guideline on UTI in febrile infants and young children. *Am Fam Physician* 2012; **86**(10): 940–946.
5. National Institute for Health and Care Excellence. Urinary tract infection in under 16s: diagnosis and management. CG54. 2018. <https://www.nice.org.uk/guidance/cg54> (accessed 23 Mar 2022).
6. Stein R, Dogan HS, Hoebeke P, et al. Urinary tract infections in children: EAU/ESPU guidelines. *Eur Urol* 2015; **67**(3): 546–558. DOI: <https://doi.org/10.1016/j.eururo.2014.11.007>
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): 1312–1313. DOI: <https://doi.org/10.1136/bmj.320.7245.1312>
8. Kaufman J, Knight AJ, Bryant PA, et al. Liquid gold: the cost-effectiveness of urine sample collection methods for young precontinent children. *Arch Dis Child* 2020; **105**(3): 253–259. DOI: <https://doi.org/10.1136/archdischild-2019-317561>
9. Hollingworth W, Busby J, Butler CC, et al. The diagnosis of urinary tract infection in young children (DUTY) study clinical rule: economic evaluation. *Value Health* 2017; **20**(4): 556–566. DOI: <https://doi.org/10.1016/j.jval.2017.01.003>
10. Boon HA, Van den Bruel A, Struyf T, et al. Clinical features for the diagnosis of pediatric urinary tract infections: systematic review and meta-analysis. *Ann Fam Med* 2021; **19**(5): 437–446. DOI: <https://doi.org/10.1370/afm.2684>

11. Hay AD, Birnie K, Busby 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. DOI: <https://doi.org/10.3310/hta20510>
12. Shaikh N, Hoberman A, Hum SW, 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): 550–556. DOI: <https://doi.org/10.1001/jamapediatrics.2018.0217>
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): 386–390. DOI: <https://doi.org/10.1001/archpedi.154.4.386>
14. Gorelick MH, Hoberman A, Kearney D, et al. Validation of a decision rule identifying febrile young girls at high risk for urinary tract infection. *Pediatr Emerg Care* 2003; **19**(3): 162–164. DOI: <https://doi.org/10.1097/01.pec.0000081238.98249.40>
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): e20163026. DOI: <https://doi.org/10.1542/peds.2016-3026>
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. DOI: <https://doi.org/10.1093/fampra/cmab118>
17. Kaufman J, Fitzpatrick P, Tosif S, et al. Faster clean catch urine collection (Quick-Wee method) from infants: randomised controlled trial. *BMJ* 2017; **357**: j1341. DOI: <https://doi.org/10.1136/bmj.j1341>
18. Stevenson M, Sergeant E. Package “epiR” version 2.0.41. 2022. <https://cran.r-project.org/web/packages/epiR/epiR.pdf> (accessed 23 Mar 2022).
19. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Cham: Springer International Publishing; 2019. DOI: <https://doi.org/10.1007/978-3-030-16399-0>
20. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011; **12**: 77. DOI: <https://doi.org/10.1186/1471-2105-12-77>
21. Van Calster B, Nieboer D, Vergouwe Y, et al. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol* 2016; **74**: 167–176. DOI: <https://doi.org/10.1016/j.jclinepi.2015.12.005>
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–e164. DOI: <https://doi.org/10.3399/bjgp13X663127>
23. Tullus K. Low urinary bacterial counts: do they count? *Pediatr Nephrol* 2016; **31**(2): 171–174. DOI: <https://doi.org/10.1007/s00467-015-3227-y>