Section 3b1

Is this evidence about a diagnostic test important?

In deciding whether the evidence about a diagnostic test is important, we will focus on a modern way of thinking about diagnosis that takes into account both components of evidence-based medicine: your individual clinical expertise and the best external evidence. The former is your prior assessment of diagnostic possibilities before you do the test ('prior or pretest probabilities') and the latter is the ability of the test to distinguish patients with and without the target disorder (both the oldfashioned concepts of sensitivity and specificity and the newfangled and more powerful ideas around likelihood ratios). We'll show you how to combine these two elements of EBM to refine your estimates of the target disorder ('posterior or post-test probabilities') and make the diagnosis. Diagnostic tests that produce big changes from pretest to post-test probabilities are important and likely to be useful to you in your practice.

Where do these pretest probabilities come from? Usually they are derived from your own accumulating clinical experience, specific for the setting in which you work and the sorts of patients you see. As a result, pretest probabilities for the same target disorder can vary widely between and within countries and between primary, secondary and tertiary care. We have summarized some published pretest probabilities in Table 3b1.1 and more are available from our Website.

Suppose that you're working up a patient with anemia and think that the probability that they have iron deficiency anemia is 50%; that is, the odds are about 50–50 that it's due to iron deficiency. When you present the patient to your boss, you ask for an educational prescription to determine the usefulness of performing a serum ferritin on your patient as a means of detecting iron deficiency anemia. Suppose further that, in filling your prescription, you find a systematic

Table 3b1.1 Some pretest probabilities

Patient problem	Clinical setting	Target disorder	Pretest probability
Melena in a 50-year-old man who drinks 25 units of alcohol a week but has no stigmata of liver disease	Emergency room in North America	Varices	5%
		Benign ulcer	55%
		Gastritis	40%
Symptomless 60–69-year-olds	Primary care	Undiagnosed colon cancer: all patients	0.5%
		positive family history	1.5%
Symptomless	Primary care	≥ 75% stenosis of	
Woman 30–39 y/o		one or more coronary	0.3%
60–69 y/o		arteries	8%
Man 30–39 y/o			2%
60–69 y/o			12%
Non-anginal chest pain			
Woman 30–39 y/o			1%
60–69 y/o			19%
Man 30–39 y/o			5%
60–69 y/o			28%
Atypical angina			
Woman 30–39 y/o			4%
60–69 y/o			54%
Man 30–39 y/o			22%
60–69 y/o			67%
Typical angina pectoris			
Woman 30–39 y/o			26%
60–69 y/o			91%
Man 30–39 y/o			70%
60–69 y/o			94%
Symptomless 50 y/o with a solitary pulmonary nodule	Primary care	Cancer for any nodules	50%
		For 3 cm nodules	65%

To find more examples, and to nominate additions to the databank of pretest probabilities, refer to this textbook's Website at: http://cebm.jr2.ox.ac.uk/

review of several studies of this diagnostic test (evaluated against the reference standard of a bone marrow stain for iron), decide that it is valid (based on the guides in Tables 3a3.2 and 3a1.1), and find their results as shown in Table 3b1.2. By the time you've tracked down and studied the external evidence, your patient's serum ferritin comes back at 60 mmol/L. How should you put all this together?

As you can see from Table 3b1.2, your patient's result places them in the top row of the table, either in cell **a** or cell **b**. From that fact you would conclude several things: first, you'd note that 90% of patients with iron deficiency have serum ferritins in the same range as your patient, (**a/(a+c)**, and that property, the proportion of patients with the target disorder who have positive test results, is called sensitivity.

Table 3b1.2 Results of a systematic review of serum ferritin as a diagnostic test for iron deficiency anemia

		Target disorder (iron deficiency anemia)		Totals
		Present	Absent	
Diagnostic test result (serum ferritin)	Positive (<65 mmol/L	731 **a**	270 **b**	1001 **a+b**
	Negative (≥65 mmol/L	78 **c**	1500 **d**	1578 **c+d**
	Totals	809 **a+c**	1770 **b+d**	2579 **a+b+c+d**

Sensitivity = **a/(a+c)** = 731/809 = 90%
Specificity = **d/(b+d)** = 1500/1770 = 85%
LR+ = sens/(1–spec) = 90%/15% = 6
LR– = (1–sens)/spec = 10%/85% = 0.12
Positive predictive value = **a/(a+b)** = 731/1001 = 73%
Negative predictive value = **d/(c+d)** = 1500/1578 = 95%
Prevalence = **(a+c)/(a+b+c+d)** = 809/2579 = 32%
Pretest odds = prevalence/(1–prevalence) = 31%/69% = 0.45
Post-test odds = pretest odds × likelihood ratio
Post-test probability = post-test odds/(post-test odds +1)

And you might also note that only 15% of patients with other causes for their anemia have results in the same range as your patient,* which means that your patient's result would be about six times as likely (90% / 15%) to be seen in someone with, as opposed to someone without, iron deficiency anemia and that's called the likelihood ratio for a positive test result. Furthermore, since you thought ahead of time (before you had the result of the serum ferritin) that your patient's odds of iron deficiency were 50–50, that's called a pretest odds of 1:1 and, as you can see from the formulae towards the bottom of Table 3b1.2, you can multiply that pretest odds of 1 by the likelihood ratio of 6 to get the post-test odds of iron deficiency anemia after the test: 1×6 = 6. Since, like most clinicians, you may be more comfortable thinking in terms of probabilities than odds, this post-test odds of 6:1 converts (as you can see at the bottom of Table 3b1.2) to a post-test probability of 6/(6+1) = 6/7 = 86%. So it looks like you've made the diagnosis and this diagnostic test looks worthwhile.

(To check yourself out on these calculations, try the same ferritin result for a patient who, like those in the table, has a pretest odds of 0.47;† you'll know you did it right if you wind up with an answer identical to its equivalent, the positive predictive value.)

Extremely high values of sensitivity and specificity are useful, but not for the reasons you may think.‡ When a test has a very high sensitivity (such as the loss of retinal vein pulsation in increased intracranial pressure), a negative result (the presence of pulsation) effectively rules out the diagnosis (of raised intracranial pressure) and one of our clinical clerks suggested that we apply the mnemonic SnNout to such findings (when a sign has a high *Sen*sitivity, a *N*egative result

* The complement of this proportion is called specificity and it describes the proportion of patients who do not have the target disorder who have negative or normal test results, **d/(b+d)**.
† The post-test odds are 0.45 × 6 = 2.7 and the post-test probability is 2.7/3.7 = 73%. Note that this is identical to the positive predictive value.
‡ On first encounter, most learners think that tests with high sensitivity rule in diagnoses and tests with high specificity rule them out; the reverse is the case.

rules *out* the diagnosis). Similarly, when a sign has a very high specificity (such as a fluid wave for ascites), a positive result effectively rules in the diagnosis (of ascites); not surprisingly, our clinical clerks call such a finding a SpPin (when a sign has a high *Sp*ecificity, a *P*ositive result rules *in* the diagnosis). We've listed some SpPins and SnNouts in Table 3b1.3 and have generated a longer list on our Website.

Although the serum ferritin determination looks impressive when viewed in terms of its sensitivity (90%) and specificity (85%), the newer way of expressing its accuracy with likelihood ratios reveals its even greater power and, in this particular example, shows how we can be misled by the fact that the old sensitivity–specificity approach restricts us to just two levels (positive and negative) of the test result. Most test results, like serum ferritin, can be divided into several levels and in Table 3b1.4 we show you a particularly useful way of dividing test results into five levels. When this is done, one extreme level of the test result can be shown to rule in the diagnosis and in this case you can SpPin 59% of the patients with iron deficiency anemia, despite the unimpressive sensitivity (59%) that would have been achieved if the ferritin results had been split at this level. Likelihood ratios of 10 or more, when applied to pretest probabilities of 33% or more (.33/.67 = pretest odds of 0.5) will generate post-test probabilities of 5/6 = 83% or more. Moreover, the other extreme level can SnNout 75% of those who do not have iron deficiency anemia (again despite a not very impressive specificity of 75%). Likelihood ratios of 0.1 or less, when applied to pretest probabilities of 33% or less (.33/.67 = pretest odds of 0.5) will generate post-test probabilities of 0.05/1.05 = 5% or less. Two other intermediate levels can move a 50% prior probability (pretest odds of 1:1) to the useful but not usually diagnostic post-test probabilities of 4.8/5.8 = 83% and 0.39/1.39 = 28%. And one indeterminate level in the middle (containing about 10% of both sorts of patients) can be seen to be uninformative, with a likelihood ratio of 1. We've shown the effects of these sorts of likelihood ratios on these sorts of pretest probabilities in Table 3b1.5.

Table 3b1.3 Some SpPins and SnNouts

Target disorder	SpPin (& specificity) [presence rules in the target disorder]	SnNout (& sensitivity) [absence rules out the target disorder]
Ascites (by imaging or tap)*	Fluid wave (92%)	History of ankle swelling (93%)
Pleural effusion†	Auscultatory percussion note loud and sharp (100%)	Auscultatory percussion note soft and/or dull (96%)
Increased intracranial pressure (by CAT scan or direct measurement)‡		Loss of spontaneous retinal vein pulsation (100%)
Cancer as a cause of lower back pain (by further investigation)§		Age >50 or cancer history or unexplained weight loss or failure of conservative therapy (100%)
Sinusitis (by further investigation)¶		Maxillary toothache or purulent nasal secretion or poor response to nasal decongestants or abnormal transillumination or history of coloured nasal discharge
Alcohol abuse or dependency**	Yes to ≥3 of the CAGE questions (99.8%)	
Splenomegaly (by imaging)††	Positive percussion (Nixon method) and palpation	
Non-urgent cause for dizziness‡‡	Positive head-hanging test and either vertigo or vomiting (94%)	

To find more examples, and to nominate additions to the databank of SpPins and SnNouts, refer to this textbook's Website at: http://cebm.jr2.ox.ac.uk/

* JAMA 1992; 267: 2645–8.
† J Gen Int Med 1994; 9: 71–4.
‡ Arch Neurol 1978; 35: 37–40.
§ JAMA 1992; 268: 760–5.
¶ JAMA 1993; 270: 1242–6.
** Amer J Med 1987; 82: 231–5.
†† JAMA 1993; 270: 2218–21.
‡‡ JAMA 1994; 271: 385–8.

Table 3b1.4 The usefulness of five levels of a diagnostic test result

Diagnostic test result		Target disorder present		Target disorder absent		Likelihood ratio	Diagnostic impact
Serum ferritin (mmol/L)		Number	%	Number	%		
Very positive	<15	474	59%	20	1.1%	52	Rule in SpPin
Moderately positive	15–34	175	22%	79	4.5%	4.8	Intermediate high
Neutral	35–64	82	10%	171	10%	1	Indeterminate
Moderately negative	65–94	30	3.7%	168	9.5%	0.39	Intermediate low
Extremely negative	≥95	48	5.9%	1332	75%	0.08	Rule out SnNout
		809	100%	1770	100%		

Table 3b1.5 Some post-test probabilities generated by five levels of a diagnostic test result

Likelihood ratio	Post-test probability of the target disorder for different pretest probabilities						Diagnostic impact
	Pre-test 5%	Pre-test 10%	Pre-test 20%	Pre-test 30%	Pre-test 50%	Pre-test 70%	
Very positive 10	34%	53%	71%	81%	91%	96%	Rule in SpPin
Moderately positive 3	14%	25%	43%	56%	75%	88%	Intermediate high
Neutral 1	5%	10%	20%	30%	50%	70%	Indeterminate
Moderately negative 0.3	1.5%	3.2%	7%	11%	23%	41%	Intermediate low
Extremely negative 0.1	0.5%	1%	2.5%	4%	9%	19%	Rule out SnNout

Finally, there's an easier way of manipulating all these probability↔odds calculations and a nomogram for doing so appears as Figure 3b1.1 and in the pocket cards that come with this book. You can check out your understanding of this nomogram by replicating the results in Table 3b1.5.

To your surprise (we reckon!) your patient's test result generates an indeterminate likelihood ratio of only 1 and the test which you thought might be very useful, based on the sensitivity and specificity way of looking at things, really hasn't been helpful in moving you toward the diagnosis, so you'll have to think about other tests (including perhaps the reference standard of a bone marrow examination) to sort this out.

More and more reports of diagnostic tests are providing multilevel likelihood ratios as measures of their accuracy. When they only report sensitivity and specificity, you can sometimes find a table with more levels and generate your own set of likelihood ratios or you can find a scatter plot (of test results versus diagnoses) that is good enough for you to be able to split into levels. Or, if all you have is sensitivity and specificity, you can generate likelihood ratios from them by reference to the formulae in Table 3b1.2 (the likelihood ratio for a positive test result = LR+ = sensitivity/[1-specificity] and the likelihood ratio for a negative test result = LR − = [1-sensitivity]/specificity).

Some reports into the accuracy of diagnostic tests go beyond even likelihood ratios and one of them deserves mention here. This extension considers multiple diagnostic tests as a cluster or sequence of tests for a given target disorder. These multiple results can be presented in different ways, either as clusters of positive/negative results or as multivariate scores, and in either case they can be ranked and handled just like other multilevel likelihood ratios.

In any event, having decided that a diagnostic test produces important changes from pretest to post-test probabilities, you might want to study the final issue, described in Section 4.1, of how to integrate the results of this critical appraisal with your individual clinical expertise and apply the results

Nomogram for interpreting diagnostic test result

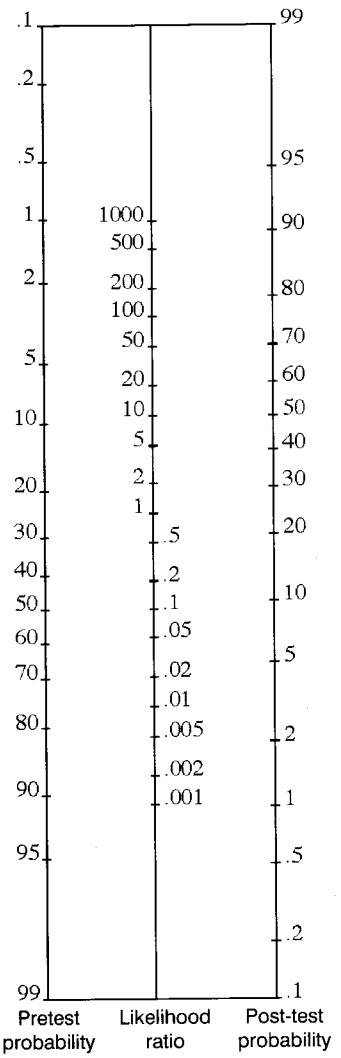

Figure 3b1.1 A likelihood ratio nomogram. Adapted from Fagan T J 1975 Nomogram for Bayes's Theorem (c). New England Journal of Medicine 293: 257

to your own patient (but if you jumped to this second step without first determining whether the evidence about this diagnostic test was valid, you'd better go back to Section 3a1 first!).

Further reading

Sackett D L, Haynes R B, Guyatt G H, Tugwell P. Clinical epidemiology: a basic science for clinical medicine, 2nd edn. Little, Brown, Boston, 1991. Chapter 4 (for interpreting diagnostic tests).

Jaeschke R, Guyatt G H, Sackett D L for the Evidence-Based Medicine Working Group. Users' guides to the medical literature. VI. How to use an article about a diagnostic test. A. Are the results of the study valid? JAMA 1994; 271: 389–91. B. What are the results and will they help me in caring for my patients? JAMA 1994; 271: 703–7.