The problems affecting the diagnosis of urinary tract infection

Temenuga Dacheva* & James Malone-Lee*

A most important contemporary challenge of medicine in the elderly is the diagnosis of urinary tract infection. Recently, a number of studies have challenged the validity of the common, routine urinalysis methods. The evidence implies that there may be a large pool of significant disease that is overlooked. This situation occurs at a time when the use of antibiotics is subject to much censure. Thus the proposition that widespread, inappropriately untreated, urinary infection may exist, does not sit comfortably with the zeitgeist. This article will examine the evidence that contradicts many widely accepted standards with respect to the diagnosis of urinary tract infection.

The practical implications

Current evidence implies that assessment of a patient in relation to urinary tract infection (UTI) should depend more on the symptoms and signs than the investigations. In either sex, hesitancy, reduced stream, intermittency and terminal dribbling seem to be the most sensitive symptoms of infection, more so than pain. The important signs are suprapubic tenderness, loin tenderness and a uriniferous odor [1].

There is now evidence implying that judgments of the appropriateness of treatment for UTI must rely more on the symptomatic response than culture data. Treatment cessation must be conditioned on vigilance for symptomatic relapse, regardless of urinalysis or course length recommendations. Urine infections can be tenacious and difficult to eradicate.

It is probable that the implications for diagnosis and treatment arising from contemporary discoveries will prove troubling, antagonistic are sometimes infuriating to many clinicians. It will take years to resolve the conflicts that must surely arise.

The context as it applies to the elderly

‘Lower urinary tract symptoms’ (LUTS) is a collective term that includes storage symptoms, such as frequency, urgency, urge incontinence; symptoms of stress urinary incontinence; voiding symptoms such as hesitancy, reduced stream and intermittency; and finally sensory symptoms that include various degrees and expressions of pain. The prevalence (Pr) of LUTS increases with age and is reported in up to 40% of men and 28% of women aged 70–79 years [2,3].

UTIs are among the most prevalent infectious diseases worldwide, leading to significant morbidity and mortality and wreaking a substantial economic cost [4]. Amongst the elderly, UTIs are the most common bacterial infection overall [5]. UTI is now incriminated in the etiology of all of the LUTS, most notably overactive bladder and voiding symptoms, these being particularly troublesome in older people. Regrettably, UTIs are often missed when patients present in GP surgeries, clinics and hospitals [6–8]. Untreated UTI has the potential to develop into chronic, recurring pathology, exacerbating LUTS and incontinence.

Thus, a crucial first step in the assessment of an elderly patient with LUTS is the exclusion of UTI, but this is not straightforward [9,10]. There is widespread confusion over the significance of urinary infection in the elderly, particularly what is termed ‘asymptomatic bacteriuria’ [11,12]. Thus, there is a need to revaluate the diagnosis of urinary infection, particularly in the elderly.

The reference standard

Currently the diagnosis of UTI is subordinate to the culture of a clean-catch, midstream urine sample (MSU), where a threshold of $10^5$ colony forming units (cfu) ml$^{-1}$ of a single species of a known urinary pathogen defines the diagnostic threshold for infection. This criterion, described by Kass in 1957 rested on a study of 74 women with acute pyelonephritis, and 335 asymptomatic controls [13]. In a further analysis of the MSU culture, Kass used pregnant women with pyelonephritis to represent significant infection [14]. Despite the very limited data sets and the highly selected samples, the Kass criterion has become the gold standard for all clinical situations. The widespread adoption of this single, dichotomous threshold for diagnosing UTI, was criticized by some, but their warnings have gone unheeded [15,16]. In recent times there has been renewed criticism of the applicability of this threshold, and the Kass criterion is appearing...
increasingly untenable [8,17,18]. In different experiments on patients with acute dysuric cystitis [17] and in chronic lower urinary tract symptoms [8], it has been found that the adoption of the Kass threshold discards at least 50% of genuine UTIs.

**Diagnostic categorization**

The use of a single diagnostic threshold of \( \geq 10^5 \) cfu ml\(^{-1} \) of a single species of a known urinary pathogen implies a dichotomous choice, ‘UTI’ or ‘no UTI’ and the use of strict categorization. Medicine has always used categories to differentiate and understand ideas but the ability of categories to reflect reality is insufficiently questioned. Immanuel Kant warned that categories are inventions of the mind that should not be confused with reality [19]. More recently, Karl Popper encouraged us to abandon the absolutism of categorization in favor of fallibilism, which adopts the principle that we could be wrong about our understanding of the world. So instead of informing a patient that ‘you do not have an infection’ we use probabilistic descriptors that take account of our uncertainty [20].

**Using probability instead of certitude**

Human cognition is known to struggle mightily with probability (P) and that is why we have problems in interpreting the test results manifested by our patients. A way of clarifying the situation is to habitually consider P as conditioned. Always ask the question “The P of A; given what?” The given is extremely important. Another worthwhile habit is to be sceptical about any Gaussian distribution. The Gaussian distribution is a product of yes/no dichotomous classifications and thus a product of the human imagination. In truth, Gaussian distributions are rare in nature.

**Figure 2** is a classical Gaussian bell-shaped curve. Let us assume that it describes the distribution of values of a test variable from a sample of the population. The mean (3.5 units), median and mode coincide in the middle. We have demarcated the point that is one standard deviation (SD = 1) from the mean. This is a point estimate and we might conjecture that results greater than this value are abnormal. It is deceptively neat and simple. Gaussian extrapolations are often used to make predictions in medicine but the justification for doing so is extremely dubious [23,24].

**Figure 3** illustrates real life data and is a plot of the urinary \( \log_{10} \) (pyuria) (white blood cell [WBC] µl\(^{-1} \)) on the x-axis (abscissa) with, on the y-axis (ordinate), the proportion with that pyuria found to be MSU culture positive. Hence this graph plots the P of a positive MSU culture given the specific \( \log_{10} \) (pyuria) (WBC µl\(^{-1} \)). In the background, colored by a gray spectrum, we have fitted a continuous P density function reflecting the properties of the population. To repeat the point; the y-axis of this function provides the P of a positive MSU culture given the specific \( \log_{10} \) (pyuria)(WBC µl\(^{-1} \)) on the x-axis.

The current categorical cut-off for normal excludes patients with an 18% P of a positive culture. The dashed arrow demonstrates a 43.5% P of a positive MSU culture, given a pyuria count of \( 10^4 \) (log \( _{10} \) [pyuria] = 4).

The double headed arrows illustrate another important point; it is not right to use a point estimate \( \log_{10} \) (pyuria) = 4. We should be far less didactic with a 95% CI instead (\( \pm \)). The related estimate of the P of a positive culture has a wider 95% confidence interval because, being extrapolated from another measure, there is more doubt about its true value.

**Figure 4** shows that we should be sceptical about the results of MSU culture.

**Figure 1** illustrates the problem that we have created: A spectrum is drawn between the two extremes of no UTI and an infection of sufficient severity to require intensive care. A single diagnostic threshold places an arbitrary boundary on the continuum and declares all below as no UTI and all above as UTI.

The presumption of imposing a categorical dichotomy onto a continuum is doubtful. Nature is inimical to categories. Biological phenomena are dispersed across continua. Charles Darwin never tired of emphasizing the gradualism in nature dispersed across continua. Charles Darwin never tired of emphasizing the gradualism in nature. Nature dichotomy onto a continuum is doubtful. Nature

---

**Figure 1. Categorization imposed on the disease spectrum.**

- 'Normal'/abnormal' border
- You have no infection < \( \geq \) You have an infection
- Fully healthy
- Pyelonephritis and moribund
- Spectrum of disease

---
The problems affecting the diagnosis of urinary tract infection – PERSPECTIVE

by Figure 3. The figure shows that 26% of patients with a positive MSU culture have zero pyuria. There is a remarkable lack of concordance between MSU culture and pyuria.

The reason for showing these data is to illustrate the errors that can be introduced by attempting to tidy up the picture through the artifice of categorization. The data are not Gaussian and methods based on the bell-shaped curve are unjustified and our patients will not fit into such constructs. These figures really do refute the advice “You do not have an infection because your urine tests are negative”. We are only likely to make sense of these difficulties if we start to use correct, probabilistic reasoning and the best option for achieving this is the application of Bayes’ theorem.

### The importance of predictive values & Bayes’ theorem

Apart from the misrepresenting continua, describing a test outcome as normal or abnormal is compromised by another very significant interpretation error. For better understanding, we will illustrate this using some simple mathematics.

Table 1 illustrates the true situation as it applies to any test. There are always four different outcomes, not two: true positive (TP); true negative (TN); false positive (FP); and false negative (FN). We compound the difficulties by using these variables to calculate derived values that can be even more misleading.

The sensitivity (Sn) of a test is the proportion of diseased patients correctly identified by the test through positive (TP) outcome. The total number of diseased patients is the sum of TPs plus FNs. The equation for the Sn can be written as:

\[
Sn = \frac{TP}{TP + FN}
\]

The specificity (Sp) of a test is the proportion of disease-free individuals correctly identified by a negative test. By the same reasoning, the equation for the Sp can be written as:

\[
Sp = \frac{TN}{TN + FP}
\]

It is important to appreciate that Sn and Sp of a test are population-level properties that say very little about the individual patient sitting in front of you. The generalizations appropriate to a population are very different to the specific circumstances of the patient, who may well be an outlier [23].

For the sake of completeness, we mention two values that are derived from the Sn and Sp. The likelihood ratios are invoked by some but their purpose is not very clear. The equations for the likelihood ratios are given below; they are only applicable at the population-level.

Positive likelihood ratio = \[
\frac{Sn}{1 - Sp}
\]  (3)

Figure 2. A classical Gaussian distribution.

<table>
<thead>
<tr>
<th>Test result</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥4.5 units abnormal</td>
<td>25000</td>
</tr>
<tr>
<td>4.5 units abnormal</td>
<td>15000</td>
</tr>
<tr>
<td>3.5 units abnormal</td>
<td>10000</td>
</tr>
<tr>
<td>2.5 units abnormal</td>
<td>5000</td>
</tr>
<tr>
<td>Current normal cut-off</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 3. Frequency distribution of the proportion of patients with a culture positive midstream urine sample, given the log10 (pyuria) (WBC µl⁻¹) count on the x-axis. A probability density function is shown in the background. The continuous spectrum, typical of nature, is emphasized by the change in the gray shading.

UTI: Urinary tract infection; WBC: White blood cell.
The two variables that a clinician really needs for specific patients are the positive and negative predictive values (NPVs). Beware of how these are calculated; there is a significant trap.

The positive predictive value (PPV) is a measure of the P that the patient has the disease given a positive test. The NPV is the P that the patient does not have the disease given a negative test.

Now for the problem; the PPV and NPV can be calculated from the TPs and FPs and TNs and FNs as described below. If this course is taken then the results are only valid at population level and say little about the specific patient.

\[
PPV = \frac{TP}{TP + FP} \quad (5)
\]

\[
NPV = \frac{TN}{TN + FN} \quad (6)
\]

It is in the interests of the patient that predictive values be calculated using the equations below, which we shall simplify a little further on in the test.

\[
PPV = \frac{Sn \times Pr}{Sn \times Pr + (1 - Sp) \times (1 - Pr)} \quad (7)
\]

The Pr is the P that you have the disease given the characteristics of the population subgroup that the patients bests fits.

\[
NPV = \frac{Sp \times (1 - Pr)}{(1 - Sn) \times Pr + Sp \times (1 - Pr)} \quad (8)
\]

These two equations are examples of Bayes’ equation [25].

The symbols P(A|B) mean “The P of A given that B has occurred”. This equation can be rewritten to express the circumstances of a clinical test:

\[
P of disease given test result = P(D pos | Test pos) \quad (9)
\]

\[
P of test result given disease = P(Test pos | D pos) \quad (10)
\]

\[
P of disease = P(D pos) = Pr \quad (11)
\]

\[
P of test result given no disease = P(Test pos | D neg) \quad (12)
\]

\[
PPV = P(D pos | Test pos) = \frac{P(Test pos | D pos) \times P(D pos)}{P(Test pos | D pos) \times P(D pos) + P(Test neg | D neg)} \quad (13)
\]

In Bayes’ terminology the PPV is called the posterior, the Pr or P(D pos) is called the prior and the P of a true positive P(D pos | Test pos), the same as the Sn is called the likelihood function. The denominator is referred to as ‘the P of the data’, meaning the P of Test pos, which is the same as the sum of the P of a TP and the P of a FP.

This very important relationship can be simplified to the following:

\[
PPV \propto Sn \times Pr \quad (14)
\]

\[
NPV \propto Sp \times (1 - Pr) \quad (15)
\]

We must emphasize that the Pr here is not referenced to the population, but to a specific subgroup that contains the individual patient. This means that it is essential to characterize the patients carefully so that they can be placed, with as much precision as possible, into a subgroup of the population and the Pr of that subgroup should be used.

The following example should clarify this point: the Pr of disease in a healthy 23-year old man, with no symptoms or signs, and taking no medication, will be very different to that applicable to an 85-year old, diabetic, female patient with frequency, urgency and incontinence and a 15-year history of recurrent UTI. These contrasts will make a substantial difference to the PPV.

An important question must follow: how do you know the Pr as it applies to an individual patient? At the moment we do not know, but must use our clinical judgment to modify the probabilistic judgments that we make. That is
no different from obtaining a careful history and examination and using our knowledge and skill to achieve an estimation. This is infinitely preferable to the crudeness of doing a test and then applying one rule to all.

One of the most important properties of Bayes’ equation is that the posterior P, output from one calculation, can be used as the prior P input for a subsequent calculation. Thus, with each step of the clinical assessment, a sequence of Bayesian calculations can build an increasingly sophisticated probabilistic appreciation of the differential diagnosis. Modern microcomputer methods and the developments catalyzed by the computer games industry have made it feasible for the appropriate Bayesian methods to be deployed as a patient is clerked and examined. At this time, we are collaborating with mathematicians at University College London to construct such a system for the assessment and management of patients with lower urinary tract symptoms. In the future, we should prove very much more adept in the way that we interpret test data.

Inherent to this method is a strong criticism of routine batteries of tests, instituted without consideration of the specific circumstances, with blind acceptance of the results. These approaches are costly, misleading and harmful to the patients.

### The sensitivities of the tests

The Sn of a test has been placed in context. Used in isolation it speaks for the population; for the individual patient the Sn should only be used as a component of a Bayesian analysis. Applied in this manner, the Sn of a test used to diagnose and the Sp of one used to exclude, need to be of high order. Regrettably, when it comes to urinalysis, the news on this front is not good.

### Urine microscopy & the quantification of pyuria

An alternative method of diagnosing urine infection is the identification of significant pyuria. The determination of the excretion rate of white blood cells in urine was first described by Hottinger [29]. However, the technique used is usually attributed to Addis and referred to as the Addis count [30].

Mond examined unspun fresh urine and counted the white cells using a hemocytometer, finding that ≥10 WBC µl⁻¹ were noted in all patients with symptomatic acute cystitis and bacteriuria (10⁵ cfu ml⁻¹) [31]. In 1968, Gadeholt compared the examination of spun and unspun specimens and showed that the calculated excretion rates on uncentrifuged specimens showed higher values [32]. This confirmed earlier suspicions that centrifugation caused loss of cells.

Gadeholt showed that all patients with ≥10 WBC µl⁻¹ of unspun urine excreted >400,000 WBC h⁻¹. This was confirmed by Mabeck in 1969 [33] and by Baerheim et al. [34]. Thus the finding of ≥10 WBC µl⁻¹ of unstained, unspun urine examined on a hemocytometer,
became established as the most effective method for diagnosing urinary infection [34]. Latham and Stamm reported that the detection of pyuria ($\geq 10^2$ WBC $\mu l^{-1}$) was the most accurate and efficient method of identifying UTIs in ambulatory women with symptoms of acute cystitis [18].

**Midstream urine culture**

In 1982 Stamm published work conducted on acutely dysuric women [17]. He discovered that in the presence of the classical symptoms of frequency/dysuria, a culture result of $10^2$ cfu ml$^{-1}$ was a more appropriate threshold for diagnosis.

Standard laboratory methods for analysing MSUs are set at a threshold of $10^5$ cfu ml$^{-1}$ [17]. Thus, Hooton and Stamm have shown that laboratory MSU cultures miss over 50% of genuine infections in women with frequency/dysuria [35].

**Contemporary data from this centre in nonacute lower urinary tract symptoms**

We tested an enhanced urine culture of catheter urine (CSU) samples by inoculating nonselective agar immediately and culturing anaerobically with a threshold of $10^2$ cfu/ml. 297 women with lower urinary tract symptoms provided specimens. A total of 93 (31%) showed pyuria with only 29 (31%) of these pyuric urines being positive by routine culture methods but 40 (43%) positive on enhanced culture [36].

Of the 204 without pyuria, 20 (9.8%) grew bacteria on enhanced culture and only 9 (4.4%) did so on routine culture. Thus enhanced culture, with an appropriate culture threshold, identified a larger number of patients with LUTS as having a UTI [36].

This experiment was followed by a prospective, blinded, observational cohort study of outpatients presenting with painless LUTS. Urine samples were analyzed as follows:

- A total of 508 MSUs were used to compare leukocyte esterase, nitrite dipstick and urine microscopy with cultures seeking $10^3$ cfu ml$^{-1}$;
- A total of 470 CSUs were similarly used to compare the same surrogates with $10^2$ cfu ml$^{-1}$ and with the enhanced culture method, seeking $10^2$ cfu ml$^{-1}$;
- Comparison of leukocyte esterase against microscopic pyuria in 508 MSUs and 470 CSUs. A total of 42 normal volunteers provided MSU specimens for comparison [8].

For an MSU culture at $10^5$ cfu ml$^{-1}$, the leukocyte esterase dipstick was 56% sensitive, nitrite 10% sensitive and microscopic pyuria 56% sensitive. Specificities were 66%, 99% and 72% respectively.

For CSU culture at $10^2$ cfu ml$^{-1}$, the leukocytes esterase dipstick was 59% sensitive, nitrite 20% sensitive, and microscopic pyuria 66% sensitive. Specificities were 84, 97 and 73%, respectively.

For the dipstick detection of significant microscopic pyuria ($\geq 10$ WBC $\mu l^{-1}$), the leukocyte esterase dipstick demonstrated a Sn of 38% and Sp of 84%.

The CSU culture at $10^2$ cfu ml$^{-1}$ was positive in 29% compared with 15% at $10^5$ cfu ml$^{-1}$.

These findings are illustrated in Figure 5, which depicts the relative proportions of patients with an infection, diagnosed on CSU at $10^2$ cfu ml$^{-1}$, that would be identified by each test.

Thus our group, Stamm [17], Deville et al. [6], and Hurlbut and Littenberg [7] have found that the three key surrogate markers of urinary infection and the routine urine culture method, were significantly deficient [8].

These tests are commonly positive with more severe infection and such concurrences are memorable. Thus, recall bias may encourage confirmation bias, which provokes a mistaken perception of test efficacy. How can a test be inaccurate when it is positive in such obviously infected patients?

**Figure 5. What the tests really detect.**

The largest circle represents the patients who would be diagnosed with a urinary tract infection on culture of a CSU isolating $10^2$ CFU ml$^{-1}$ of a single species of a known urinary pathogen. The smaller circles represent the subset of this larger group that would be identified by the test that is named in the relevant subcircle.

CFU: Colony forming units; CSU: Catheter urine sample; MSU: Midstream urine sample; WBC: White blood cell.
Diagnosis by symptoms

Accepting the shortcomings of the tests, research is focused on finding alternatives. Progress is being made and the future does seem hopeful [37–40]. On this occasion, novel diagnostic markers must be validated properly across the whole disease spectrum, which takes time. Inevitably, we are asked what to do in the meantime? In answer to this question, we describe our own approach below, although currently subject to critical evaluation.

The most sensitive symptoms of urinary infection in patients without acute disease are not pain but hesitancy, reduced stream, intermittency and terminal dribble, and this applies to both genders [1].

Our data imply that a uriniferous odor is almost invariably associated with evidence of UTI. Suprapubic tenderness and loin tenderness do appear to be important markers of disease activity in patients with chronic low grade infections.

Systemic markers of infection, such as C-reactive protein and erythrocyte sedimentation rate, seem unresponsive to many UTIs.

If a patient has symptoms, then the P that they have infection is high, regardless of routine urinalysis [1,41,42].

We treat patients on their symptoms. If the urine tests prove positive, the P of urine infection is very high. We are forced to judge the efficacy of treatment on the patients’ response description. We have reported success with these methods but we do use protracted courses of antibiotics, which is controversial [43]. We have initiated randomized control study work.

Future perspective

The science that we do and which has contributed to this review was described by Karl Popper as “Scientific evolutionary epistemology” [20,44]. It is also recognized as an example of ‘Cultural evolution’ [45]. Nassim Taleb describes the discipline of ‘Sceptical empiricism’, which is a very important component of the method [23,24]. We believe that these methods will become increasingly evident in medical science and Bayesian methods will start to feature in clinical practice.

Executive summary

Reference standard

- The current gold standard reference used to diagnose urinary infection in acute and chronic disease appears to contain very significant errors.
- The surrogate indicators of urinary tract infection (UTI), dipsticks and urine microscopy, have been calibrated to this dubious standard set by Kass.

Diagnostic categorization

- The use of dichotomized categorization ‘UTI’ versus ‘No UTI’ when interpreting results does not comply with the underlying biology.

Using probability instead of certitude

- The judgement of the presence or otherwise of UTI should really be expressed as a probability lying somewhere between the lower and upper margins of an interval.
- It is important to be conscious of the biological continua that underlie all diseases.
- Nature does not conform to the human expedient of categorization.

The importance of predictive values & Bayes’ theorem

- A test result should be interpreted by giving due consideration to the specific circumstances that affect the individual patient.

The sensitivities of the test

- The widespread supposition that, given a negative test, a patient cannot have a urine infection is not correct.

Urinary tract infection diagnosis by urinary tests

- Urinalysis tests results are conditioned on patients in hospital with pyelonephritis; these methods are not valid for acute cystitis nor chronic lower urinary tract symptoms.
- Rigorous studies have shown that urine dipstick analysis and midstream urine sample culture are very insensitive and provide flawed data.

Urinary tract infection diagnosis by symptoms

- The symptoms that a patient describes are very good indicators of disease.
- Hesitancy, reduced stream, intermittency and terminal dribbling are the most sensitive markers of urinary tract infection in patients with nonacute lower urinary tract symptoms.
- New diagnostic methods are in development but for the time being the arbiter is the bedside clinical assessment with palpation for suprapubic and loin tenderness.

Treatment decisions

- Treatment decisions have to be based on symptoms, signs and response; there are no tests capable of providing this function.
Financial & competing interests disclosure
The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties. No writing assistance was utilized in the production of this manuscript.


