Quantifying how tests reduce diagnostic uncertainty

Malcolm G Coulthard

Arch Dis Child 2007;**92**:404–408. doi: 10.1136/adc.2006.111633

Correspondence to:
Dr Malcolm G Coulthard,
Royal Victoria Infirmary,
Newcastle NE1 4LP, UK;
malcolm.coulthard@nuth.
nhs.uk

Accepted 30 November 2006
Published Online First
11 December 2006

Background: Diagnostic tests are commonly evaluated from sensitivity and specificity, which are robust and independent of prevalence, but clinically not intuitive. Many clinicians prefer to use positive and negative predictive values (PPV and NPV), but they are frequently applied as if they are independent of prevalence, whereas this may make an important difference.

Methods and results: We present graphs that allow easy reference to appropriate values and demonstrate that PPV and NPV are not independent of prevalence. PPV and NPV figures reflect the prior probability of the case having a positive diagnosis, estimated clinically from the history, examination and other results, as well as the impact of the test result. To avoid the common error of allowing for prior probability twice, and to interpret the impact of the test result alone, we present graphs of the proportionate reduction in uncertainty score (PRU), calculated from sensitivity, specificity and prevalence. These plots show the extent to which either a positive or negative test result affects the remaining degree of uncertainty about a diagnosis in either direction, according to likely clinical prevalence.

Conclusions: PRU plots demonstrate the discriminatory value of tests more clearly than sensitivity and specificity from which they are derived, and should be published alongside them.

Clinicians usually judge the likelihood of a patient having a particular diagnosis by considering information from several sources and integrating these separate assessments. Much of this evaluation is reached apparently unconsciously and instantaneously from a knowledge of populations and diseases, such as the immediately obvious widely different diagnostic implications of jaundice when it occurs in a well 2 day old baby girl compared to an elderly alcoholic man. However, the diagnostic implications of laboratory and imaging test results are generally considered more consciously and formally based on their published performance data.

The power of diagnostic tests is usually reported either as sensitivity and specificity, or their derived likelihood or diagnostic odds ratios, or as their positive and negative predictive values (PPV and NPV). These allow clinicians to evaluate how well a test can distinguish between patients from different diagnostic groups. However, there are problems with all these ways of summarising data. PPV and NPV best address the questions that clinicians intuitively want to ask but are of limited usefulness because they can only be applied to groups of patients with the same disease prevalence (or individuals with a similar disease probability) as the population from which the data were derived.

I therefore propose a method of plotting the diagnostic implications of positive or negative test results which may make their interpretation more intuitive and convenient to clinicians, allowing them to appreciate their implications easily and quickly from a graph, for any level of estimated disease probability.

SENSITIVITY AND SPECIFICITY

The sensitivity is the proportion of true cases that register a positive test result, and the specificity is the proportion of unaffected individuals found to be test-negative. Thus, each term only applies either to the affected group or to the healthy group, and does not take patients in the other group into account. They are therefore robust parameters, and independent of the prevalence of true cases in the populations in which they are evaluated. Using table 1, sensitivity is calculated as TP/(TP+FN) and specificity as TN/(FP+TN), and may be expressed as a proportion or percentage. However, sensitivity and specificity do not necessarily provide a convenient or intuitive answer for the clinician's questions.

LIKELIHOOD RATIOS AND DIAGNOSTIC ODDS RATIOS

Other measures can be derived by combining both the sensitivity and specificity values, including the likelihood ratios (LRs), which are determined according to whether the test result is positive ($LR+ = \text{sensitivity}/(1 - \text{specificity})$), or negative ($LR- = (1 - \text{specificity})/\text{sensitivity}$). These values can be further combined to produce the diagnostic odds ratio, calculated as $LR+/LR-$. However, clinicians may be more familiar with the concept of probability (degrees of diagnostic certainty and uncertainty) than with assessing odds ratios, and may therefore find predictive values more intuitive.

PPV AND NPV

Predictive values address the chances of a particular diagnosis from the perspective of a known test result. The PPV is the proportion of genuinely positive cases among all the patients with positive test results, which includes the true and false positives. It is calculated (table 1) as $PPV = TP/(TP+FP)$. Similarly, the NPV estimates the number of truly negative cases among the total negative results, and is calculated as $NPV = TN/(FN+TN)$. These values also may be expressed as proportions or percentages.

A major limitation is that predictive values can only be applied to populations where the condition has a similar prevalence to the population tested, or to individuals with a

Table 1 Categories of true diagnoses, and test results

		True diagnostic category		Totals
		+	-	
Test result	+	TP	FP	TP+FP
	-	FN	TN	FN+TN
Totals		TP+FN	FP+TN	All cases

FN, false negatives; FP, false positives; TN, true negatives; TP, true positives.

Table 2 Expected frequencies for positive nitrite tests in a series of infants admitted to hospital with unexplained fever, malaise and vomiting, of whom half were found to have a UTI, assuming a sensitivity of 0.52 and specificity of 0.99

		UTI		Totals
		+	-	
Positive nitrite test stick	+	52	1	53
	-	48	99	147
Totals		100	100	200

UTI, urinary tract infection.

similar risk of a positive result. However, values are frequently published, quoted, or referred to in debate without reference to this, as if they are intrinsic to the test itself. This suggests they may not always be applied appropriately in clinical practice, even though population differences can have a huge impact on their interpretation. For example, when stick-testing urine for the presence of UTI in children, consider the PPV and NPV values generated by the observation that the nitrite test is positive in just over half of infected urines, but only in around 1% of sterile samples.^{1,2} The sensitivity and specificity of nitrite sticks as a diagnostic test for UTI are therefore 0.52 and 0.99, respectively. Tables 2 and 3 show the expected frequencies of positive nitrite tests and UTIs in two very different clinical settings, from which very different predictive values can be calculated.

Table 2 shows the experience in an acute paediatric ward where half of the infants admitted with unexplained swinging fever, malaise and vomiting are found to have a UTI; the PPV = 52/53, or 0.98, and the NPV = 99/147, or 0.67. Since a PPV of 98% is relatively high, it might suggest that nitrite sticks could provide a useful screening test for childhood UTI, with only a 2% chance of falsely identifying an unaffected child. Indeed, nitrite sticks are commonly used for this in general practice, with test-negative urines being discarded.³

However, table 3 shows that when nitrite sticks were used to screen children in a general paediatric outpatient clinic, where the prevalence of UTI was just two cases in 400, the PPV fell to just 20%. This means that a well child found on screening to have a positive nitrite test is about four times more likely not to have a UTI than to have one. By contrast, the NPV value in this population is dramatically higher than among the acutely ill infants, at 0.997. This suggests that a clinically well child whose urine was nitrite-negative would have less than one chance in

Table 3 Expected frequencies for positive nitrite tests in a series of children whose urines were screened routinely in a general paediatric outpatient clinic, assuming a sensitivity of 0.52 and specificity of 0.99

		UTI		Totals
		+	-	
Positive nitrite test stick	+	1	4	5
	-	1	394	395
Totals		2	398	400

UTI, urinary tract infection.

300 of having a UTI, whereas an acutely ill infant with a nitrite-negative urine would still have virtually a 19% chance of having infected urine.

ESTIMATING LIKELY LEVELS OF PREVALENCE, OR PRIOR PROBABILITIES

In order to interpret PPV and NPV figures appropriately, therefore, it is essential for clinicians to estimate the likely prevalence of the condition in their particular population, or for the individual case they are dealing with. When tests are being used widely for screening, this estimate is likely to be close to the overall population prevalence but may be much higher in specific settings such as specialist clinics. For individual patients, estimates may be very different, according to particular factors in the clinical history and other test results. This clinical assessment of the pre-test risk is essentially the same concept as defining the prior probability in Bayesian statistics. Although intrinsically imprecise, even if clinical estimates of probability merely provide a "ball-park figure", they will still allow more sensible interpretation of test results. Using the example of screening for childhood UTI with nitrite sticks, the prevalence among well nursery school children is likely to be under 0.01 but several-fold higher among children who have had previous UTIs, and might be as high as 0.95 for a child with known reflux nephropathy who suddenly develops a fever and rigors, offensive smelling urine, frequency and stranguria.

CALCULATING PPV AND NPV AT DIFFERENT LEVELS OF PREVALENCE

Given the sensitivity and specificity, it is simple to calculate PPV and NPV values for any range of likely population prevalences, as was done for two particular rates of UTI in tables 2 and 3.

Figure 1 Plots of (A) the positive and negative predictive values (PPV and NPV), (B) the value-added PPV and NPV, and (C) the proportionate reduction in uncertainty scores (PRU) for diagnosing a urinary tract infection (UTI) in a child as a result of carrying out a urinary nitrite stick test, according to the estimated likely prevalence of UTI given that individual's clinical circumstances.

Table 4 The impact of positive and negative urinary nitrite stick test results (assuming a sensitivity of 0.52 and specificity of 0.99) on the PPV and NPV of a child having a UTI, according to the prior probability (likely clinical prevalence) of them having a positive diagnosis

Likely clinical prevalence (%)	0.1	0.5	1	5	10	25	50	75	90	95
Positive test – “standard” PPV	0.050	0.207	0.344	0.732	0.853	0.946	0.981	0.994	0.998	0.999
Negative test – “standard” NPV	1.000	0.998	0.995	0.975	0.949	0.861	0.674	0.407	0.186	0.098

NPV, negative predictive value; PPV, positive predictive value; UTI, urinary tract infection.

For a prevalence of P (between 0 and 1), and with a total number of patients of N , the proportions of results in table 1 will be:

$$TP = N \times P \times \text{sensitivity}$$

$$FN = N \times P \times (1 - \text{sensitivity})$$

$$TN = N \times (1 - P) \times \text{specificity}$$

$$FP = N \times (1 - P) \times (1 - \text{specificity})$$

These are shown as “standard” PPV and NPV figures in table 4, and are also plotted in fig 1A for the whole range of clinically likely prevalences to make the usefulness of the tests immediately obvious at different incidences of the condition. Presenting these data in graphic form is likely to reinforce for clinicians the crucial importance of interpreting predictive values in the light of at least approximate estimates of the starting probability or risk of their particular cases being positive.

ONLY COUNTING THE VALUE ADDED FROM THE TEST RESULT

Standard PPV and NPV values can be misleading because they are constructed from two factors: the prevalence of the condition in the population they were generated in and the impact of the test result itself. If this is not appreciated, it can lead to the contribution from the prevalence being inadvertently counted twice. For example, the NPV of a negative nitrite stick result from an apparently healthy child in a general paediatric outpatient clinic means that the probability the child does not have a UTI is impressively high at 99.8%. However, that is mostly because the starting probability against that child having a UTI was 99.5% (in table 3, 398 of 400 children did not have one). Only 0.3% of extra certainty is contributed by the negative urine test. An unwary general paediatrician reviewing a well child would be intrinsically aware that the probability against them having a UTI would start off very small, say, at least 99% or more against. Knowing that a negative nitrite test gave an NPV for a UTI of 99.8% might lead them to assume that the true probability for the child not having a UTI could be compounded from both the “at least 99%” starting position and the 99.8% NPV odds, giving a final risk of 50 000 to 1 against, or virtually nil.

To overcome this risk of double-counting, and make it clearer that the effect of a test result should be compounded with the prior odds for that diagnosis, it is possible to calculate the value-added predictive value figures. The value-added PPV is computed as $(PPV - \text{expected prevalence})$, and the value-added NPV as $(NPV - (1 - \text{expected prevalence}))$. Table 5 shows these

figures for predicting a UTI from nitrite stick results across a range of prevalences, and fig 1B demonstrates this graphically.

PLOTTING THE PROPORTIONATE REDUCTION IN UNCERTAINTY (PRU)

Although “value-added predictive values”, adjusted for the appropriate estimated prevalences, give a much clearer picture of the diagnostic implications of test results than sensitivity and specificity, likelihood ratios, or standard PPV and NPV values, they may still be misleading. This is because they show the absolute improvement in diagnostic probability that the test result makes rather than the fraction or proportion of the outstanding uncertainty that it removes. It is probably more useful and intuitive to think of the impact of test results in a relative way. A small absolute improvement in diagnostic certainty may make little difference in the face of great uncertainty, whereas the same small absolute clarification could make a dramatic difference to the decision making process if the diagnosis was already thought to be very likely.

For example (table 5 and fig 1B), in a group of children in whom at most 0.5% would be thought likely to have a UTI, a positive nitrite test adds 20.2% to the chances of them being affected, but the final probability only rises to 20.7%, still leaving them with an approximate 80% likelihood of not having a UTI. By contrast, where strong clinical suspicion gives an estimated prior risk of 95%, the added value of just 4.9% from finding a positive nitrite test reduces the uncertainty 50-fold from 5% to just 0.1%, or very close to certainty.

I have therefore calculated PRU scores to express the proportion by which a positive or negative test result reduces the outstanding diagnostic uncertainty, rather than its absolute reduction. The scores for nitrite sticks in diagnosing childhood UTI can be seen in table 6, and are plotted against the clinical estimate of probability in fig 1C. For an estimated prevalence of P , the PRU for a positive test result is calculated as $PRU = (PPV - P)/(1 - P)$, and for a negative test result the PRU is calculated as $PRU = (NPV - (1 - P))/P$.

It is immediately clear from the graph that when the prevalence of UTI is likely to be relatively low, finding a positive nitrite test should have little influence on diagnostic decision making, but when the clinical situation suggests that a UTI is fairly likely, a positive test result has a greater diagnostic impact. At about a 1% probability of a UTI it reduces the diagnostic uncertainty by around 40%, but hugely shortens the odds when a UTI is clinically likely. The impact of a negative nitrite test can also be appreciated from the plot. In a group of children whose

Table 5 The impact of positive and negative urinary nitrite stick test results (assuming a sensitivity of 0.52 and specificity of 0.99) on the value-added PPV and value-added NPV figures of a child having a UTI, according to the prior probability (likely clinical prevalence) of them having a positive diagnosis

Likely clinical prevalence (%)	0.1	0.5	1	5	10	25	50	75	90	95
Positive test – value-added PPV	0.049	0.202	0.334	0.682	0.753	0.700	0.481	0.244	0.098	0.049
Negative test – value-added NPV	0.001	0.003	0.005	0.025	0.049	0.111	0.174	0.157	0.086	0.048

NPV, negative predictive value; PPV, positive predictive value; UTI, urinary tract infection.

Table 6 The impact of positive and negative urinary nitrite stick test results (assuming a sensitivity of 0.52 and specificity of 0.99) on the PRU figures of a child having a UTI, according to the prior probability (likely clinical prevalence) of them having a positive diagnosis

Likely clinical prevalence (%)	0.1	0.5	1	5	10	25	50	75	90	95
Positive test – PRU	0.049	0.203	0.338	0.718	0.836	0.927	0.962	0.975	0.979	0.980
Negative test – PRU	0.515	0.514	0.513	0.502	0.489	0.444	0.347	0.210	0.096	0.050

PRU, proportionate reduction in uncertainty; UTI, urinary tract infection.

probability of having a UTI is around 10%, a negative nitrite test reduces the likelihood from its already low level by about half, but when the prevalence is likely to be high the discriminatory value falls off dramatically. These patterns of varying usefulness of tests with likely prevalence are immediately visually apparent from the plot but difficult to appreciate just from knowing the sensitivity and specificity data from which they are derived.

EXAMPLES OF PRU PLOTS

Figure 2 shows four PRU plots for tests with combinations of high (0.99) and low (0.70) sensitivities and specificities. The impacts of positive (black lines) and negative (grey lines) test results can be easily appreciated across the whole range of pre-test diagnostic probabilities, or estimated prevalences of a positive diagnosis. Clearly, when tests are both highly sensitive and highly specific (test A), either a positive or a negative result will greatly reduce the remaining diagnostic uncertainty at all but the most extreme values of estimated prior probability.

Conversely, when the sensitivity and specificity are both relatively low (test C), no test result can make a powerful contribution to decision making, but it is clear that a positive test is most valuable when the pre-test probability is high, and a negative test is most useful in screening situations.

It is well known that negative results from highly sensitive tests (like A and B) are especially powerful at correctly ruling out diagnoses in unaffected patients, but that positive results from such highly sensitive tests will only rule in affected or true-positive cases strongly if they are also highly specific. This is clearly expressed by the “SPIN and SNOUT” mnemonic: highly SPecific tests are needed to rule diagnoses IN, and highly SeNsitive tests are needed to rule them OUT.⁴ These plots also make this concept clear. Similarly, while positive results of highly specific tests like A and D can powerfully rule in affected cases, negative results from these tests can only powerfully rule them out if they are also highly sensitive. The interactions between sensitivity and specificity are also obvious in these plots.

Figure 2 Plots of the proportionate reduction of uncertainty (PRU) according to finding a positive test (black lines) or a negative test (grey lines) result, for tests with various combinations of sensitivity and specificity.

What is already known on this topic

- Sensitivity and specificity are of limited clinical usefulness alone, and are commonly converted to positive and negative predictive values.
- Although predictive values also depend upon prevalence, they are commonly used as if there is a single value for a particular test and this frequently leads to serious misinterpretation.

What this study adds

- The proportional reduction in uncertainty (PRU) is also calculated from prevalence, sensitivity and specificity, but describes the extent that a test result reduces the outstanding degree of clinical uncertainty.
- The PRU is best presented graphically, so the impact on diagnostic uncertainty can be appreciated for any likely clinical prevalence.
- PRU plots should be presented alongside all sensitivity and specificity data.

In situations where the clinically likely range of prevalences includes very low figures, the graph of PRU scores may be seen more clearly if the prior prevalence is plotted on a logarithmic scale.

CONCLUSIONS

PRU plots provide a picture of the additional diagnostic value provided by clinical signs or tests which can be immediately appreciated far more intuitively than the sensitivity and specificity data from which they are derived. They make it much easier to judge how and when to apply tests, and to interpret the significance of their results. They make it clearer whether particular tests are of greater value to rule in or rule out diagnoses, and how that power varies according to the clinically likely prevalence of the condition for the individuals in question. PRU plots should replace PPV and NPV and should be published alongside sensitivity and specificity figures to assist clinicians in interpreting the usefulness of tests.

ACKNOWLEDGEMENTS

I am very grateful to Professor John NS Matthews for his helpful comments during the preparation of this manuscript.

Contribution and competing interests statement: The author declares that the ideas and writing of this manuscript were entirely his own, and that there are no competing interests.

REFERENCES

- 1 **Powell HR**, McCredie DA, Ritchie MA. Urinary nitrite in symptomatic and asymptomatic urinary infection. *Arch Dis Child* 1987;**62**:138–40.
- 2 **American Academy of Pediatrics**. Practice parameter: the diagnosis, treatment, and evaluation of the initial urinary tract infection in febrile infants and young children. *Pediatrics* 1999;**103**:843–52.
- 3 **Vernon S**, Foo CK, Coulthard MG. How general practitioners manage children with urinary tract infection: an audit in the former Northern Region. *Br J Gen Pract* 1997;**47**:297–300.
- 4 **Sackett DL**, Haynes RB, Guyatt GH, et al. *Clinical epidemiology: a basic science for clinical medicine*, 2nd edn. Boston: Little, Brown, 1991.

Submit an eLetter, and join the debate

eLetters are a fast and convenient way to register your opinion on topical and contentious medical issues. You can find the “submit a response” link alongside the abstract, full text and PDF versions of all our articles. We aim to publish swiftly, and your comments will be emailed directly to the author of the original article to allow them to respond. eLetters are a great way of participating in important clinical debates, so make sure your voice is heard.