Validating the prediction of lower urinary tract infection in primary care: sensitivity and specificity of urinary dipsticks and clinical scores in women

Paul Little, Sheila Turner, Kate Runnby, Rachel Jones,* Greg Warner, Michael Moore, J Andrew Lowes, Helen Smith, Catherine Hawke, Geraldine Leydon, and Mark Mullee

ABSTRACT

Background
Dipsticks are one of the most commonly used near-patient tests in primary care, but few clinical or dipstick algorithms have been rigorously developed.

Aim
To confirm whether previously documented clinical and dipstick variables and algorithms predict laboratory diagnosis of urinary tract infection (UTI).

Design of study
Validation study.

Setting
Primary care.

Method
A total of 434 adult females with suspected lower UTI had bacteriuria assessed using the European Urinalysis Guidelines.

Results
Sixty-six percent of patients had confirmed UTI. The predictive values of nitrite, leucocyte esterase (+ or greater), and blood (haemolysed trace or greater) were confirmed (independent multivariate odds ratios = 5.6, 3.5, and 2.1 respectively). The previously developed dipstick rule — based on presence of nitrite, or both leucocytes and blood — was moderately sensitive (75%) but less specific (66%; positive predictive value [PPV] 81%, negative predictive value [NPV] 57%). Predictive values were improved by varying the cut-off point: NPV was 76% for all three dipstick results being negative; the PPV was 92% for having nitrite and either blood or leucocyte esterase. Urine offensive smell was not found to be predictive in this sample; for a clinical score using the remaining three predictive clinical features (urine cloudiness, dysuria, and nocturia), NPV was 67% for none of the features, and PPV was 82% for three features.

Conclusion
A clinical score is of limited value in increasing diagnostic precision. Dipstick results can modestly improve diagnostic precision but poorly rule out infection. Clinicians need strategies to take account of poor NPVs.

Keywords
algorithms, clinical scoring; diagnosis, urinary tract infection; primary care; urinalysis.

INTRODUCTION
Acute urinary tract infection (UTI) is one of the most common acute bacterial infections among adult females.1,2

Empiric antibiotic treatment has been advocated as cost-effective,* but unselective antibiotic use will result in a growing problem of antibiotic resistance, which has been identified in 20% of laboratory specimens.4,5 There are current proposals to make courses of trimethoprim available over the counter. This potential overuse of antibiotics for UTI creates an urgent need to address the question: can we use history and physical examination, or near-patient tests, for better diagnosis and targeting of antibiotics?

A systematic review of the role of symptoms in diagnosis identified few high-grade studies (those with ≥50 consecutive patients, and independent blind comparison of symptoms and signswith a gold standard among patients with suspected UTI), none

*The authors would like to dedicate this paper to Dr Rachel Jones who recently died unexpectedly.

P Little, MD, FRCP; MRCP; professor; S Turner, senior research fellow; K Runnby, BA, R Jones, H Smith, MSc, MD, MRCGP, FFPHM, professor; M Mullee, MSc, CStat; G Leydon, senior research fellow; Community Clinical Sciences Division (CCS), Southampton University, Southampton. G Warner, MA, MRCGP, FCH, DRCOG, DSEM, GP, Nightingale Surgery, Ramsey. M Moore, BMed Sci, MSc, MRCP, MRCGP, GP; Three Swans Surgery, Salisbury. JA Lowes, BSc, MRCGP, FRCP, Southampton Health Protection Agency Laboratory, Southampton. C Hawke, public health physician, Hastings and St Leonard Primary Care Trust, St Leonards-on-Sea.

Address for correspondence
Professor P Little, University of Southampton, Aldermoor Health Centre, Aldermoor Close, Southampton, SO16 5ST.
E-mail: p.little@soton.ac.uk


DOI: 10.3399/bjgp10X514747
of which were based in primary care. Only one study, which had poor methodology, assessed the predictive value of combining symptoms; and all previous studies used a very insensitive cut-off for laboratory diagnosis, whereas recent laboratory diagnosis standards suggest colony counts down to 10^3 cfu/ml (colony-forming units per millilitre) are classified as UTI. Clinicians commonly use dipsticks to rule out infection, and they are the most widely used simple near-patient test in primary care. Although summary data are available for studies that assessed nitrite and leucocyte esterase separately, primary data are needed to assess the independent predictive value of all dipstick results. The evidence base for dipstick use in primary care is poor, due to the paucity of studies and ‘spectrum bias’. Studies from primary care have a range of one or more limitations: they have either not assessed the independent value of dipstick results and symptoms (hence potentially over-complicating clinical decision rules); and/or not used the range of dipstick variables (most include nitrite and leucocyte but not blood); and/or failed to develop and then test algorithms in separate samples (McIsaac et al being the exception); and/or had low power. Only the most recent dipstick studies have used the recent more rigorous laboratory guidelines for diagnosis.

The current authors have reported a study where a clinical score and a dipstick score were developed for women presenting with suspected UTI. The independent predictive values of symptoms and of dipsticks results were assessed. Based on accumulating evidence and recent international consensus, the latter study used more sensitive laboratory gold standards to include lower colony counts. However, the predictive value of any scoring system that is tested in the same sample used to calculate the scoring system is likely to have artificially inflated predictive values: both a training and a validation set are needed. To estimate the more realistic predictive values of these scores, this study assessed the predictive value of the scores and the component variables of the scores in a new validation sample.

**METHOD**

**Setting**
Between January 2002 and February 2005, 117 primary care clinicians (doctors or practice nurses) from 62 practices in the south of England recruited 434 patients following informed, written consent. The clinicians recruited consecutive patients, and most recruited only a few patients before stopping recruitment.

**Data collection**
Structured clinical information was recorded by the clinician at the time of consultation. Patients were asked to rate each symptom they experienced as a slight problem, a moderately severe problem, or a severe problem. Midstream urine (MSU) was examined for cloudiness to the naked eye and offensive smell prior to performing a dipstick test (Bayer Multistix® 8 SG), which was performed according to the manufacturer’s instructions.

**Inclusion criteria**
Inclusion criteria were adult female patients (aged 18 to 70 years) with suspected UTI — which in practice usually meant patients with a history of dysuria and frequency. Exclusions were where other diagnoses were considered likely; (for example, patients with vaginal symptoms); pregnant women; age over 70 years since the relation between symptoms and bacteriuria is likely to be different in older age groups; and also current severe mental problems (for example, dementia) where patients would have difficulty with consenting and answering questions.

**Laboratory analysis**
MSU was transported as in routine practice (which takes normally 1 day to reach the lab); 10 µl of MSU was cultured onto cystine-lactose-electrolyte-deficient agar and incubated overnight at 37°C. The European Urinalysis Guidelines were used as the main reference standard. MSU was analysed by individuals who were blind to the clinic dipstick results.

**Rationale for laboratory diagnosis**
The American Society for Microbiology guidelines suggest reporting down to 10^3 cfu/ml of *E. coli*; however, European Urinalysis Guidelines acknowledge the problem of preventing multiplication of bacteria in transit, and advocate reporting counts down to 10^3 cfu/ml, or pure growth of *E. coli*, and
suggest using higher counts for more unusual organisms or mixed growths.\(^7\)

Postal questionnaire
The postal questionnaire asked patients about their demographics and medical history (including past history of UTI).

Sample size
For the sample size \((\alpha = 0.05; \beta = 0.2; \text{NQuery sample size programme})\) it was assumed 50% of urine samples are infected,\(^4\) and the prevalence of predictive variables is 20–70%; to detect an odds ratio (OR) of 2 required 403 patients. For a sensitivity or specificity of 50% a total of 400 patients will provide 95% confidence intervals [CIs] of 43% to 57%, and for 80% sensitivity or specificity 74% to 86%.

Analysis
The variables found to be predictive from the previous study were assessed in multivariate logistic regression using Stata 9. Previously developed clinical scores were also assessed by cross-tabulation; any new scores were recalculated based on simple counts of the rounded logistic coefficients, and the receiver operator curve was determined for each score. Performance of each score for different cut-offs in the score was assessed. At each cut-off, the sensitivity, specificity, positive and negative predictive values, 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 were determined.

RESULTS

Study population
Less than 10% of eligible patients who were approached declined to participate. Of the 434 who agreed to participate, dipstick information was available for 429 (99%), and clinical information for 431 (99%); 219 (50%) patients were found to have high colony counts \((\geq 10^5 \text{ cfu/ml})\), and 287 (66%) fulfilled the more sensitive criteria of lower colony counts \((\geq 10^3 \text{ cfu/ml})\) according to European Urinalysis Guidelines.\(^7\) Of the 269 patients who returned the demographic questionnaire, 200 (74%) reported a previous UTI, 152 (57%) were married, and 152 (57%) were reported as having an educational qualification of GCSE (secondary intermediate qualification) or equivalent.

Assessing the predictive value of dipstick variables
Nitrites were found to be most predictive of UTI, followed by blood and then leucocytes (Table 1), with ORs very similar to those of the previous derivation

| Table 1. Dipstick predictors of laboratory diagnosis of urinary tract infection (UTI). |
|-----------------|----------------|----------------------|----------------|----------------|-----------------|-----------------|
|                 | UTI, n (%)     | No UTI, n (%)        | Likelihood ratios (LR+, LR–) | Crude odds ratio (95% CI) | Adjusted odds ratio* (95% CI) | P-value |
| Nitrite         | 98 (35)        | 9 (6.2)              | 5.5, 0.7                        | 8.0 (3.9 to 16.3)         | 5.6 (2.7 to 11.7)              | <0.001         |
| Leucocyte: + or greater | 246 (87) | 77 (53)              | 1.6, 0.3                        | 5.7 (3.6 to 9.2)          | 3.5 (2.1 to 5.8)               | 0.002          |
| Blood: haemolysed trace or greater | 205 (72) | 61 (42)              | 1.7, 0.5                        | 3.6 (2.4 to 5.4)          | 2.1 (1.3 to 3.4)               | <0.001         |

*Adjusted mutually for other significant variables in the final model (these were nitrite, leucocyte, and blood, but not protein).

| Table 2. Validation of dipstick score to predict laboratory diagnosis of urinary tract infection (UTI).* |
|-----------------|----------------|-----------------|----------------|-----------------|-----------------|----------------|
| Cut-off point on dipstick score (% at or above cut-off point) | Sensitivity, % | Specificity, % | PPV, % | NPV, % | % correctly classified | LR +ve test | LR –ve test |
| ≥0 (100)        | 100            | 0               | –     | –    | 66              | 1              | –             |
| ≥1 (84)         | 94             | 37              | 74    | 76   | 75              | 1.5            | 0.2           |
| ≥1.5 (77)       | 89             | 46              | 76    | 68   | 75              | 1.7            | 0.2           |
| ≥2 (61)         | 75             | 66              | 81    | 57   | 72              | 2.2            | 0.4           |
| ≥2.5 (60)       | 74             | 67              | 81    | 56   | 71              | 2.2            | 0.4           |
| ≥3 (24)         | 33             | 94              | 92    | 42   | 54              | 6.1            | 0.7           |
| ≥3.5 (23)       | 32             | 94              | 92    | 42   | 53              | 5.8            | 0.7           |
| ≥4.5 (19)       | 26             | 95              | 91    | 40   | 49              | 5.3            | 0.8           |
| >4.5 (0)        | 0              | 100             | –     | –    | 34              | –1             | 1             |

*Score weighted according to the rounded logistic coefficients based on the sum of nitrite = 2, leucocyte = 1.5, blood = 1. LR = likelihood ratio. NPV = negative predictive value. PPV = positive predictive value.
The previously developed dipstick rule — based on having nitrite, or both leucocytes and blood — was moderately sensitive at 75% (95% CI = 71% to 78%) but less specific (66%; 95% CI = 60% to 72%) with a positive predictive value (PPV) of 81%, (95% CI = 0.7 to 2.1) and a negative predictive value (NPV) of 57% (95% CI = 0.7 to 1.5).

Predictive values were improved by varying the cut-off point: NPV was 76% (95% CI = 66% to 84%) for all three dipstick results being negative (see Table 2: cut-off point ≥1); PPV was 92% (95% CI = 86% to 96%) for having nitrite and either blood or leucocyte esterase (see Table 2: cut-off point ≥3).

**Clinical variables**

Only two of the original four predictive variables that were found to predict laboratory diagnosis from the derivation sample independently predicted UTI: cloudy urine, and dysuria rated as a moderately severe problem.

**Table 3. Clinical predictors of laboratory diagnosis of urinary tract infection (UTI).**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>UTI, n (%)</th>
<th>No UTI, n (%)</th>
<th>Likelihood ratios (LR+, LR-)</th>
<th>Crude odds ratio (95% CI)</th>
<th>Adjusted odds ratio (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urine cloudy on examination</td>
<td>141 (49)</td>
<td>39 (27)</td>
<td>1.8, 0.7</td>
<td>2.6 (1.7 to 4.1)</td>
<td>2.5 (1.6 to 3.9)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Urine smell offensive on examination</td>
<td>82 (29)</td>
<td>28 (19)</td>
<td>1.5, 0.9</td>
<td>1.7 (1.0 to 2.7)</td>
<td>1.2 (0.7 to 2.1)</td>
<td>0.560</td>
</tr>
<tr>
<td>Dysuria: reported a moderately severe problem</td>
<td>189 (66)</td>
<td>70 (48)</td>
<td>1.4, 0.7</td>
<td>2.1 (1.4 to 3.1)</td>
<td>2.0 (1.3 to 3.0)</td>
<td>0.001</td>
</tr>
<tr>
<td>Nocturia: reported a moderately severe problem</td>
<td>133 (47)</td>
<td>64 (44)</td>
<td>1.0, 0.9</td>
<td>1.10 (0.7 to 1.6)</td>
<td>0.99 (0.7 to 1.5)</td>
<td>0.960</td>
</tr>
<tr>
<td>Any nocturia</td>
<td>224 (78)</td>
<td>98 (68)</td>
<td>1.2, 0.7</td>
<td>1.7 (1.1 to 2.7)</td>
<td>1.60 (1.0 to 2.6)</td>
<td>0.047</td>
</tr>
</tbody>
</table>

*Adjusted mutually for other variables in the model (cloudy urine, dysuria, nocturia). The estimate for any nocturia quoted above is adjusted for cloudy urine and moderately bad dysuria; if any nocturia and any dysuria are included in the model, for simplicity the estimates are: cloudy urine 2.4 (95% CI = 1.5 to 3.8); nocturia 1.6 (95% CI = 1.00 to 2.5); dysuria 2.7 (95% CI = 1.6 to 4.4).

**Table 4. Validation of clinical score to predict laboratory diagnosis of urinary tract infection (UTI).**

<table>
<thead>
<tr>
<th>Cut-off point on dipstick score (% at or above cut-off point)</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
<th>PPV, %</th>
<th>NPV, %</th>
<th>% correctly classified</th>
<th>LR +ve test</th>
<th>LR –ve test</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥0 (100)</td>
<td>100</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>66</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>≥1 (86)</td>
<td>91</td>
<td>22</td>
<td>70</td>
<td>54</td>
<td>68</td>
<td>1.2</td>
<td>0.4</td>
</tr>
<tr>
<td>≥2 (57)</td>
<td>65</td>
<td>59</td>
<td>76</td>
<td>46</td>
<td>63</td>
<td>1.5</td>
<td>0.6</td>
</tr>
<tr>
<td>≥3 (24)</td>
<td>28</td>
<td>83</td>
<td>67</td>
<td>37</td>
<td>47</td>
<td>1.7</td>
<td>0.9</td>
</tr>
<tr>
<td>≥4 (6)</td>
<td>7</td>
<td>97</td>
<td>83</td>
<td>35</td>
<td>37</td>
<td>2.5</td>
<td>1</td>
</tr>
<tr>
<td>&gt;4 (0)</td>
<td>0</td>
<td>100</td>
<td>–</td>
<td>–</td>
<td>34</td>
<td>–1</td>
<td>1</td>
</tr>
</tbody>
</table>

*Score weighted according to the rounded logistic coefficients based on the sum of: urine cloudiness = 1, urine smell = 1, moderately severe dysuria = 1, moderately severe nocturia = 1. LR = likelihood ratio. NPV = negative predictive value. PPV = positive predictive value.

**Table 5. Modified clinical score based on cloudy/burning any degree/night frequency any degree.**

<table>
<thead>
<tr>
<th>Cut-off point on dipstick score (% at or above cut-off point)</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
<th>PPV, %</th>
<th>NPV, %</th>
<th>% correctly classified</th>
<th>LR +ve test</th>
<th>LR –ve test</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥0 (100)</td>
<td>100</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>66</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>≥1 (96)</td>
<td>98</td>
<td>8</td>
<td>68</td>
<td>67</td>
<td>68</td>
<td>1.1</td>
<td>0.3</td>
</tr>
<tr>
<td>≥2 (71)</td>
<td>80</td>
<td>46</td>
<td>74</td>
<td>54</td>
<td>69</td>
<td>1.5</td>
<td>0.4</td>
</tr>
<tr>
<td>≥3 (29)</td>
<td>36</td>
<td>84</td>
<td>82</td>
<td>40</td>
<td>52</td>
<td>2.3</td>
<td>0.8</td>
</tr>
<tr>
<td>&gt;3 (0)</td>
<td>0</td>
<td>100</td>
<td>–</td>
<td>–</td>
<td>34</td>
<td>–1</td>
<td>1</td>
</tr>
</tbody>
</table>

*Score weighted according to the rounded logistic coefficients based on the sum of: urine cloudiness = 1, burning dysuria any degree = 1, nocturia any degree = 1. LR = likelihood ratio. NPV = negative predictive value. PPV = positive predictive value.
severe problem (Table 3). Moderately severe nocturia and offensive smell of urine were no longer significant. The original clinical decision rule from the derivation sample based on two or more of the above features was now found to have a sensitivity of 65% (95% CI = 62% to 68%), which was previously 65%; and a specificity of 59% (95% CI = 53% to 65%), which was previously 69% (see Table 4; cut-off point ≥2).

However, as the presence of nocturia to any degree was independently predictive (OR = 1.60, 95% CI = 1.01 to 2.55), a modified score was assessed so that simply the presence of the symptoms nocturia and dysuria were included without the need for severity rating. This resulted in increased sensitivity, but NPVs were still poor (Table 5): NPV was 67% for none of the features, and the PPV was 82% for three features.

DISCUSSION
Summary of main findings
This study confirms both the potential and the limitations of using dipstick and clinical information in routine clinical practice to predict laboratory diagnosis.

The previous clinical decision rule did not perform as well as in this derivation sample. PPVs remained quite similar to those found in the derivation study, but the NPV was poor. Not all the variables found to be predictive in the first study were as predictive in this study (urine smell was not found to be predictive in this sample). However, even using a modified score (Table 5) based on the variables confirmed to be predictive in this study (cloudiness, dysuria, nocturia) did not greatly improve the predictive values. The implications of this for practice are that clinicians can be reasonably confident that patients with suspected UTI who have dysuria, nocturia, and cloudy urine do have UTI, but they should be cautious about excluding patients based on the absence of these features.

Three dipstick variables identified previously to be most independently predictive of UTIs in the derivation sample were tested in a new dataset by multivariate analysis, and the multivariate ORs were similar to the previous study. The dipstick score performed significantly better than the clinical score. At a cut-off point in the score of ≥2 (equivalent to having nitrite, or both leucocytes and blood), both the sensitivity and specificity of the score was very similar to that found previously, as was PPV, but the NPV decreased from 65% (derivation sample) to 57% in this sample.

Although the predictive values could be improved by varying the cut-off points, the NPVs remained low. Thus, in practice, clinicians cannot rule out the diagnosis of UTI using either clinical information or dipstick results, and will need to use appropriate strategies, such as delayed prescription, to take account of the relatively low NPVs.

Strengths and limitations of the study
To the authors’ knowledge, this study is the first to confirm the predictive value of a rigorously developed dipstick algorithm and of a clinical score in an adequately powered primary care sample, and the first to combine this with use of a rigorous laboratory standard for diagnosis. The recommended group was chosen, that is, those patients where UTI was the suspected diagnosis, and the sample had similar characteristics to UK national attending samples. The sample had a similar incidence of UTI to the authors’ previous derivation study and to previous primary care studies.

Limitations include the fact that multiple variables were used in developing the models, but type I error is less likely since the results were highly significant for most variables that were tested, and this is the second sample in which these findings apply. Results from this study may not apply to other groups (for example, where vaginal infection is suspected).

Comparison with existing literature
Two recent studies in primary care have used the more rigorous lower colony count standards for the diagnosis of UTI. The validation study by McIsaac et al, which combined clinical information and dipstick results (two or more of dysuria, leucocytes, and nitrites) did not weight these variables, did not use dipstick haematuria, and demonstrated modest predictive values (LR +ve test = 1.73 and LR –ve test = 0.43). This supports the present authors’ previous findings that clinical information is unlikely to add greatly to the predictive value when all three predictive variables from dipsticks are used.

The study by Hummers-Pradier et al had similar limitations (it did not use dipstick haematuria, did not weight variables, and had no separate derivation and validation samples) and also found modest predictive values.

However, even though the present study has confirmed modestly better predictive values, they are still far from optimal. Predictive values can be maximised by varying the cut-off points in the score: with all three dipstick variables being negative, it would be reasonable to say that UTI would be unlikely (NPV = 76%); however, even with this higher NPV, 24% of patients would be told they have no UTI when in fact they do.
Implications for clinical practice and future research

The pattern of clinical information in suspected UTI is of limited value in increasing diagnostic precision: although UTI is likely among patients with dysuria, nocturia, and urine cloudiness, the absence of these features performs poorly in ruling out UTI. A dipstick rule does improve diagnostic precision; but in applying the results of dipsticks, clinicians will still need to take account of the limited NPVs, which are low: even when all results are negative, 24% of female patients will still have UTI. This means that in practice clinicians should consider using strategies such as delayed prescribing for such patients, or alternatively advising a review consultation, if symptoms are not settling. Research into the practice implications of such strategies is also needed.

Funding body
This study was funded by the Health Technology Programme of UK NHS Research and Development (reference 97/14/06).

Ethical approval
The study had South West MREC ethical committee approval (reference 03/6/11).

Competing interests
All authors except J Andrew Lowes have no competing interests and therefore have nothing to declare. J Andrew Lowes has been paid to attend consultancy workshops by Bayer and is currently working in collaboration with Bayer in an unpaid capacity.

Acknowledgements
We are very grateful for the time given by GPs and nurses and all the patients who agreed to take part. We are also grateful to Bayer for provision of the BSG strips.

Discuss this article
Contribute and read comments about this article on the Discussion Forum: http://www.rcgp.org.uk/bjgp-discuss

REFERENCES