The Use of Urinary Dipstick Tests to Exclude Urinary Tract Infection

A Systematic Review of the Literature

Andrew St John, PhD,[1] James C. Boyd, MD,[2] Andrew J. Lowes, MD,[3] and Christopher P. Price, PhD[4]

Key Words: Dipstick tests; Urinary tract infection; UTI; Rule-out tests; Leukocyte esterase; Nitrite; Posttest probability

DOI: 10.1309/C69RW1BT7E4QAFPV

Abstract

Several systematic reviews have examined the use of dipstick tests to diagnose or rule in urinary tract infection (UTI). We examined the evidence relating to the use of urine leukocyte esterase and nitrite tests in adults to exclude or rule out UTI. A search of the literature from 1966 to 2003 revealed 30 studies as containing relevant and suitable information, and 23 of these, which used a cutoff of 10^8 colony-forming units per liter, were combined in a meta-analysis. The leukocyte esterase or nitrite test combination, with one or the other test positive, was used in 14 studies and showed the highest sensitivity and the lowest negative likelihood ratio. While there was significant heterogeneity between the studies, 7 of 14 demonstrated significant decreases in pretest to posttest probability with a pooled posttest probability of 5% for the negative result. In certain circumstances, there is evidence for the use of urinalysis as a rule-out test for UTI.

Urinary tract infection (UTI) is a common complaint. A telephone survey in the United States revealed that 11% of women 18 years or older reported at least 1 presumed UTI during the last year,[1] and an estimated 60% of women have a UTI at some stage in their life.[2] Approximately 23% of all hospital-acquired infections are due to UTI.[3] The incidence is increased in elderly men and women, particularly among those living in institutions, where it can be up to 53% and 37% in women and men, respectively.[4]

In the past, UTI often has been defined for research purposes as the presence of at least 10^8 colony-forming units (CFU) per liter in freshly voided urine, although symptomatic infection can occur with 10^6 CFU/L,[4] and some would argue for a cutoff of 10^5 CFU/L in symptomatic patients when urine can be cultured without delay.[5] Diagnosis through bacterial culture is important in some circumstances because failure to detect a UTI can have serious consequences, particularly in certain patients such as pregnant women. However, uncomplicated UTI in nonpregnant women rarely causes severe illness or has significant long-term consequences, and in 50% of patients, the condition improves without antimicrobials within 3 days.[6]

Nevertheless, suspected UTI is one of the most common indications for antimicrobial use, with much prescribing of antimicrobials based on clinical symptoms and signs without confirmation by culture.[7] As antimicrobial resistance rates increase, there is growing concern about inappropriate and unnecessary antimicrobial use.[8] Although empirical treatment of women with suspected uncomplicated UTI was shown in 1 decision analysis to be the most cost-effective strategy, adding a dipstick test and using the results to rule out treatment could be justified if reducing antimicrobial treatment is an additional objective.[9] The emergence of multidrug-resistant

extended-spectrum β-lactamase-producing *Escherichia coli* emphasizes the increasing risks associated with unnecessary antimicrobial use,[10,11] which already was known to increase the risk of subsequent UTIs.[12]

The frequency of UTI generates a significant workload for the laboratory, with large laboratories analyzing from 200 to 300 urine specimens per day.[13] Most specimens sent to a laboratory will show no evidence of infection when tested, and, consequently, there has been considerable interest in ways to screen out negative specimens before processing them for culture by a rule-out test strategy.[14] There have been suggestions that this could be particularly appropriate for screening out asymptomatic bacteriuria in a population with a low prevalence of UTI, such as pregnant women, as a cost-effective alternative to culture[15,16]; however, a recent review commissioned by the National Institute for Clinical Excellence rejected this advice on the basis of current evidence, although it highlighted the problem as an important research topic.[17]

In recent times, there has been an increased understanding of how to evaluate the clinical usefulness of diagnostic tests.[18] Although many studies have compared dipstick testing with culture, few have taken into account the wide variation in prevalence or pretest probability of UTI in different populations. The prior or pretest probability of a condition is an important factor in determining the effect of a positive or negative test result on the posttest probability of ruling a diagnosis in or out.[19] Furthermore the clinical significance of missing a true-positive result will depend on the population being examined and ranges from possible fetal mortality in pregnancy, in which there is a low pretest probability of UTI, to little evidence of significant excess morbidity in uncomplicated UTIs in women, in whom there is a high pretest prior probability of UTI. In addition there is the issue of choice of thresholds of colony counts in urine for confirmation of infection.[13,20] Commonly used colony count thresholds to establish cutoffs for distinguishing normal and abnormal dipstick test results were set many years ago, often for purposes that are different from those for which the tests are currently used. Even when the dipstick tests have been evaluated in an appropriate context, the comparator test has been a urine culture using a cutoff of 10^8 and sometimes 10^7 CFU/L to define UTI.

Systematic reviews and meta-analyses of studies using these tests in adults and children have identified other deficiencies in many of the included studies.[21-23] Gorelick and Shaw[21] showed that in children, the Gram stain and dipstick analysis for leukocyte esterase (LE) and nitrite performed similarly in detecting UTI (used as rule-in tests) and were superior to microscopic analysis for pyuria. A more recent systematic review and meta-analysis in children included a larger number of articles than reviewed by Gorelick and Shaw[21] and concluded that pyuria of 10 or more cells per high-power field (or microliter) and bacteriuria as detected by microscopy were best suited for diagnosing UTI in children, whereas there were insufficient data to draw conclusions about the value of urinary dipstick tests when used as a rule-in test.[22]

At the time this study was begun, the only systematic review of UTI dipstick data specifically from adults concluded that a positive LE or nitrite test result was the best index or rule-in test but a negative dipstick result (rule-out test) could not exclude UTI in patients with a high prior probability of contracting the condition.[23] Since publication of that study, several new studies have focused on using dipstick tests to exclude (rule out) rather than diagnose (rule in) UTI in adults, and this prompted the systematic review described herein. As part of the present study, we looked at factors that might influence the performance of the tests, eg, pretest probability and methodological issues. While we carried out our review, another related systematic review by Deville et al[24] was published.

Materials and Methods

Identification of Relevant Literature

Although the primary purpose of the review was to examine the use of dipstick tests in adult UTI, the literature search was extended to cover all age groups. Thus, the National Library of Medicine's MEDLINE and Embase databases were used to conduct a search for articles between January 1966 and December 2002 related to the diagnosis of UTI. The search terms included the following: urinary tract infections, bacteriuria, pyuria, LE, nitrite, urinary protein, hematuria, dipstick, reagent strip, screening, rule out, cost-effectiveness, urine culture, laboratory diagnosis, and urinalysis.

In addition to the database searches, the reference lists of articles that were reviewed for possible inclusion in the study and other review articles were searched for further possible references. The articles were reviewed by 2 of us (A.J.L. and A.S.J.) for possible relevance; discrepancies were resolved by further discussion. The initial criteria for inclusion of articles into the review were as follows:

- Contained original data that had not been published previously (no narrative reviews)
- Used dipstick tests for LE, nitrite, blood, and protein
- Compared the dipstick tests with a laboratory culture method as the reference standard that, together with the dipstick tests, was performed on all specimens
- Used a definition of a UTI based on the culture method by a stated number of colonies per liter or a cutoff value
- Reported sufficient data for dipstick tests and culture standard to permit the creation of a table containing figures for true- and false-positive and true- and false-negative results (2 × 2 table)

Additional factors for assessing the quality of the articles were as follows: the clinical setting of the study, ages of the

DOI: 10.1309/C69RW1BT7E4QAFPV

subjects or age ranges (ie, adults or children) given, whether a prospective or a retrospective study, whether patients were selected at random or consecutively, the prevalence of UTI in the cohort studied, details of patient symptoms, whether the operators responsible for the culture tests were blinded to the results of the dipstick tests, and how contaminated specimens and mixed cultures were treated in the analysis.

Extraction of Data and Combining the Results

Data were extracted from all reviewed articles that met the selection criteria. For each dipstick test and, when appropriate, combinations of tests, a 2 × 2 table was created as follows:

	UTI	No UTI
Positive test result	True-positive	False-positive
Negative test result	False-negative	True-negative

For articles reporting multiple thresholds for the reference method cutoff, a separate table was extracted for each threshold. From the tables, the sensitivity and specificity were calculated for each test, together with the confidence intervals, which were calculated according to Gardner and Altman.[25] The results and confidence intervals were compared visually for each study by using a forest plot.[26] In addition, for each test in each study, the calculated sensitivity was plotted against 1 – specificity to graphically visualize possible heterogeneity that might be present.

The homogeneity of diagnostic parameters (sensitivity, specificity, positive and negative likelihood ratios [LRs], and the diagnostic odds ratios [DORs]) across studies was tested by using the method of Cochrane, which provides a χ^2 statistic computed as the weighted sum of the square differences between the overall summary estimate and the results of individual studies.[27] Pooled estimates of mean values and confidence intervals were generated by using fixed-effects and random-effects modeling for sensitivity, specificity, positive and negative LRs, and the DOR. The statistical analyses were carried out using the SAS MIXED procedure (SAS Institute, Cary, NC).

Results

The various searches generated a total of 1,450 titles. After review of the abstracts, 250 were selected for a full review because they dealt with the study of UTI. Only 118 articles met the initial basic criteria, and these were allocated into 4 groups: (1) insufficient description of the patient population, primarily technical evaluations, 33 articles; (2) studies in children, 29 of which were included in 2 previous systematic reviews,[21,22] 32 articles; (3) studies in adults and children, 15 articles; and (4) studies in adults with additional information about the cohort studied, 38 articles.

Of the 38 articles in group 4, 3 reported data from 2 subject groups, so there were 41 studies in total. Within this group, there was considerable variation in how contaminated specimens and mixed cultures were treated in the analysis. In particular, 11 studies indicated that mixed cultures and contaminated specimens were excluded from the analysis, so these articles were excluded from further review. Another major discriminator between the studies was an indication of whether the operators of the dipstick or reference tests were unaware of the results of the other test, designated as blinded, or whether no such indication was given, designated as unblinded. Details of the tests used, cutoff, and whether the study was blinded or unblinded are given in **Table 1**.

The most common tests used were LE and nitrite, separately or in combination in which a result was assigned as positive if one or the other test indicated a trace or greater (LE or nitrite positive) or in which both tests had to indicate a trace or greater for a positive result (LE and nitrite positive). Although other dipstick tests such as blood and protein were used in certain studies, the low numbers of studies did not justify combining the data. Of the

Table 1
Numbers of Studies by Colony Count Used to Define Urinary Tract Infection, Tests Performed, and Level of Operator Blinding in Articles Meeting Other Selection Criteria

	Colony Count							
	>10^8 CFU/L		>5 × 10^7 CFU/L		>10^7 CFU/L		>10^6 CFU/L	
Test*	Blinded	Unblinded	Blinded	Unblinded	Blinded	Unblinded	Blinded	Unblinded
LE	12	3	—	—	2	—	1	—
N	11	5	—	—	2	—	1	—
LE or N	9	5	1	—	1	2	1	—
LE and N	6	1	—	—	1	—	1	—
LE or N or blood	2	—	—	—	1	—	—	—
LE or N or blood or protein	1	—	—	—	—	1	—	1
LE or N or protein	1	—	—	—	—	—	—	—

CFU, colony-forming units; LE, leukocyte esterase; N, nitrite.
* Dipstick test combinations for diagnosis of disease. When 1 test is listed, only that test result was positive; when 2 or more are joined by "or," one of the test results was positive; when 2 are joined by "and," the results of both tests were positive.

DOI: 10.1309/C69RW1BT7E4QAFPV

30 sets of data, 23 used 10^8 CFU/L as the cutoff for definition of UTI and 7 used a variety of lower cutoffs. The selected studies included 8 data sets from pregnant women, and the remainder described a variety of clinical settings and UTI prevalence.[28-47] The low numbers of studies in each of the clinical settings involving nonpregnant patients prevented meta-analysis of the data from these individual groups. **Table 2** shows the studies that used 10^8 CFU/L as the cutoff, classified according to clinical setting, prevalence of UTI, dipstick test, and the sensitivity and specificity for each of the tests.

Meta-analysis of the data initially was performed on the blinded and unblinded subgroups of the cohort; the summary statistics are given in **Table 3**. Statistical tests showed significant heterogeneity for all tests in the blinded and unblinded

Table 2
Features and Sensitivity and Specificity of Studies Using a Cutoff of 10^8 CFU/L

					LE		N		LE or N		LE and N	
Study	Sex	No. of Subjects	Patient Setting	UTI Prevalence (%)	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
Audurier et al,[28] 1988	M/F	2,183	Laboratory	16	0.856	0.762	0.479	0.958	0.897	0.749	0.438	0.971
Bachman et al,[29] 1993	F	1,047	Antenatal	2	0.167	0.972	0.458	0.997	0.500	0.969	0.125	1.000
Bowman and Riley,[30] 1991	M/F	1,020	Laboratory	22	NR	NR	NR	NR	0.969	0.757	NR	NR
Chernow et al,[31] 1984	M/F	203	Hospital, including catheterized patients	24	1.000	0.760	0.265	0.987	NR	NR	NR	NR
Etherington and James,[15] 1993	F	898	Antenatal	3	0.593	0.861	0.667	0.997	0.741	0.859	NR	NR
Evans et al,[32] 1991	M/F	50	Hospital inpatients	36	0.722	0.813	0.833	1.000	NR	NR	NR	NR
Evans et al,[32] 1991	M/F	50	Day-hospital patients	40	0.600	0.867	0.900	1.000	NR	NR	NR	NR
Graninger et al,[33] 1992	F	1,000	Antenatal	13	0.587	0.872	0.540	1.000	0.786	0.872	NR	NR
Hagay et al,[34] 1996	F	313	Antenatal	8	0.529	0.903	0.375	0.993	NR	NR	NR	NR
Jellheden et al,[35] 1996	F	795	General practice, excluding pregnancy	83	NR	NR	0.620	0.889	NR	NR	NR	NR
Lachs et al,[36] 1992	M/F	107	Hospital ED and clinic	50	NR	NR	NR	NR	0.556	0.780	NR	NR
Lachs et al,[36] 1992	M/F	259	Hospital ED and clinic	7	NR	NR	NR	NR	0.925	0.704	NR	NR
Lammers et al,[37] 2001	F	331	Hospital ED and clinic	46	0.908	0.408	NR	NR	0.849	0.531	0.270	0.927
Lenke and Van Dorsten,[38] 1981	F	146	Antenatal	12	NR	NR	0.222	1.000	NR	NR	NR	NR
McNair et al,[39] 2000	F	528	Antenatal	7	NR	NR	NR	NR	0.472	0.803	NR	NR
Monane et al,[40] 1995	F	427	Elderly housing and long-term care	17	0.863	0.559	NR	NR	NR	NR	NR	NR
Pallares et al,[41] 1988	M/F	180	Health care center	54	0.857	0.976	0.510	0.951	0.929	0.915	NR	NR
Preston et al,[42] 1999	F	228	Hospital gynecology ward	12	0.750	0.885	0.643	0.990	NR	NR	0.964	0.885
Robertson and Duff,[43] 1988	F	750	Antenatal	8	0.774	0.975	0.435	0.988	0.919	0.951	0.323	0.942
Sewell et al,[44] 1985	M	459	Laboratory	25	NR	NR	NR	NR	0.783	0.436	NR	NR
Soisson et al,[45] 1985	F	1,062	Antenatal	6	NR	NR	0.452	0.996	NR	NR	NR	NR
Tissot et al,[46] 2001	M/F	339	ICU, including catheterized patients	17	0.842	0.730	0.509	0.848	0.877	0.617	0.421	0.957
Tuel et al,[47] 1990	M/F	169	Spinal-injuries, with catheters	28	0.660	0.918	0.681	0.943	0.872	0.869	0.468	0.992

CFU, colony-forming units; ED, emergency department; ICU, intensive care unit; LE, leukocyte esterase; N, nitrite; NR, not reported; Sens, sensitivity; Spec, specificity.

Table 3
Summary Statistics for Sensitivity, Specificity, LR, and DOR for Each Test in the Two Cohorts Separately and Combined

	Sensitivity	CI	Specificity	CI	LR+	CI	LR–	CI	DOR	CI
Blinded and unblinded studies										
LE	0.72	0.61-0.84	0.82	0.74-0.90	4.87	3.26-7.29	0.31	0.18-0.51	16.8	9.93-28.5
N	0.54	0.44-0.64	0.98	0.96-0.99	29.3	14.4-59.7	0.48	0.37-0.62	63.4	29.6-136
LE or N	0.81	0.71-0.90	0.77	0.69-0.86	4.27	2.82-6.47	0.22	0.14-0.35	19.9	9.84-40.0
LE and N	0.43	0.23-0.64	0.96	0.93-0.99	9.61	5.44-17.0	0.54	0.26-1.13	23.7	8.63-65.3
Blinded studies										
LE	0.70	0.57-0.82	0.82	0.72-0.93	5.21	3.10-8.75	0.34	0.25-0.47	16.2	8.71-30.0
N	0.57	0.44-0.70	0.98	0.95-1.00	30.6	13.2-71.0	0.44	0.31-0.64	70.8	28.1-179
LE or N	0.82	0.70-0.93	0.80	0.70-0.90	5.16	2.90-9.20	0.22	0.13-0.37	24.4	10.4-56.9
LE and N	0.43	0.19-0.67	0.95	0.92-0.99	8.64	4.56-16.4	0.53	0.22-1.25	25.1	6.99-89.8
Unblinded studies*										
LE	0.82	0.57-1.00	0.80	0.72-0.88	3.94	3.35-4.63	0.14	0.01-1.40	20.8	5.80-74.4
N	0.48	0.35-0.60	0.97	0.94-1.00	29.6	6.20-141	0.53	0.44-0.65	54.6	11.5-260
LE or N	0.79	0.62-0.97	0.72	0.58-0.87	3.11	1.87-5.16	0.22	0.08-0.60	14.1	3.84-51.5

CI, confidence interval; DOR, diagnostic odds ratio; LE, leukocyte esterase; LR, likelihood ratio; LR+, rule-in test; LR–, rule-out test; N, nitrite.
* For LE and N, there were insufficient studies.

DOI: 10.1309/C69RW1BT7E4QAFPV

groups; this heterogeneity was confirmed further by the wide confidence intervals for many of the parameters, including LRs and DORs. A forest plot of the LRs derived from the studies for blinded and unblinded groups is shown in **Figure 1**.

To study the possible causes of this heterogeneity and to see whether it would be acceptable to combine the blinded and unblinded groups and to combine the studies in pregnancy with those in other clinical settings, further statistical analysis was performed according to Lijmer et al.[48] This approach is an extension of the meta-analytic regression model originally described by Moses et al[49] that allows examination of differences in diagnostic accuracy between tests to be examined by adding [0,1] indicator variables to the regression equation. When blinding and pregnancy were added as indicator variables, neither exerted a significant effect on the resultant DOR. Accordingly, the various groups were combined, and the summary statistics are given in Table 3.

Table 3 shows the mean values and confidence intervals obtained by random-effects analysis for sensitivity, specificity, positive and negative LRs, and DORs. Because of the heterogeneity in the studies, the data for the blinded and unblinded studies were analyzed together and separately (Table 3).

Of the 4 test combinations, the "LE or nitrite, one or the other positive" combination had the highest sensitivity with similar values in the blinded and unblinded cohort of studies, and the nitrite test had the highest specificity. The negative LRs were largely similar for all tests, but the lowest values were again for the LE or nitrite test combination and for LE alone in the unblinded group. However, in all cases, the confidence intervals for all LE or nitrite values were very wide. The highest positive LR was observed for the nitrite test, which also had the highest DOR (63.4). Although the meta-analysis was confined to studies using the 10^8 CFU/L cutoff, we combined the LE or nitrite data from these studies with data from studies using lower cutoffs to produce a summary receiver operating characteristic curve **Figure 2**.

Table 4 shows the pretest and posttest probabilities calculated from the LRs for the individual studies, blinded and unblinded, that used the LE or nitrite dipstick test combination compared with the pooled result. An alternative approach to reviewing these data is to plot the data as a graphic representation of Bayes theorem as shown in **Figure 3**.[50] The latter approach allows easy visualization of the change in pretest to posttest probability when using the test.

Discussion

The purpose of this systematic review was to determine whether there is sufficient evidence in the literature to indicate that in certain populations it is possible to use dipstick tests as

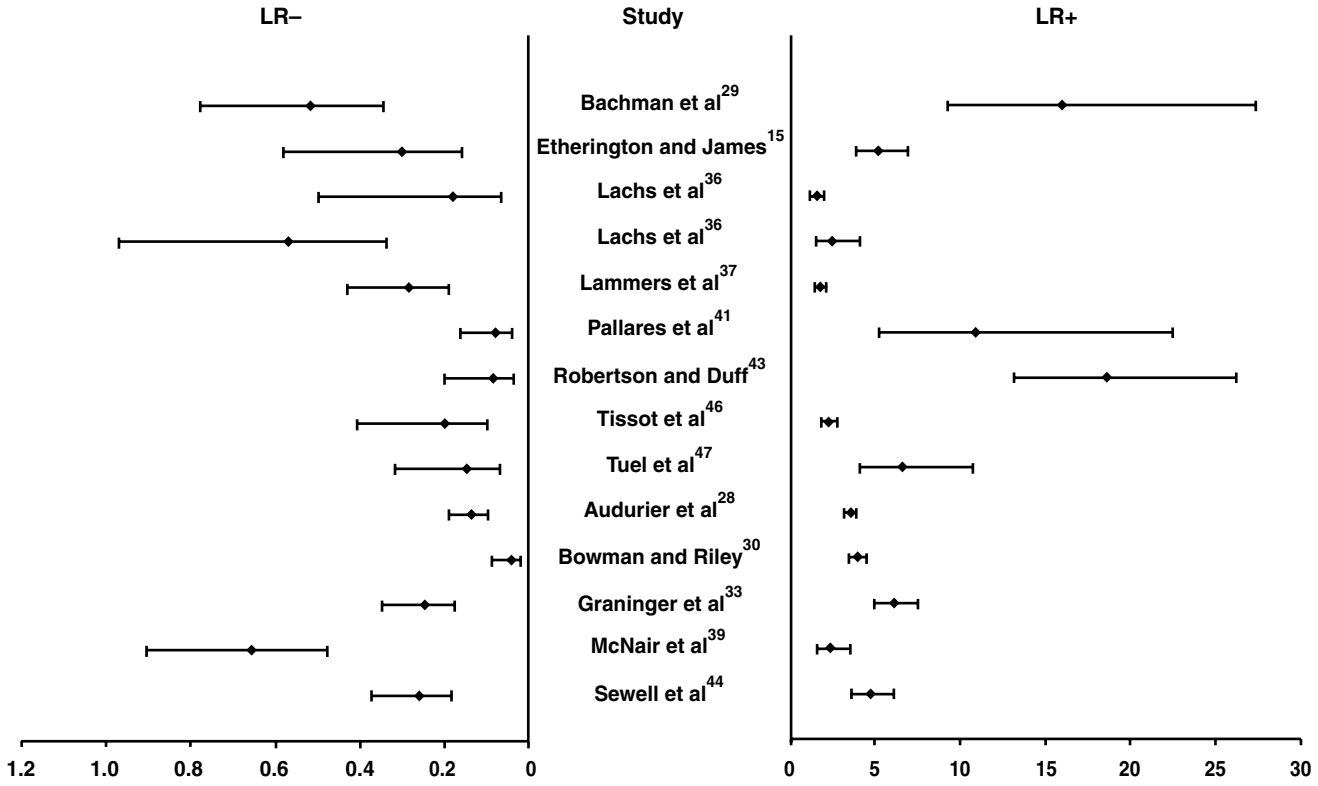

Figure 1 Forest plot of likelihood ratios (LRs) for the leukocyte esterase or nitrite combination as a rule-in (LR+) and a rule-out (LR–) test.

DOI: 10.1309/C69RW1BT7E4QAFPV

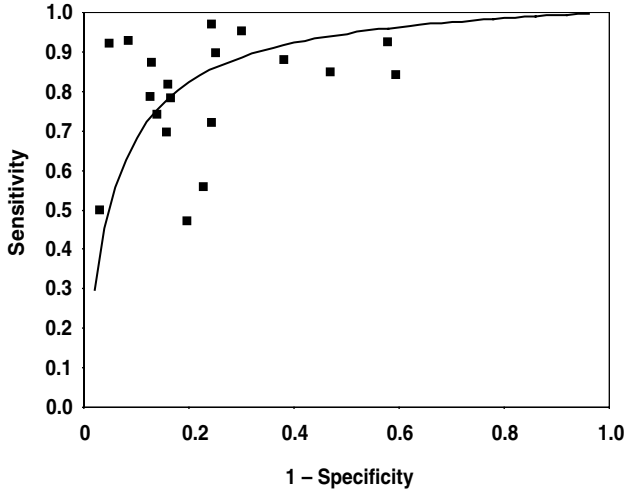

Figure 2 Summary receiver operating characteristic curve for all studies using the leukocyte esterase or nitrite test combination.

Table 4
Pretest and Posttest Probabilities of Dipstick Leukocyte Esterase and Nitrite Combinations (One or Both Positive) in 14 Studies and the Pooled Results

Study	Prevalence (Pretest Probability)	Posttest Probability: Positive Result	Posttest Probability: Negative Result
Audurier et al[28]	0.16	0.40	0.02
Bachman et al[29]	0.02	0.25	0.01
Bowman and Riley[30]	0.22	0.53	0.01
Etherington and James[15]	0.03	0.14	0.01
Graninger et al[33]	0.13	0.47	0.03
Lachs et al[36]	0.50	0.61	0.15
Lachs et al[36]	0.07	0.16	0.04
Lammers et al[37]	0.46	0.61	0.19
McNair et al[39]	0.07	0.15	0.05
Pallares et al[41]	0.54	0.93	0.08
Robertson and Duff[43]	0.08	0.63	0.01
Sewell et al[44]	0.25	0.61	0.08
Tissot et al[46]	0.17	0.32	0.04
Tuel et al[47]	0.28	0.72	0.05
Pooled results*	0.20	0.52	0.05

* Pooled prevalence is the mean prevalence across all the 14 studies. The pooled posttest probabilities are calculated from the summary likelihood ratios shown in Table 3 (blinded and unblinded studies section) and posttest odds using the following formulas: (1) Pretest Odds = Prevalence/1 – Prevalence; (2) Posttest Odds = Pretest Odds × Likelihood Ratio; (3) Posttest Probability = Posttest Odds/1 + Posttest Odds.

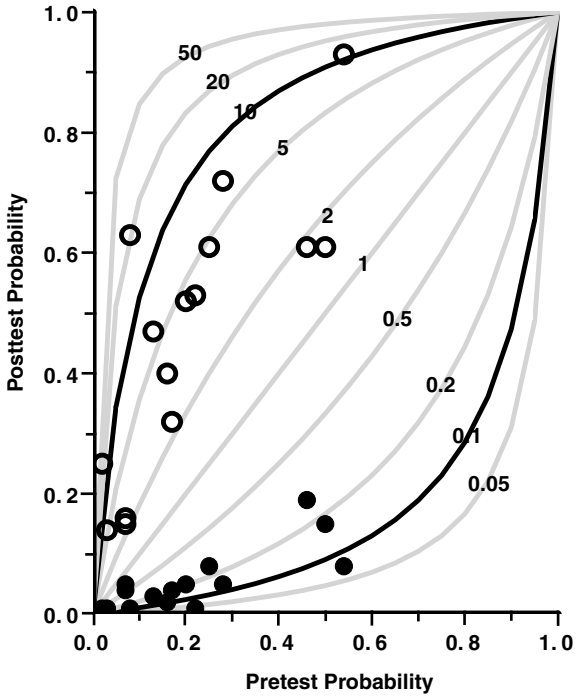

Figure 3 The relationship of pretest and posttest probabilities and likelihood ratios according to the Bayes theorem. The solid lines (isocontours) represent the pretest and posttest probabilities for various likelihood ratios ranging from 0.05 to 50. The individual data points show the pretest and posttest probabilities for the individual studies chosen in this review, using the leukocyte esterase or nitrite test combination. Open circles represent the relationships between pretest and posttest probabilities for positive test results (rule-in tests) and the closed circles for negative test results (rule-out tests). An optimal rule-in test is characterized by having a high likelihood ratio (eg, plots above the isocontour corresponding to a likelihood ratio of 10) and shows a large *increase* of pretest to posttest probability. An optimal rule-out test is characterized by having a low likelihood ratio (eg, plots below the isocontour corresponding to a likelihood ratio of 0.1) and shows a large *decrease* of pretest to posttest probability.

a screening process to rule out UTI and, thus, reduce the number of negative urine samples sent for culture. A number of parameters can be used to assess the ability of a test to rule out a condition, including sensitivity, negative LR, and DOR. For a test to be effective at ruling out a condition, it needs to have high sensitivity to minimize the number of false-negative results.[19] Negative LRs less than 0.2 indicate useful diagnostic evidence, and LRs less than 0.1 provide convincing evidence that with a negative test result the disease is absent.[51] However, these guideline LRs depend on whether the pretest probability of disease is high or low.

The DOR expresses the odds of a positive result in patients with disease compared with the odds of positive results in patients without disease. Although it has the advantage of summarizing the accuracy of a test as a single value and has less variability from study to study than LRs, it is a less interpretable value from a clinical perspective.[52]

Two previous systematic reviews of UTI studies in children focused on the use of dipstick tests to diagnose or rule in UTI.[21,22] At the beginning of the present review, the only

DOI: 10.1309/C69RW1BT7E4QAFPV

systematic review performed in adults was that by Hurlbut and Littenberg,[23] which focused on evidence to support using dipstick tests to diagnose UTI and found that the best rule-in test was the LE or nitrite combination. The review did not identify the sources of heterogeneity, nor did it clearly identify which studies were included; the citations of articles accepted for the study include articles describing secondary research. The data derived from Hurlbut and Littenberg[23] were analyzed further in a review by Bent et al,[53] who looked at the accuracy and precision of history-taking and physical examination for the diagnosis of UTI in women. By using the data from Hurlbut and Littenberg,[23] Bent et al[53] calculated the LRs for the LE or nitrite combination (one or the other positive) to be 4.2 and 0.3 for the positive and negative test results, respectively, and the posttest probabilities to be 81% for the positive result and 23% for the negative result. This latter figure is not sufficiently low to rule out disease, and, consequently, Bent et al[53] agree with the conclusion of Hurlbut and Littenberg[23] that dipstick tests cannot be used to exclude UTI.

In the present review, the DOR results indicated that the most useful diagnostic test seems to be nitrite with a DOR value of 63.4 but with extremely wide confidence intervals. The LE or nitrite combination showed the highest sensitivity, with a pooled negative LR of 0.22 and a posttest probability of 5% for the negative result. This is a lower result than that derived from the review by Hurlbut and Littenberg[23] and Table 4 shows that 8 of 14 studies included in the meta-analysis demonstrated posttest probabilities of less than 5%, and in some of these the reduction in pretest to posttest probability was significant.

Although our review might be interpreted as providing more convincing evidence for the value of dipstick testing to exclude UTI than the previous systematic review by Hurlbut and Littenberg,[23] our pooled data also had wide confidence intervals, which may be due to a number of factors. We demonstrated significant heterogeneity between the studies but were unable to elucidate the causes, although variation in the diagnostic threshold for the reference test is likely to be a cause. Another cause is suggested in the findings of another systematic review of UTI testing that was published during the present review. Deville et al[24] used a different approach to the selection of articles to be included, relying on a grading or marking system to assess the quality of articles. In contrast, we used absolute criteria to include or exclude articles with the consequence that fewer articles met our criteria for inclusion in the meta-analysis, and the possibility for subclass analysis, such as that performed by Deville et al,[24] was very limited. Given the different approaches taken by us and Deville et al,[24] it is perhaps not surprising that we produced different evidence. The 2 systematic reviews of UTI in children mentioned previously also arrived at different conclusions in relation to the use of dipsticks to diagnose UTI.[21,22]

Although it might be argued that less stringent selection criteria for included articles might cast some doubt on the accuracy of the conclusions, the findings of Deville et al[24] support the concept of using dipstick testing to rule out UTI. In a variety of clinical situations, they found that combining the results of LE and nitrite increased the sensitivity from 68% to 88%, and the corresponding posttest probabilities or predictive values of the negative test ranged from 84% to 98%, with the value being greater than 95.5% in the majority of the settings studied. Corresponding figures in our 14 studies using the LE or nitrite test combination were 81% to 99% (Table 2). Deville et al[24] concluded that a negative dipstick result excluded the presence of infection in most studies, contrary to the findings of Hurlbut and Littenberg.[23]

Thus, although negative LRs greater than 0.1 may not be regarded as strong evidence, the magnitude of the decreases in pretest to posttest probability of UTI conferred by a negative result as shown by us and Deville et al[24] indicates the potential value of urinalysis as a rule-out test. Even though convincing evidence (as defined by Jaeschke et al[51]) has been hard to come by, anecdotal evidence suggests that many laboratories use dipstick testing as part of an algorithm that also includes the appearance of the urine and clinical symptoms and history to determine whether a urine culture needs to be performed. On the other hand, analysis of the studies looking at the transition of pretest to posttest probability suggests that the use of an LE or nitrite combination in a rule-out strategy would give an acceptable posttest probability of 5%, especially from a situation in which there is a low- to mid-range pretest probability.

The use of history and symptom scoring systems in the diagnosis of UTI has been published and might add additional value to a UTI algorithm.[53,54] Dobbs and Fleming[54] found that a combination of symptom history and dipstick testing (using nitrite, protein, and hemoglobin) in a wide spectrum of patients identified 89% of infected cases and included 33% of noninfected cases. Bent et al[53] found that the presence of one or more symptoms was associated with a probability of infection of approximately 50%, the figure at which Dobbs and Fleming[54] considered it appropriate to institute antibiotic therapy. The combination of certain symptoms, eg, dysuria and frequency without vaginal discharge, raised the probability to 90%. However, these authors also found that history, physical examination, and urinalysis did not lower the posttest probability to a level at which UTI could be ruled out when a patient had one or more symptoms.

Although all of the individual symptoms yielded summary positive LRs between 0.2 and 1.7 (compared with 4.2 for urinalysis), the combination of symptoms yielded a value of 24.6, equating to a posttest probability of only 77%. The negative LRs for individual symptoms yielded summary negative LRs between 0.8 and 2.6, and when symptoms were combined, a value of 0.3, equating to a posttest probability of 4%.

It is perhaps worth noting that the prevalence of UTI in the studies reviewed by Bent et al[53] was high, ranging from 12% to 59%. On the basis of their review, Bent et al[53] proposed an algorithm for the diagnosis of UTI, including history,

DOI: 10.1309/C69RW1BT7E4QAFPV

symptoms, and urinalysis, effectively the conclusion drawn by Dobbs and Fleming[54] and by Smith et al[12] in a more recent study using urinalysis and phase contrast microscopy. However Bent et al,[53] by taking the data from their review and inserting it into an algorithm that included history, symptoms, and urinalysis, found that a complete profile of negative results still implied a posttest probability of about 20%, whereas our data would suggest that figure is closer to 5%.

The data from the pretest and posttest probabilities, when plotted on a graphic representation of Bayes theorem, indicate that when used as a rule-out test (ie, negative test results), the posttest probabilities tend to be small, with those for most studies less than 0.05. Furthermore, the majority of the posttest probabilities lie below the likelihood line equivalent to 0.2. On the other hand, for the use of the LE or nitrite test combination as a rule-in test (ie, positive test results), several of the studies do not achieve a posttest probability of 95%, and most of the points lie below the LR line equivalent to 10.

All systematic reviews of this area have demonstrated considerable heterogeneity in the accuracy of dipstick testing, which is due to a variety of causes. One of the major causes is the different prevalence of UTI or pretest probability of the disease in the studied population. However, pretest probability can be assessed approximately from the type of patient and the clinical history. The latter features including symptoms have not always been well documented in the studies of the accuracy of urinalysis. This information, combined with an LR derived from the dipstick test, can be used to derive a posttest probability.[54] Such an approach can be used by a general or family practitioner to guide decisions such as whether a culture is needed and what type of treatment. The LRs identified by the present review and the review by Deville et al[24] suggest that this approach is valid, and, perhaps of greater significance, is the significant transition of the pretest to posttest probability when using the LE or nitrite in a rule-out test strategy.

There is evidence supporting the use of urinalysis as a rule-out test for UTI when a cutoff for a culture result of 10^8 CFU/L is appropriate, eg, suspected acute pyelonephritis, when a negative dipstick result might prompt consideration of other explanations for a patient's illness, especially when considering the transition from pretest to posttest probability. Unfortunately, the current evidence suggests that urinalyses should not be used in a screening scenario such as antenatal screening for asymptomatic bacteriuria, when it may be important to detect all positive results. The definitive answer to establishing the role of urinalysis in a rule-out strategy is to evaluate the use of a diagnostic algorithm in a prospective, randomized controlled trial in individual populations; one such Health Technology Assessment Authority–sponsored trial in family practice is in progress.

From [1]ARC Consulting, Perth, Western Australia; [2]Department of Pathology, University of Virginia Health System, Charlottesville; [3]Southampton HPA Laboratory, Southampton General Hospital, Southampton, England; and [4]Department of Clinical Biochemistry, University of Oxford, Oxford, England.

Address reprint requests to Dr St John: ARC Consulting, 14 Learoyd St, Mt Lawley, W Australia 6050.

References

1. Foxman B, Barlow R, D'Arcy H, et al. Urinary tract infection: self-reported incidence and associated costs. *Ann Epidemiol*. 2000;10:509-515.
2. Foxman B. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *Am J Med*. 2002;113(suppl 1A):5S-13S.
3. Emmerson AM, Enstone JE, Griffin M, et al. The Second National Prevalence Survey of infection in hospitals: overview of the results. *J Hosp Infect*. 1996;32:175-190.
4. Neild GH. Urinary tract infection. *Medicine*. 2003;21:85-90.
5. Stamm WE, Hooton TM. Management of urinary tract infections in adults. *N Engl J Med*. 1993;329:1328-1334.
6. Christiaens TC, De Meyere M, Verschraegen G, et al. Randomised controlled trial of nitrofurantoin versus placebo in the treatment of uncomplicated urinary tract infection in adult women. *Br J Gen Pract*. 2002;52:729-734.
7. Fahey T, Webb E, Montgomery AA, et al. Clinical management of urinary tract infection in women: a prospective cohort study. *Fam Pract*. 2003;20:1-6.
8. SMAC. Standing Medical Advisory Committee (SMAC) report: the path of least resistance: occasional report. 1998. Available at http://www.advisorybodies.doh.gov.uk/smac1.htm. Accessed May 5, 2005.
9. Fenwick EA, Briggs AH, Hawke CI. Management of urinary tract infection in general practice: a cost-effectiveness analysis. *Br J Gen Pract*. 2000;50:635-639.
10. Colodner R, Rock W, Chazan B, et al. Risk factors for the development of extended-spectrum beta-lactamase-producing bacteria in nonhospitalized patients. *Eur J Clin Microbiol Infect Dis*. 2004;23:163-167.
11. Woodford N, Ward ME, Kaufmann ME, et al. Community and hospital spread of *Escherichia coli* producing CTX-M extended-spectrum beta-lactamases in the UK. *J Antimicrob Chemother*. 2004;54:735-743.
12. Smith HS, Hughes JP, Hooton TM, et al. Antecedent antimicrobial use increases the risk of uncomplicated cystitis in young women. *Clin Infect Dis*. 1997;25:63-68.
13. Graham JC, Galloway A. ACP Best Practice No. 167: the laboratory diagnosis of urinary tract infection. *J Clin Pathol*. 2001;54:911-919.
14. Smith PJ, Morris AJ, Reller BL. Predicting urine culture results by dipstick testing and phase contrast microscopy. *Pathology*. 2003;35:161-165.
15. Etherington IJ, James DK. Reagent strip testing of antenatal urine specimens for infection. *Br J Obstet Gynaecol*. 1993;100:806-808.
16. Rouse DJ, Andrews WW, Goldenberg RL, et al. Screening and treatment of asymptomatic bacteriuria of pregnancy to prevent pyelonephritis: a cost-effectiveness and cost-benefit analysis. *Obstet Gynecol*. 1995;86:119-123.
17. National Collaborating Centre for Women's and Children's Health. Antenatal care: routine care for the healthy pregnant woman: Clinical Guideline (NICE). 2003:81. Available at http://www.rcog.org.uk/index.asp?PageID=693. Accessed May 5, 2005.

DOI: 10.1309/C69RW1BT7E4QAFPV

18. Knottnerus JA, van Weel C, Muris JW. Evaluation of diagnostic procedures. *BMJ*. 2002;324:477-480.
19. Sackett DL, Haynes RB. The architecture of diagnostic research. *BMJ*. 2002;324:539-541.
20. European Confederation of Laboratory Medicine. European urinalysis guidelines. *Scand J Clin Lab Invest Suppl*. 2000;231:1-86.
21. Gorelick MH, Shaw KN. Screening tests for urinary tract infection in children: a meta-analysis. *Pediatrics*. 1999;104:e54. Available at http://pediatrics.aappublications.org/cgi/content/full/104/5/e54. Accessed July 16, 2006.
22. Huicho L, Campos-Sanchez M, Alamo C. Metaanalysis of urine screening tests for determining the risk of urinary tract infection in children. *Pediatr Infect Dis J*. 2002;21:1-11.
23. Hurlbut TA III, Littenberg B. The diagnostic accuracy of rapid dipstick tests to predict urinary tract infection. *Am J Clin Pathol*. 1991;96:582-588.
24. Deville WL, Yzermans JC, van Duijn NP, et al. The urine dipstick test useful to rule out infections: a meta-analysis of the accuracy. *BMC Urol*. 2004;4:4.
25. Gardner MJ, Altman DG. *Statistics With Confidence: Confidence Intervals and Statistical Guidelines*. London, England: BMJ Books; 1989.
26. Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. *BMJ*. 2001;322:1479-1480.
27. Deeks JJ, Altman DG, Bradburn MJ. Systematic reviews of evaluations of diagnostic and screening tests. In: Egger M, Davey Smith G, Altman DG, eds. *Systematic Reviews in Health Care: Meta-Analysis in Context*. London, England: BMJ Books; 2001:248-282.
28. Audurier A, Burdin JC, Darbas H, et al. Evaluation of a screening test for urinary infection [in French]. *Pathol Biol (Paris)*. 1988;36:921-924.
29. Bachman JW, Heise RH, Naessens JM, et al. A study of various tests to detect asymptomatic urinary tract infections in an obstetric population. *JAMA*. 1993;270:1971-1974.
30. Bowman RA, Riley TV. Evaluation of Clinitek 200 and Rapimat II/T for screening for urinary tract infection. *J Clin Pathol*. 1991;44:58-60.
31. Chernow B, Zaloga GP, Soldano S, et al. Measurement of urinary leukocyte esterase activity: a screening test for urinary tract infections. *Ann Emerg Med*. 1984;13:150-154.
32. Evans PJ, Leaker BR, McNabb WR, et al. Accuracy of reagent strip testing for urinary tract infection in the elderly. *J R Soc Med*. 1991;84:598-599.
33. Graninger W, Fleischmann D, Schneeweiss B, et al. Rapid screening for bacteriuria in pregnancy. *Infection*. 1992;20:9-11.
34. Hagay Z, Levy R, Miskin A, et al. Uriscreen, a rapid enzymatic urine screening test: useful predictor of significant bacteriuria in pregnancy. *Obstet Gynecol*. 1996;87:410-413.
35. Jellheden B, Norrby RS, Sandberg T. Symptomatic urinary tract infection in women in primary health care: bacteriological, clinical and diagnostic aspects in relation to host response to infection. *Scand J Prim Health Care*. 1996;14:122-128.
36. Lachs MS, Nachamkin I, Edelstein PH, et al. Spectrum bias in the evaluation of diagnostic tests: lessons from the rapid dipstick test for urinary tract infection. *Ann Intern Med*. 1992;117:135-140.
37. Lammers RL, Gibson S, Kovacs D, et al. Comparison of test characteristics of urine dipstick and urinalysis at various test cutoff points. *Ann Emerg Med*. 2001;38:505-512.
38. Lenke RR, Van Dorsten JP. The efficacy of the nitrite test and microscopic urinalysis in predicting urine culture results. *Am J Obstet Gynecol*. 1981;140:427-429.
39. McNair RD, MacDonald SR, Dooley SL, et al. Evaluation of the centrifuged and Gram-stained smear, urinalysis, and reagent strip testing to detect asymptomatic bacteriuria in obstetric patients. *Am J Obstet Gynecol*. 2000;182:1076-1079.
40. Monane M, Gurwitz JH, Lipsitz LA, et al. Epidemiologic and diagnostic aspects of bacteriuria: a longitudinal study in older women. *J Am Geriatr Soc*. 1995;43:618-622.
41. Pallares J, Casas J, Guarga A, et al. Rapid diagnostic methods as predictors of urinary infection in primary health care [in Spanish]. *Med Clin (Barc)*. 1988;91:775-778.
42. Preston A, O'Donnell T, Phillips CA. Screening for urinary tract infections in a gynaecological setting: validity and cost-effectiveness of reagent strips. *Br J Biomed Sci*. 1999;56:253-257.
43. Robertson AW, Duff P. The nitrite and leukocyte esterase tests for the evaluation of asymptomatic bacteriuria in obstetric patients. *Obstet Gynecol*. 1988;71:878-881.
44. Sewell DL, Burt SP, Gabbert NJ, et al. Evaluation of the Chemstrip 9 as a screening test for urinalysis and urine culture in men. *Am J Clin Pathol*. 1985;83:740-743.
45. Soisson AP, Watson WJ, Benson WL, et al. Value of a screening urinalysis in pregnancy. *J Reprod Med*. 1985;30:588-590.
46. Tissot E, Woronoff-Lemsi MC, Cornette C, et al. Cost-effectiveness of urinary dipsticks to screen asymptomatic catheter-associated urinary infections in an intensive care unit. *Intensive Care Med*. 2001;27:1842-1847.
47. Tuel SM, Meythaler JM, Cross LL, et al. Cost-effective screening by nursing staff for urinary tract infection in the spinal cord injured patient. *Am J Phys Med Rehabil*. 1990;69:128-131.
48. Lijmer JG, Bossuyt PMM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med*. 2002;21:1525-1537.
49. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med*. 1993;12:1293-1316.
50. Habbema JDF, Eijkemans R, Krijnen P, et al. Analysis of data on the accuracy of diagnostic tests. In: Knottnerus JA, ed. *The Evidence Base of Clinical Diagnosis*. London, England: BMJ Books; 2002:117-143.
51. Jaeschke R, Guyatt GH, Sackett DL, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, VI: how to use an article about a diagnostic test. B: what are the results and will they help me in caring for my patients? *JAMA*. 1994;271:703-707.
52. Boyd JC, Deeks JJ. Analysis and presentation of data. In: Price CP, Christenson R. *Evidence-Based Laboratory Medicine: From Principles to Outcomes*. Washington, DC: AACC Press; 2003:115-136.
53. Bent S, Nallamothu BK, Simel DL, et al. Does this woman have an acute uncomplicated urinary tract infection? *JAMA*. 2002;287:2701-2710.
54. Dobbs FF, Fleming DM. A simple scoring system for evaluating symptoms, history and urine dipstick testing in the diagnosis of urinary tract infection. *J R Coll Gen Pract*. 1987;37:100-104.

DOI: 10.1309/C69RW1BT7E4QAFPV