Supplementary Online Content

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*. Published online April 16, 2018. doi:10.1001/jamapediatrics.2018.0217

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix. Additional Details Regarding Development of the Predictive Models

Training database construction

We obtained electronic records of consecutive children evaluated at the Emergency Department of Children's Hospital of Pittsburgh between 2007 and 2013 who, at the time of presentation, were less than 2 years of age, were febrile and who had no known abnormalities of the urinary tract. We excluded specimens obtained using a urine collection bag (Figure 1). The database included demographic information, presenting temperature and results of the urinalysis and urine culture; however, it did not include information about the child's presenting signs and symptoms. We included children with fever defined as a temperature of at least 38° C or a diagnosis code indicating fever (e.g., fever, febrile seizure). We restricted eligible visits to those with a single urinalysis (UA) and a single urine culture, each obtained within 3 hours of the other. We excluded children with known abnormalities of the urinary tract (e.g., spina bifida, neurogenic bladder). When a child had more than one eligible visit during the study period, we randomly selected a single episode.

Data abstraction

We abstracted information on the following predictors: duration of fever in hours; maximum temperature (reported by parent or measured in the ED); other source for fever (yes/no; coded yes if any of the following diagnoses/symptoms were present: acute otitis media, upper respiratory tract infection [any cough or congestion], gastroenteritis, pneumonia, meningitis, bronchiolitis, viral syndrome), foul smelling urine (yes/no), vomiting (yes/no); diarrhea (yes/no); and abdominal tenderness on exam (yes/no). We considered symptoms as being present only if they were noted within 24 hours of the ED visit. Except for circumcision status, when a sign or symptom was not mentioned in the visit notes, we presumed it to be absent (for circumcision, we recorded this as "unknown").

Model building

For each model, we started with a "baseline" model in which we included variables identified in previous studies as predictors of UTI. We then tested whether dropping or adding any variables were significant statistically (using the likelihood ratio test) and/or clinically (considering changes in the area under the receiver operating characteristic curve and, at various cutoffs, the false positive and false negative rates and net reclassification improvement).[1] The net reclassification index (NRI) represents a measure for evaluating incremental gain in prediction performance that results from adding a marker to a set of baseline predictors. Specifically, the NRI is the sum of 1) the proportion of subjects with UTI who are correctly moved into higher predicted risk categories with addition of the variable(s) of interest minus the proportion of children with UTI who are incorrectly moved into lower risk categories, and 2) the proportion of subjects without UTI who are correctly moved into lower predicted risk categories with addition of the variable(s) of interest minus the proportion of children without UTI who are incorrectly moved into higher risk categories.

The baseline "Clinical Model" included 8 previously identified risk factors of UTI:[2] age in months (continuous), race (not black vs. black), sex, circumcision status, fever duration (continuous, hours), maximum temperature in Celsius (continuous), other source of fever (no other source vs. other source including meningitis, pneumonia, AOM, gastroenteritis, URI without AOM, bronchiolitis, viral syndrome), and history of UTI (yes vs. no). Additional variables considered were abdominal tenderness on exam (yes vs. no), diarrhea (yes vs. no), vomiting (yes vs. no), and foul-smelling urine (yes vs. no). Because the probability of UTI was similar in females and uncircumcised males, we combined them into one group. Likewise, we combined "unequivocal source of fever" and "possible source of fever". We dropped "duration of fever" and "history of UTI" from the Clinical Model because dropping them decreased the predictive ability of the Clinical Model only marginally (AUC decreased by ~1% when both were dropped; NRI suggested little incremental value in retaining these variables). In contrast, dropping age, other source of fever, or sex/circumcision decreased the AUC of the model by 2 .0, 2.1 and 6.7 percent respectively. With duration and history removed from the model, dropping either race or maximum temperature decreased the AUC of the model by 1.6% and 1.9%, respectively. Dichotomizing temperature and age (using cutoffs previously established in the literature) did not change the accuracy of the model appreciably (AUC decreased by <1%). We did not add any of the 4 other candidate variables (abdominal tenderness, diarrhea, vomiting, and foul-smelling urine), because adding them did not improve the model significantly (AUC decreased by <1% and NRI suggested little incremental value in adding these tests). SAS version 9.4 and STATA 14 was used for all analyses.

The baseline "Dipstick Model" included the variables from our final "Clinical Model" plus leukocyte esterase (none, trace, 1+, 2+, 3+) and nitrite (present vs. absent). In contrast to a recent study,[3] the addition of specific gravity did not significantly improve the accuracy of the Dipstick Model. The baseline "Dipstick + Gram

Stain Model" included the variables from our final Dipstick Model plus the results of the Gram stain (positive vs. negative). The presence of any organism was considered a positive Gram-stained smear. No other variables were considered. The baseline "Hemocytometer Model" included the variables from our final Dipstick Model plus WBC/mm^3 (continuous variable). No other variables were considered. The baseline "Enhanced Urinalysis Model" included the variables from our final Hemocytometer Model plus urine Gram stain results. No other variables were considered.

Three points regarding development of the models warrant further explanation. First, to determine the accuracy of models that included laboratory tests, we excluded children in whom a urine sample was unlikely to have been obtained (i.e., we excluded those with a low pre-test probability on the Clinical Model, i.e., <2% probability). Second, because the clinical and laboratory features of UTI do not differ in children <2 months of age, these children were included in the training database. Third, because we expected the probability of disease to vary according to clinical variables even in children with the same laboratory test profile, we included variables from our final Clinical Model in all models that included laboratory tests (Dipstick, Dipstick + Gram-Stain, Hemocytometer, and Enhanced Urinalysis Models). To test whether this presumption was true, we tested versions of the latter 4 models with and without clinical variables.

Multilevel Likelihood ratios

Having developed the models in a random sample of children presenting with fever (in whom the prevalence of UTI was approximately 6%), we would expect them to perform well when applied to a similar population of children. However, to allow generalization of our findings to populations with a substantially different baseline pre-test probability of UTI[4] we also report multilevel likelihood ratios (eTable 3).

eTable 1. Accuracy of Symptoms in the Diagnosis of Urinary Tract Infection (UTI) From the Present Study Compared to Previous Studies

Finding	Diagnosis[a]		Likelihood ratio in present study (Training database)		Likelihood ratio in previous studies[b]	
	UTI (n=542)	No UTI (n=1144)	Positive likelihood ratio (CI)	Negative likelihood ratio (CI)	Positive likelihood ratio (CI)	Negative likelihood ratio (CI)
Age <12 months	431	798	1.14 (1.08, 1.21)	0.68 (0.56, 0.82)	--	--
Female	483	733	1.39 (1.32, 1.47)	0.30 (0.24, 0.39)	--	--
Non-black race	461	786	1.24 (1.17, 1.30)	0.47 (0.37, 0.58)	1.4 (1.1, 1.8)	0.52 (0.29, 0.73)
Fever ≥48 hours	238	327	1.52 (1.33, 1.73)	0.78 (0.72, 0.85)	1.3 (0.8, 1.9)	0.95 (0.85, 1.06)
No other source of fever	443	602	1.55 (1.45, 1.66)	0.39 (0.32, 0.47)	1.4 (1.1, 1.8)	0.69 (0.55, 0.80)
Max reported temperature ≥39° C	336	455	1.41 (1.29, 1.55)	0.65 (0.57, 0.74)	1.4 (1.2, 1.7)	0.78 (0.65, 0.81)
Foul smelling urine (>slight)	42	12	7.39 (3.92, 13.92)	0.93 (0.91, 0.96)	5.1 (3.7, 6.9)[5]	0.59 (0.46, 0.75)[5]
History of UTI	35	28	2.64 (1.62, 4.29)	0.96 (0.94, 0.98)	2.9 (1.2, 7.1)	0.95 (0.89, 1.02)
Uncircumcised	31	42	4.94 (3.45, 7.07)	0.45 (0.32, 0.63)	2.8 (1.9, 4.3)	0.33 (0.18, 0.63)
Vomiting	142	291	1.03 (0.87, 1.23)	0.99 (0.93, 1.05)	0.89 (0.43, 1.25)	1.07 (0.88, 1.21)
Diarrhea	76	188	0.85 (0.67, 1.09)	1.03 (0.99, 1.07)	0.64 (0.32, 1.26)[6]	1.3 (1.0, 1.7)[6]
Abdominal tenderness on exam	6	16	0.79 (0.31, 2.01)	1.00 (0.99, 1.01)	--	--
Leukocyte esterase (≥trace)	504	82	12.9 (10.5, 16.0)[c]	0.07 (0.05, 0.10)[c]	5.5 (4.1, 7.3)[7]	0.26 (0.18, 0.36)[7]
Positive nitrite test	200	26	16.1 (10.9, 24.0)	0.65 (0.61, 0.69)	15.9 (10.7, 23.7)[7]	0.51 (0.43, 0.60)[7]
≥10 WBC/mm^3	383	77	10.1 (8.11, 12.5)[c]	0.11 (0.08, 0.15)[c]	5.9 (4.1, 8.5)[7]	0.27 (0.20, 0.37)[7]

CI = 95% confidence interval; -- = could not find likelihood ratios in the literature

[a]Number of children

[b]Unless otherwise indicated, used values from one meta-analysis[2]

[c]The accuracy values (sensitivity, specificity, and likelihood ratios) reported for the leukocyte esterase and urine microscopy tests (WBC/mm^3) are artificially elevated because, in the definition of UTI currently endorsed by the AAP, these tests were required to be positive in patients with UTI. In other words, the tests being evaluated were also part of the gold standard used to evaluate them. This type of bias, known as incorporation bias, tends to artificially inflate both the sensitivity and specificity of the tests being evaluated.[8] Likewise the accuracy values for models that include these tests (reported in Table 2 and eTable 2) are also artificially high. Assignment to risk categories by UTICalc, however, was seldom affected because changes in the predicted probabilities rarely crossed the cutoffs selected.

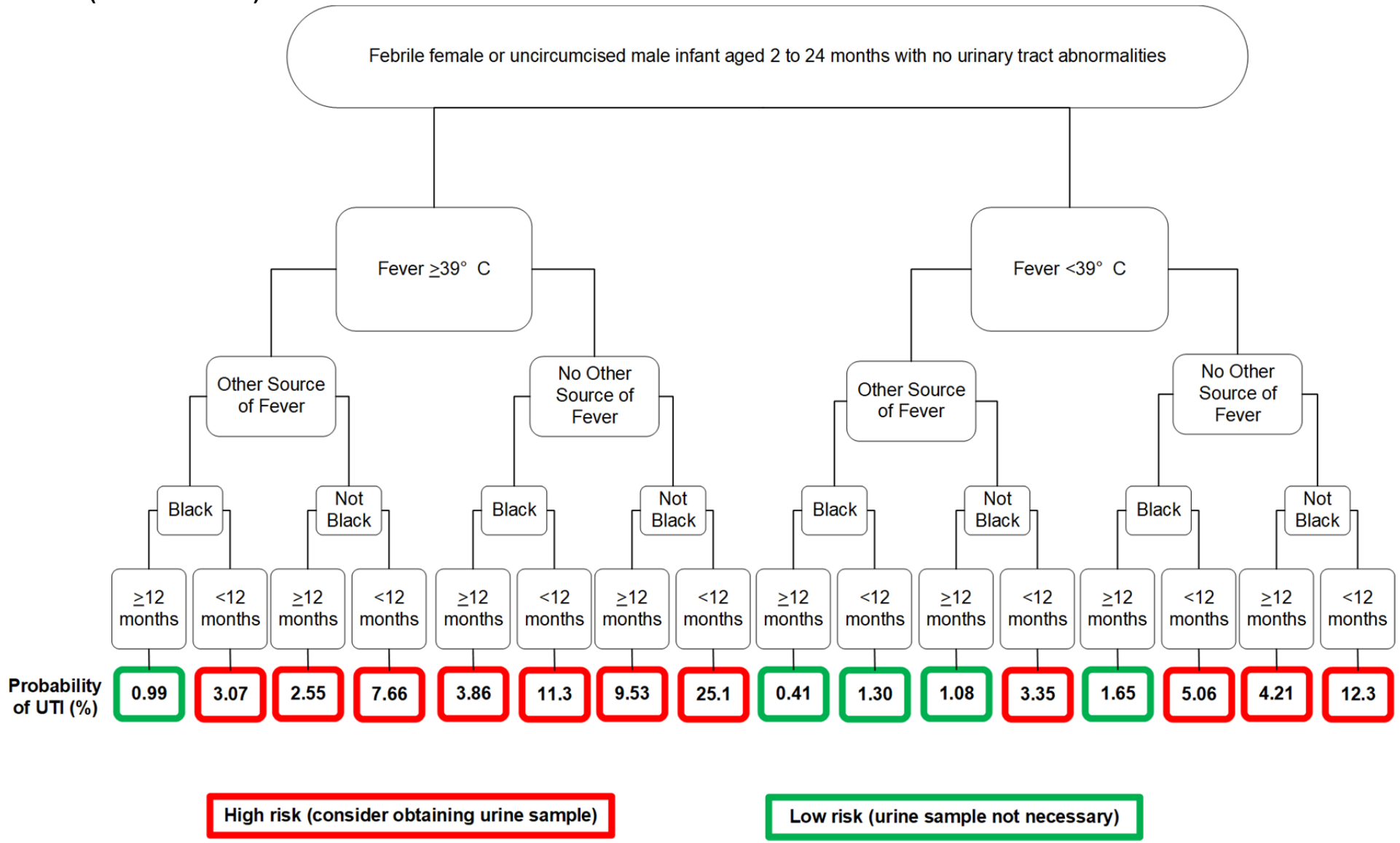
eTable 2. Overview of the Variables Included (and Their Odds Ratios) in Each of the Five Multivariable Models Developed in the Training Database

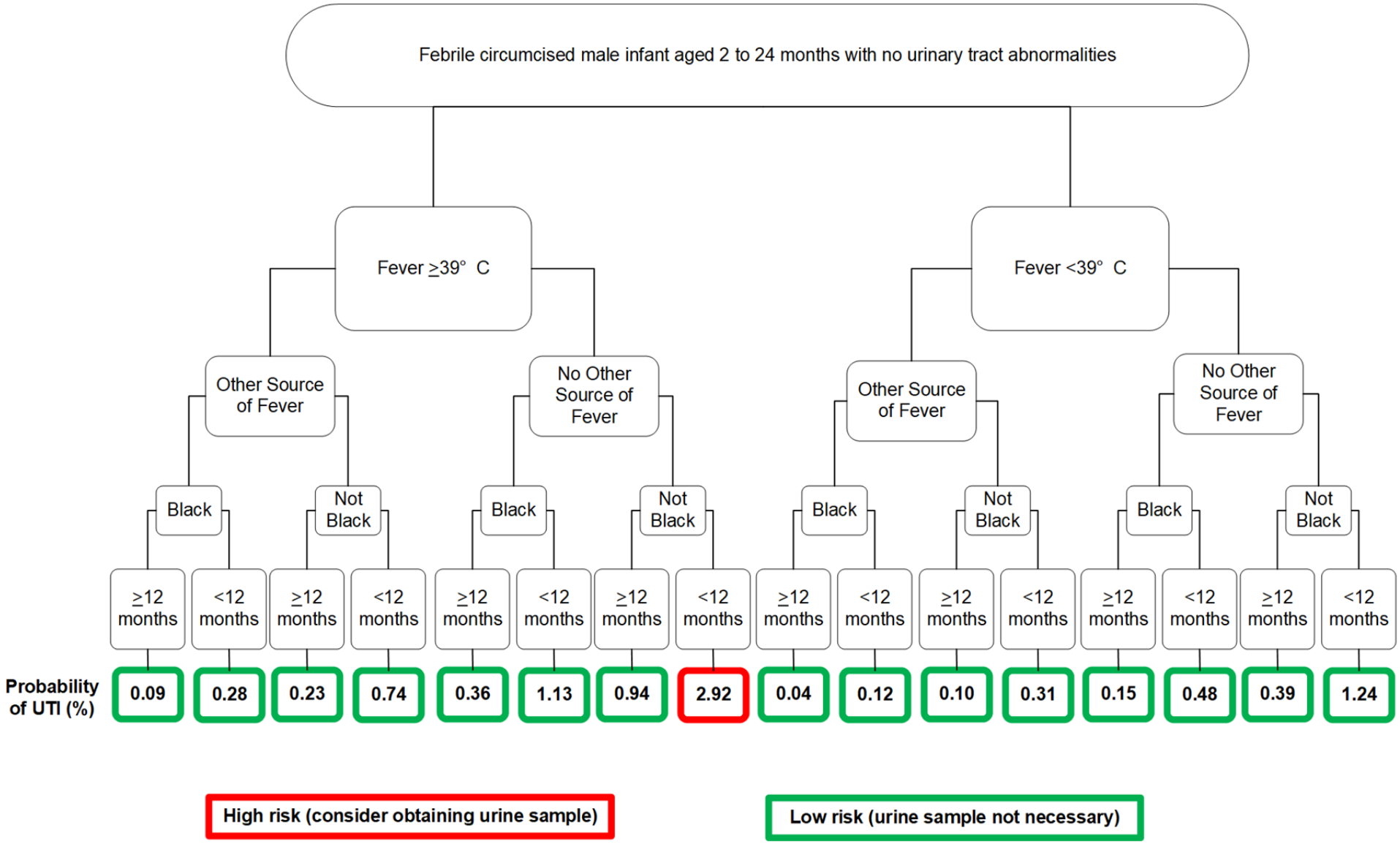
Variables included in the final versions of the models	Levels of the predictor variable[a]	Clinical Model	Dipstick Model	Dipstick + Gram Stain Model	Hemocytometer Model	Enhanced Urinalysis Model
		N=1593	N=1190[b]	N=901[b]	N=904[b]	N=900[b]
		OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Age in months	< 12 vs. ≥12	3.18 (2.40, 4.21)	1.74 (0.96, 3.18)	1.76 (0.83, 3.74)	1.87 (0.93, 3.76)	1.76 (0.76, 4.09)
Maximum temperature (°C)	≥39° C vs. <39° C	2.40 (1.84, 3.12)	1.74 (0.97, 3.12)	2.78 (1.33, 5.80)	2.08 (1.07, 4.05)	2.78 (1.26, 6.14)
Race	Not black vs. black	2.63 (1.94, 3.56)	4.15 (2.03, 8.50)	3.96 (1.64, 9.55)	3.36 (1.46, 7.75)	5.54 (2.00, 15.3)
Sex/circumcision status	Female or uncircumcised male vs. circumcised male	11.1 (6.85, 18.1)	6.20 (1.17, 32.8)	9.41 (0.88, 101)	5.80 (0.80, 42.3)	12.6 (0.62, 256)
No other fever source	Yes vs. No	4.03 (3.07, 5.29)	1.87 (1.03, 3.39)	1.44 (0.70, 2.98)	1.66 (0.84, 3.28)	1.49 (0.67, 3.28)
Nitrite	Yes vs. No		11.9 (4.88, 29.3)	6.89 (2.13, 22.26)	13.0 (4.49, 37.8)	4.65 (1.27, 17.0)
Leukocyte esterase	3+ vs. 2+ vs. 1+ vs. trace vs. none					
	Trace		11.9 (3.42, 41.2)	18.6 (4.52, 76.8)	10.0 (2.68, 37.5)	17.8 (4.11, 77.1)
	1+		37.5 (20.2, 69.7)	23.3 (10.8, 50.5)	13.9 (6.69, 29.0)	9.73 (4.03, 23.5)
	2+		214 (98.0, 468)	69.8 (25.0, 195)	29.40 (10.4, 82.8)	7.40 (2.05, 26.7)
	3+		1039 (349, 3089)	174 (53.0, 571)	65.6 (18.9, 227)	17.7 (4.38, 71.3)
WBC/mm^3	Continuous				1.04 (1.02, 1.05)	1.04 (1.03, 1.06)
Bacteria on Gram stain	Yes vs. No			24.7 (11.8, 51.7)		32.0 (14.5, 70.9)

OR = Odds ratio; CI = Confidence interval; WBC = White blood cell count
[a]For the categorical variables the reference category is listed last
[b]Only children who were deemed as "high risk" according to the Clinical Model were included

eTable 3. Multilevel Likelihood Ratios of Combination of Findings in the Training Database					
Age < 12 months	**Black**	**Other source of fever**	**Maximum temperature ≥39° C**	**Likelihood Ratio (95% CI)**	
				Female or uncircumcised male	**Circumcised male**
Yes	No	No	Yes	6.32 (4.85-8.22)	0.32 (0.14-0.76)
			No	2.07 (1.58-2.70)	0.14 (0.05-0.37)
		Yes	Yes	1.23 (0.85-1.77)	0.05 (0.01-0.37)
			No	0.32 (0.18-0.56)	0.04 (0.005-0.27)
	Yes	No	Yes	1.13 (0.71-1.81)	0.13 (0.02-1.01)
			No	0.89 (0.45-1.74)	0.12 (0.02-0.88)
		Yes	Yes	0.43 (0.22-0.85)	0.24 (0.07-0.79)
			No	0.32 (0.13-0.82)	0
No	No	No	Yes	1.10 (0.82-1.50)	0.36 (0.08-1.63)
			No	1.00 (0.52-1.93)	0.50 (0.06-4.46)
		Yes	Yes	0.50 (0.29-0.86)	0
			No	0.18 (0.04-0.77)	0
	Yes	No	Yes	0.91 (0.50-1.66)	0
			No	0.18 (0.02-1.40)	0
		Yes	Yes	0.18 (0.05-0.57)	0
			No	0	0

eFigure. Predicted Pre-test Probability of UTI for Individual Children According to Their Clinical Characteristics in the Training Dataset (Clinical Model)





Black race denotes that parent identifies child as Black (fully or partially). "Not Black" includes all other races (Caucasian, Asian, other).
Other source for fever includes (but is not limited to): acute otitis media, upper respiratory tract infection, gastroenteritis, pneumonia, meningitis, viral syndrome and bronchiolitis.

eReferences

1. Pencina MJ, D'Agostino RB, Pencina KM, Janssens AC, Greenland P. Interpreting incremental value of markers added to risk prediction models. *Am J Epidemiol*. 2012;176(6):473-481.
2. Shaikh N, Morone NE, Lopez J, et al. Does this child have a urinary tract infection? *JAMA*. 2007;298(24):2895-2904.
3. Chaudhari PP, Monuteaux MC, Shah P, Bachur RG. The importance of urine concentration on the diagnostic performance of the urinalysis for pediatric urinary tract infection. *Ann Emerg Med*. 2017;70(1):63-71.
4. Irwig ML, Bossuyt MP, Glasziou PP, Gastonis C, Lijmer JG. Designing studies to ensure that estimates of test accuracy will travel. In: Knottnerus JA, ed. *The Evidence Base of Clinical Diagnosis: Theory and Methods of Diagnostic Research*. Hoboken, NJ: Wiley; 2008:95-117
5. Hay AD, Sterne JA, Hood K, et al. Improving the diagnosis and treatment of urinary tract infection in young children in primary care: results from the DUTY Prospective Diagnostic Cohort Study. *Ann Fam Med*. 2016;14(4):325-336.
6. Hoberman A, Chao HP, Keller DM, Hickey R, Davis HW, Ellis D. Prevalence of urinary tract infection in febrile infants. *J Pediatr*. 1993;123(1):17-23.
7. Whiting P, Westwood M, Watt I, Cooper J, Kleijnen J. Rapid tests and urine sampling techniques for the diagnosis of urinary tract infection (UTI) in children under five years: a systematic review. *BMC Pediatr*. 2005;5(1):4.
8. Kohn MA, Carpenter CR, Newman TB. Understanding the direction of bias in studies of diagnostic test accuracy. *Acad Emerg Med*. 2013;20(11):1194-1206.

