

# Debate & Analysis

## The leaf plot:

a novel way of presenting the value of tests

### INTRODUCTION

The power and accuracy of clinical tests is usually reported either in terms of their sensitivity and specificity, their predictive values, or their likelihood ratios, but these concepts can be difficult for many GPs to apply to real-life clinical situations.<sup>1</sup>

### Sensitivity and specificity

These are independent of the prevalence of the condition (or its equivalent in an individual patient, your estimate of their pre-test probability of having the condition), and so cannot answer the clinician's question of 'How much does a positive or a negative result for this test or sign influence the probability of my provisional diagnosis?' Correctly interpreting these values is difficult, and requires us to grasp non-intuitive concepts with 'both sides of our brains'.<sup>2</sup>

### Positive (PPV) and negative (NPV) predictive values

These seem to make more sense, but are misleading because they can only be applied to populations with the same prevalence of the condition as was present in the study that generated them. For example, studies in special educational facilities show that finding a child with a single-palmar-crease gives a PPV of them having Down's syndrome of about 75%, but if you notice this pattern in the setting of a normal infant having a 6-week check it then would only have a PPV of about 10%.

### Positive and negative likelihood ratios

These seem more helpful because they determine how a test result will alter the pre-test odds, but it is not straightforward to quantify their impact for an individual patient. The clinician has to estimate that person's pre-test odds of having the diagnosis (= probability/1 - probability), and then multiply that by the appropriate likelihood ratio to find their new odds.

### AN ALTERNATIVE IS NEEDED

Because these methods are difficult to apply accurately in real practice, they may cause doctors to make vast errors when estimating the significance of screening results.<sup>3</sup> Very few GPs use them in any formal way, instead relying on other techniques such as their previous experience of that test.<sup>4</sup> Here we introduce the 'leaf plot' — a novel, visual way to estimate the impact that a positive or

negative test result will have on your patient's chance of having the diagnosis you suspect. We have designed it to avoid the pitfalls of previous methods of evaluating tests, and hope it will help clinicians interpret the value of tests more accurately in real-life practice.

### THE LEAF PLOT

The leaf plot gives you a visually intuitive and accurate estimate of the impact that a positive or a negative test or clinical finding will have on your patient's chances of having a diagnosis. It can be easily generated by entering the sensitivity and specificity of a test into an Excel document, and this is freely available on the charity website [childhealthafrica.org/downloads](http://childhealthafrica.org/downloads).

### How to use the leaf plot

The starting probability of a diagnosis is shown diagonally along the leaf 'vein' from nil at the bottom left, to complete certainty at the top right (Figure 1). This is your best guess of approximately how likely it is that your patient has that condition; the precise position is not critical.

Once you have decided where your patient's pre-test probability sits on the leaf's vein, then the impact that a positive test result will have on that probability is shown by the height of the vertical jump up to the red line directly above. Similarly, the impact of a negative test is shown by how far the probability drops down as it falls to the blue line. It follows that if the pink and blue areas are close to the central vein like a willow leaf, the test is weak and will make little difference to your decision making, whereas a test that produces a broad-leaved plot that reaches towards the corners of the graph will be much more useful.

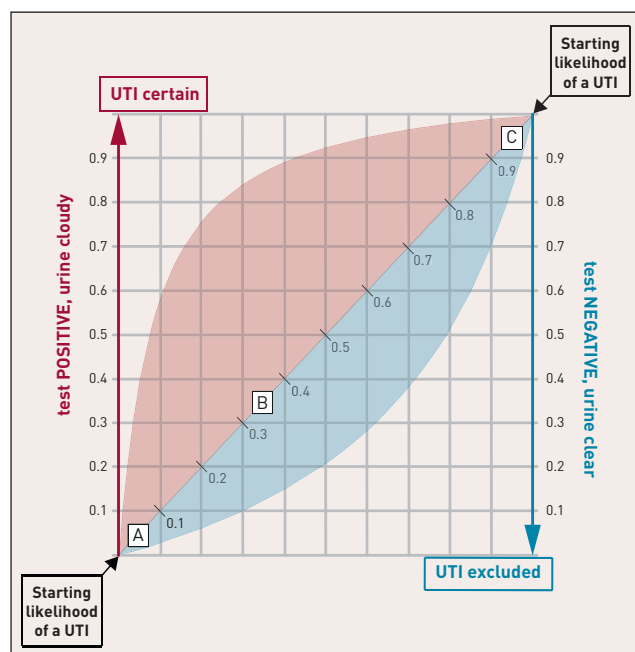
### A worked example

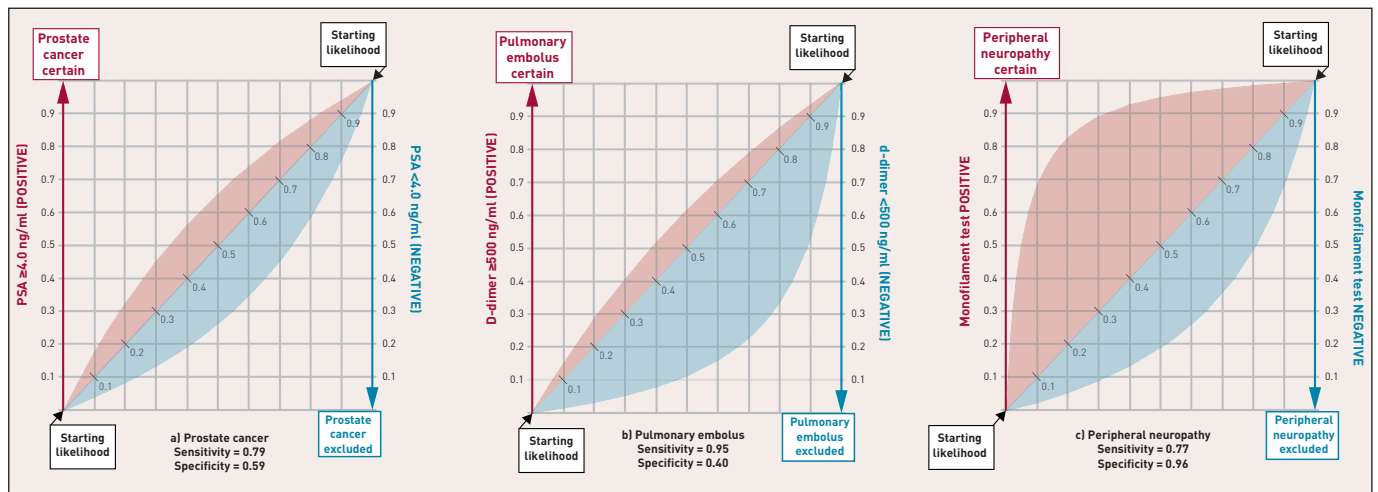
Here we will see how

to find out whether it is useful to check if a child's urine looks cloudy when you are considering the diagnosis of them having a urinary tract infection (UTI). If three-quarters of children with a UTI have cloudy urine (sensitivity 0.75), and 94% of healthy children pass clear samples (specificity 0.94), the test would generate the leaf plot shown in Figure 1. It is immediately obvious that the red area is bigger than the blue, indicating that a cloudy urine (positive test) has greater power to rule in UTIs than a clear one (negative test) does to rule them out. Now let us consider three clinical scenarios that might present in primary care, corresponding to points A, B, and C on the leaf vein.

Point A would be what would happen if you decided to screen a healthy child for a UTI by checking if they had cloudy urine. The chances of a UTI in children with crystal-clear urine would fall from an already very low level to even closer to zero, and the chances of a child with cloudy urine having a UTI would still be less than evens. Not a useful screening test.

Figure 1. Leaf plot to see how useful it would be to check for urine cloudiness to help diagnose a child's urine infection (UTI), assuming that the test has a sensitivity of 0.75 and specificity of 0.94. The initial estimated probability of the diagnosis of UTI can be anywhere on the diagonal black line from 0 at the bottom left to 1 at the top right, depending on the clinical details. The impact of a positive test (a cloudy urine) is shown by the red line and shaded area, and is easy to read from the left-hand axis, and the impact of a negative test (a clear urine) is shown in blue on the right. Points A, B, and C are used to illustrate three clinical examples given in the text.





**Figure 2.** Leaf plots for three commonly used screening tests. The prostate-specific antigen test for prostate cancer<sup>5</sup> produces a 'willow leaf' appearance showing only a limited ability to confirm or exclude prostate cancer (a), the d-dimer assay for pulmonary embolus<sup>6</sup> has a broad blue side which shows that a negative test helps to rule out a pulmonary embolus, but that a positive test makes little impact on making the diagnosis (b), and the monofilament tests in diabetes<sup>7</sup> has a very broad pink side that indicates a positive result makes a peripheral neuropathy very likely, but a negative test has less power to rule one out (c).

Point B could represent the starting probability of a UTI for an otherwise well 6-year-old female presenting to her GP with slight stinging on micturition, after passing a concentrated urine on a hot day. You might have a moderate (say, one-in-three) concern that she could have a first UTI. Here, a clear urine would reduce her probability of having a UTI to about one in eight, enabling you to watch and wait, whereas a cloudy sample would increase her probability of having a UTI to over 85%, which might prompt you to culture a mid-stream urine. Possibly a useful test in these circumstances.

Point C might be a 2-year-old female who you know has bilateral renal scarring caused by recurrent febrile UTIs and vesicoureteric reflux, and who has become febrile again and started vomiting. Here, because her starting probability of having another UTI is high (say, about 95%), finding a clear urine would still leave her with about an 85% chance of having an infection, and a cloudy test would merely increase her probability from 95% to near certainty. Neither of these mild alterations to an already high risk would alter your decision to immediately culture a urine sample and commence antibiotic treatment while awaiting microbiological confirmation. It would therefore be a waste of time for you to look at the urine clarity in this setting. Although assessing the turbidity of urine is a trivial task, other tests might be time consuming, cause delay, and be costly.

### Other examples

Other leaf plots of commonly used screening tests are shown in Figure 2. The prostate-specific antigen test for prostate cancer<sup>5</sup>

only provides weak additional diagnostic help, as shown by its 'willow' leaf shape. The d-dimer test for pulmonary embolus<sup>6</sup> is often misused.<sup>7</sup> With a narrow pink side and a broad blue side, the leaf plot makes it clear that a positive result is not useful for making the diagnosis and that the main use of the test is excluding a pulmonary embolus in patients with low baseline risk. The leaf plot of the 10 g Semmes-Weinstein monofilament examination has a broad pink side, which means a positive test in a diabetic patient makes that a peripheral neuropathy much more likely,<sup>8</sup> but a negative test does not strongly rule a neuropathy out.

### IMPLICATIONS FOR RESEARCH AND PRACTICE

It is to be hoped that future research which evaluates the value and impact of signs and tests will not only publish sensitivity and specificity data, but also produce leaf plots to provide an easy-to-understand graphic aid.

#### Malcolm G Coulthard,

Honorary Consultant Paediatric Nephrologist, Department of Paediatric Nephrology, Great North Children's Hospital, Newcastle upon Tyne.

#### Tom Coulthard,

GP Lecturer, School of Medical Education, Newcastle University, Newcastle upon Tyne.

#### Provenance

Freely submitted; externally peer reviewed.

#### Competing interests

The authors have declared no competing interests.

The authors would like to add that donations to Child Health Africa can be made on the charity website: [www.childhealthafrica.org](http://www.childhealthafrica.org).

### ADDRESS FOR CORRESPONDENCE

#### Malcolm G Coulthard

Department of Paediatric Nephrology, Great North Children's Hospital, Queen Victoria Road, Newcastle upon Tyne NE1 4LP, UK.

Email: [jeanandmalc@gmail.com](mailto:jeanandmalc@gmail.com)

DOI: <https://doi.org/10.3399/bjgp19X702113>

### REFERENCES

- Steurer J, Fischer JE, Bachmann LM, *et al*. Communicating accuracy of tests to general practitioners: a controlled study. *BMJ* 2002; **324(7341)**: 824–826.
- Loong TW. Understanding sensitivity and specificity with the right side of the brain. *BMJ* 2003; **327(7417)**: 716–719.
- Casscells W, Schoenberger A, Graboy TB. Interpretation by physicians of clinical laboratory results. *N Engl J Med* 1978; **299(18)**: 999–1001.
- Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med* 1998; **104(4)**: 374–380.
- Catalona WJ, Smith DS, Rattliff TL, *et al*. Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. *N Engl J Med* 1991; **324(17)**: 1156–1161.
- Perkins BA, Olaleye D, Zinman B, Bril V. Simple screening tests for peripheral neuropathy in the diabetes clinic. *Diabetes Care* 2001; **24(2)**: 250–256.
- Le Gal G, Righini M, Wells PS. D-dimer for pulmonary embolism. *JAMA* 2015; **313(16)**: 1668–1669.
- Smith C, Mensah A, Mal S, Worster A. Is pretest probability assessment on emergency department patients with suspected venous thromboembolism documented before Simplied D-dimer testing? *CJEM* 2008; **10(6)**: 519–523.