Developing clinical rules to predict urinary tract infection in primary care settings: sensitivity and specificity of near patient tests (dipsticks) and clinical scores

Paul Little, Sheila Turner, Kate Rumsby, Greg Warner, Michael Moore, J Andrew Lowes, Helen Smith, Catherine Hawke and Mark Mullee

INTRODUCTION

Acute urinary tract infection (UTI) is one of the most common acute bacterial infections among women.\(^1\),\(^2\) Conventional diagnosis relies on identifying a potential urinary pathogen from culture of a mid-stream specimen of urine (MSU) in a symptomatic patient. The standard for reporting UTI in most previous research and clinical practice was 10^5 colony-forming units per ml (cfu/ml).\(^3\) However, lower colony counts are associated with symptoms and respond to treatment;\(^4\) only 5% of low counts remit, the rest remain symptomatic; and 50% progress to high counts with symptoms.\(^5\)\(^6\) Although guidelines from the American Society of Microbiology and the European urinalysis guidelines\(^6\) have recently recommended reporting much lower colony counts (10^2 and 10^3 cfu/ml respectively), few studies have used these standards.

In clinical practice, the universal use of MSUs is not cost-effective and empiric antibiotic treatment is...
advocated.\(^7\) However, universal antibiotic use is associated with the growing problem of antibiotic resistance\(^8,9\) that has been identified in 20% of laboratory specimens. This study examines whether history and physical examination or near patient tests can be used for better diagnosis and targeting of antibiotics.

**Symptoms**
A recent systematic review identified nine studies that related symptoms and signs to diagnosis. However, significant limitations were documented.\(^3\)

- Few studies were identified with ≥50 consecutive patients and independent blind comparison of symptoms and signs with laboratory diagnosis among patients with suspected UTI. No such study had been conducted in primary care, and a sample of 50 patients is too small to be adequately powered for symptom prevalence of 20–70%;
- Predictive value depended on setting (for example, secondary care) and inclusion (for example, some studies included suspected vaginal infection);
- Only one study, which had poor methodologically, assessed the predictive value of combining symptoms;
- No study explored the implications of the severity of reported symptoms or used recent laboratory diagnosis standards.\(^6\)

**Near patient tests**
Dipsticks are the most widely used simple near patient tests in primary care.\(^10–13\) Summary data are available from studies that assessed nitrite and leucocyte esterase separately, but primary data are needed to assess the independent predictive value of all dipstick results.\(^14\) A systematic review suggested that the evidence base for dipstick use in primary care is poor, due to the paucity of studies and ‘spectrum bias’.\(^13,15\)

Evidence from emergency settings suggests that dipsticks may be particularly helpful where clinical assessment indicates a moderate probability of infection.\(^16\) Other studies from primary care have not assessed the independent value of dipstick results (hence over-complicating clinical decision rules), and/or mixed clinical and dipstick variables, and/or had low power.\(^11,13,17,18\) As with clinical studies, the current authors are not aware of any dipstick study that has used the recent guidelines of colony counts of 10³ cfu/ml.\(^6\) An adequately powered study was therefore needed:

- For women presenting in primary care with suspected UTI;
- To assess the independent predictive value of clinical symptoms, dipsticks results, and their combination;
- To develop clinical scoring methods for clinicians that are clear and determine the most predictive variables using multivariate methods;
- To use more sensitive laboratory gold standards.

**METHOD**

**Setting**
Between April 2002 and May 2003, 117 GPs and practice nurses from 67 practices in the south of England recruited 427 patients following written informed consent. Most doctors/nurses recruited only a few patients before stopping recruitment.

**Inclusion and exclusion criteria**
Adult female patients, aged 18 and over, with suspected UTI, usually due to a history of dysuria and frequency, were included in the study. Patients were excluded if other diagnoses were considered to be likely, for example, women with vaginal symptoms.\(^3\) Males, children, pregnant women, people aged over 70 years,\(^19–23\) and those with current severe mental problems (such as dementia) were also excluded.

**Data collection**
Structured clinical information was recorded by the clinician at the time of consultation. Patients were asked to rate each symptom according to severity: slight problem, moderately-severe problem, or a severe problem.\(^24,25\) The doctor or nurse documented whether an MSU was cloudy to the naked eye or smelled offensive,\(^11\) and performed a dipstick test (Bayer 8 SG) according to the manufacturer’s instructions.

**Laboratory analysis**
MSU was transported as in routine practice, and 10µl of MSU specimen were cultured onto cystine-lactose-electrolyte-deficient (CLED) agar and incubated overnight at 37°C.
to complete a questionnaire to be returned by post which documented demographics and past history, including past history of UTI.

**Sample size**

Sample size was calculated using NQuery sample size program (α = 0.05; 1-β = 0.8). Assuming that 50% of urine samples are infected,14 and that the prevalence of predictive variables is 20–70%, detecting an odds ratio (OR) of 2 required 403 patients. For sensitivity and specificity of 50–80%, 400 patients would estimate sensitivity or specificity with 95% confidence intervals (CIs) of ±6–7% (for 50%, 43.1 to 56.9%; for 80%, 74.5 to 85.5%); to achieve ±5% (that is, for 50%, 45 to 55%) would require 770 complete results.

**Analysis**

Developing clinical scores. Ordered categorical variables were dichotomised using cut-off points for an OR of 2 or close to 2; similar cut-offs were used for different symptoms to simplify any resultant clinical score. In multivariate logistic regression, significant variables were entered stepwise; they were retained if still significant at the 5% level and with ORs of 2 or near 2. All other variables were checked. Scores were based on simple counts of the rounded logistic coefficients using the coefficients from each separate model developed for each score (a clinical model, a dipstick model, and a combined model), and these determined the receiver operating characteristic (ROC) curve for each score.

Developing clinical prediction rules. The performance of each score for different cut-offs in the score was assessed to develop the best cut-off point for a clinical prediction rule. At each cut-off the following were determined: sensitivity, specificity, positive and negative predictive values (PPV and NPV), likelihood ratios (LRs) for a positive test (sensitivity/[1-specificity]), LRs for a negative test ([1-sensitivity]/specificity), and the number above the cut-off.

**RESULTS**

**Study population**

Less than 5% of eligible patients approached declined to participate. Of the 427 who agreed to participate, clinical information and laboratory tests were available for 408/427 (96%). Comparing patients from clinicians who recruited more than 10 patients (‘high recruiters’; n = 162) to those from lower recruiters (n = 246) showed no significant difference in the number of UTI diagnoses (65% versus 61%) or dipstick results. This suggests that major selection bias is unlikely.

---

**Rationale for laboratory diagnosis**

Laboratory evidence of bacterial growth is the best evidence of infection.14 Recent reports of intracellular infection remain of uncertain diagnostic significance.15 The American Society of Microbiology guidelines suggest reporting as low as 10^2 cfu/ml of *Escherichia coli*, whereas European urinalysis guidelines acknowledge the problems inherent in preventing the potential multiplication of bacteria in transit; they advocate reporting as low as 10^3 cfu/ml or pure growth of *E. coli*, and suggest reporting higher counts for more unusual organisms or mixed growths.6 The current study used the European guideline of 10^3 cfu/ml, but also reported the results with the standard of 10^2 cfu/ml used in the majority of previous evaluations of symptoms, signs and dipsticks.1

**Postal questionnaire**

Patients were asked at the recruitment consultation

---

### Table 1. Dipstick predictors of laboratory diagnosis of urinary tract infection according to European guidelines.

<table>
<thead>
<tr>
<th></th>
<th>UTI (n = 254)</th>
<th>No UTI (n = 154)</th>
<th>Crude odds ratio (95% CI)</th>
<th>Adjusted odds ratio (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nitrite (n = 28)</td>
<td>72 (28)</td>
<td>7 (9)</td>
<td>8.31 (3.71 to 18.6)</td>
<td>6.36 (2.77 to 14.6)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Leucocyte (+ or greater)</td>
<td>217 (85)</td>
<td>72 (47)</td>
<td>6.68 (4.17 to 10.7)</td>
<td>4.52 (2.72 to 7.50)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Blood (haemolyzed trace or greater)</td>
<td>186 (73)</td>
<td>71 (46)</td>
<td>3.20 (2.10 to 4.87)</td>
<td>2.23 (1.38 to 3.61)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Protein (+ or greater)</td>
<td>119 (47)</td>
<td>47 (31)</td>
<td>2.00 (1.32 to 3.06)</td>
<td>1.12 (0.69 to 1.83)</td>
<td>0.643</td>
</tr>
</tbody>
</table>

*10^3 colony-forming units per ml. Adjusted mutually for other variables in the model (nitrite, leucocyte and blood). UTI = urinary tract infection.

### Table 2. Dipstick decision rule performance in predicting laboratory diagnosis of urinary tract infection according to European guidelines.

<table>
<thead>
<tr>
<th></th>
<th>Test</th>
<th>Dipstick rule-</th>
<th>Dipstick rule+</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>UTI-</td>
<td></td>
<td>108</td>
<td>46</td>
<td>154</td>
</tr>
<tr>
<td>UTI+</td>
<td>58</td>
<td>196</td>
<td>254</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>166</td>
<td>242</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Dipstick decision rule is based on having either nitrite, or blood and leucocytes. *10^3 colony-forming units per ml. Dipstick negative = neither nitrite, nor leucocyte and blood combined. Sensitivity = 196/254 (77%); 95% CI = 72.0 to 82.4. Specificity = 108/154 (70%); 95% CI = 62.9 to 77.3. Positive predictive value = 196/242 (81%); 95% CI = 76.1 to 85.9. Negative predictive value 108/186 (65%); 95% CI = 57.8 to 72.4%. Likelihood ratio for a positive test = 2.58 (95% CI = 2.01 to 3.32). Likelihood ratio for a negative test = 0.33 (95% CI = 0.25 to 0.42). UTI = urinary tract infection.
The mediantime between test and standard was 6 hours. Time did not predict diagnosis (time in hours OR = 1.01, 95% CI = 0.99 to 1.02,  
$z = 0.95$, $P = 0.34$). High colony counts were detected in $177/408$ (43%; $\geq 10^5$ cfu/ml), and $254$ (62%) when the more rigorous criteria of low colony counts ($\geq 10^3$ cfu/ml) was used according to European urinalysis guidelines. 

Demographic questionnaires were returned by $270/408$ (63%) participants. There was no significant difference between those who did and did not return the questionnaire (laboratory diagnosis of UTI 65%, 58%; nitrite 20%, 18%; leucocytes 74%, 66%; urine cloudy 38%, 34%; moderately severe dysuria 60%, 60%, respectively). Of the 270 returned questionnaires, 195 (72%) reported a previous UTI, 150 (56%) were married, 174 (64%) were in employment or at college, and 172 (64%) reported having some educational qualifications. These demographics are similar to national attending samples. 

**Dipstick testing**

Three variables independently predicted diagnosis: nitrite was most predictive followed by leucocytes and blood (Table 1). A cut-off score of 2 or more based on the sum of the rounded logistic coefficients — equivalent to a clinical decision rule based on patients having either nitrite or leucocyte and blood — had sensitivity of 77%, and specificity of 70% (see Tables 2 and 3). Each end of the score could be used to improve performance by varying the cut-off point. Thus the NPV was 73% (LR-ve test 0.22) for having none of dipstick nitrite, blood, or leucocyte esterase, and the PPV was 92% (LR+ve test 7.2) for having nitrite and either blood or leucocyte esterase (Table 3).

**Clinical variables**

Four variables independently predicted UTI (Table 4): cloudy urine, offensive smell, and dysuria and/or nocturia of moderate severity. Severity was an important aspect of prediction: symptoms rated as slight problems were much less predictive. A cut off of 2 or more of a score based on the sum of the rounded logistic coefficients (a clinical decision rule based on 2 out of 4 features) had sensitivity 65% and specificity 69% (see Tables 5 and 6). Each end of the score could be used to improve performance by varying the cut-off point. Thus the NPV was 71% for none of the four clinical features, and the PPV 84% for three or more features (see Table 6).

**Implications of other approaches**

The performance of the scores was not improved by combining dipstick and clinical variables (Supplementary Table 1), by using a sequential approach to the use of dipsticks (reserving dipsticks for those with intermediate clinical scores), or by using a different laboratory standard (Supplementary Tables 2 and 3).

**DISCUSSION**

**Summary of main findings**

This study shows the potential and the limitations of using dipstick testing and clinical information in practice to predict laboratory diagnosis. A dipstick decision rule was developed based on having nitrite, or both leucocytes and blood, which was moderately sensitive (77%) and specific (70%) but
had a moderately low NPV (65%). The predictive values were improved by varying the cut-off point: NPV was 73% for all three dipstick results being negative, and PPV was 92% for having nitrite and either blood or leucocyte esterase. A clinical decision rule was also developed based on having two of the following: urine cloudiness, offensive smell, and dysuria and/or nocturia of moderate severity. The clinical decision rule was less sensitive than the dipstick decision rule (65%) and had a lower NPV (54%). The predictive value of the clinical decision rule could be improved by modifying the cut-off point: for none of the four clinical features NPV was 71%, and for three or more features PPV was 84%. When using these rules in practice, clinicians need to use appropriate strategies to take into account relatively low NPV; that is, the lower proportion of negative UTI results that are correctly diagnosed.

**Strengths and limitations of the study**

**Strengths.** This is the first adequately powered study to assess the independent predictive value of dipstick results and of clinical symptoms in a primary care sample. The sample had similar characteristics to UK national attending samples and an incidence of UTI similar to previous primary care studies. Patients were included for whom UTI was the suspected diagnosis.

**Limitations.** Results from this study may not apply to other groups (for example, where vaginal or urinary infection is suspected). There was variability in transit time to the laboratory, but there was no evidence this affected the likelihood of laboratory diagnosis. Although multiple variables were used in developing the models, type I error is less likely as the results were highly significant for most variables. The performance of clinical decision rules in the same population was estimated. Further prospective validation is required.

**Comparison with the existing literature**

Clinical variables that predict laboratory diagnosis. Four clinical variables independently predicted diagnosis: cloudy urine, offensive smell, and dysuria and/or nocturia of moderate severity. It appears that this study is the only adequately powered ‘level I’ study to date of women with presumed UTI to identify the independent predictive value of symptoms, and uses lower colony count as a better laboratory standard. The finding that duration of symptoms for one day predicted diagnosis in a moderately sized study (n = 231) in primary care could not be confirmed. Key findings in comparison with similar literature were that:

- Symptoms severity is important and the presence of symptoms is less predictive.
- An examination of urine for cloudiness and smell provides important information.
The use of low colony counts improves prediction of UTI, which supports the validity of lower counts. If low colony counts provided non-differential measurement error, predictive values would be worse when low colony counts were included as part of the gold standard.

Three key dipstick variables independently predict laboratory diagnosis: nitrite, leucocytes, and blood. A dipstick decision rule performed slightly better than a clinical decision rule. Previous studies in primary care have had limited power or did not assess the independent value of dipstick results using multivariate analysis. These findings demonstrate the importance of multivariate analysis and contradict previous findings about protein which did not independently predict UTI in the current study. The dipstick decision rule performed marginally better than the clinical decision rule, and dipsticks have the potential to target treatment and lower costs depending on the strategy used (see below). Results also suggest significant limitations in the performance of urinalysis, particularly, lower than expected sensitivity and NPVs.

**Implications for clinical practice**

Given the current debate about the appropriateness of antibiotics for uncomplicated UTI, there are likely to be different opinions on how to use clinical or dipstick decision rules. The main limitation is the number of women with UTI that are ‘missed’, as in this study 35% (n = 90), and of these 38% (n = 34) had low colony counts for the clinical decision rule.

Most women with symptoms of cystitis do not contact a health professional and can treat themselves conservatively. Placebo groups of randomised controlled trials suggest that women not treated with antibiotics mostly get better (albeit more slowly), suffer complications rarely, and will not suffer greater recurrence. Thus the utility of a clinical decision rule is not that it can perfectly target antibiotics (which is not strictly necessary) but that it can target antibiotics more appropriately rather than either empiric treatment or self management, and that it is less likely to encourage belief in the importance of seeing the doctor when compared with routinely performing MSUs in all patients. A clinical decision rule could also be potentially useful as part of telephone or internet-based triage. Given the moderately low sensitivity of the rule, a reasonable approach would be to advise women who have less than two of the four features to return if their symptoms are not settling with conservative treatment, or to offer a backup (delayed) prescription of antibiotics, as is used for respiratory infection. For dipsticks, a reasonable approach would be to ask women with negative dipstick results to return if their symptoms are not settling, or to provide a delayed prescription. Such pragmatic strategies require further testing in randomised controlled trials.

**Maximising predictive value: varying the cut-off points.** Clinicians may wish to vary their threshold for empiric management using the cut-off points at either extreme of the clinical scores. If dipstick testing reveals none of nitrite, blood or leucocytes, UTI is unlikely (NPV 73%; LR-ve test 0.22) and symptomatic advice and/or a delayed prescription would be reasonable. For those with nitrite and either blood or leucocytes, UTI is very likely (PPV 92%; LR+ve test 7.2) and empiric antibiotics are sensible.
The remaining patients could be targeted for investigation and/or given a delayed prescription. A similar strategy could be used for the clinical score, with symptomatic advice for patients having none of the four features (NPV 71%) and empiric antibiotics for those with three or more features (PPV 84%).

Simple decision rules could improve targeting of investigation and treatment. Strategies to use decision rules need to take account of their limited sensitivity and negative predictive value, which is lower than expected from previous research. Research is needed to confirm the validity of these findings in a separate sample.

Supplementary information
Additional information accompanies this article at http://www.rcgp.org.uk/Default.aspx?page=2482

Funding
This study is funded by the Health Technology Programme of UK NHS Research and Development (reference 97/14/08)

Ethics committee
South West MReC ethical committee approval (reference 03/6/11)

Competing interests
All authors except JAL have no competing interests and therefore have nothing to declare. JAL has been paid to

of UK NHS Research and Development (reference 97/14/06)

Research is needed to confirm the validity of these

findings in a separate sample.

REFERENCES