Diagnostic accuracy of a bedside D-dimer assay and alveolar dead-space measurement for rapid exclusion of pulmonary embolism
Kline JA, Israel EG, Michelson EA, O'Neil BJ, Plewa MC, Portelli DC
JAMA 285: 761-768

Critical Appraisal of a Diagnostic Test

Are the results of the study valid

1. Was there an independent 'blind' comparison with a reference standard?

   A proposed new diagnostic test needs to be compared with the 'truth'. The best determinant of the truth is an appropriate reference standard – something as perfect as possible (biopsy, surgery, PM, long-term follow-up), sometimes known as a gold standard, that can be used on every patient as well as the new test.

   You will first need to ask if the reference standard is acceptable. If not then the paper won't give you a lot of information about the new test.

   There were two reference methods: V/Q scans and CT scans and everyone had at least one of these. Depending on the physician, V/Q scans and CT scans were supplemented with other tests such as sonographs for DVT and pulmonary angiography. Additionally there was a six month follow up to detect possible non-diagnosed incidences of PE.

   There was concern by the authors about the additional testing leading to over classification, and other concerns that CT scanning might produce a higher rate of false positives. The aim of the gold standard is to define the truth and so the additional tests give result in a better picture of the truth. It would probably have been better to stick to a single standard, particularly as CT scanning was much less used
   .
   Next consider if each test – reference and new – has been carried out by different people (independently) who are also unaware (blind) of the results obtained by the other person.

   As soon as the diagnostic standard was requested then the diagnostic tests were carried out by members of the team in A&E. The results of these were not transmitted to the radiologists and others who carried out the standard tests. On the other hand the diagnostic tests were carried out after collection of clinical information and so the diagnostic test should be viewed in this context.

2. Did the patient sample include an appropriate spectrum of patients to whom the test will be applied in clinical practice?

   It is relatively easy for a test to distinguish someone who is very ill from someone who is healthy but the test needs to be able to work inside those evtremes.

   The hospital were urban academic centres so that might attract patients with slightly different characteristics from the normal population, even though those patients who had been referred were excluded. It was suggested that the probability of PE in someone over 18 coming into the A&E was about 17% was typical of a teaching hospital, with most patients being at moderate risk of PE, others being either high or low risk – so a broad spectrum.

3. Did the results of the test influence the decision to perform the reference standard?

   The apparent usefulness of a diagnostic test can be distorted if the result influences whether or not the reference standard is used. This is known as 'verification' bias.

   In the study once a patient was identified as needing diagnosis the standard was always performed.

4. Were the methods of performing the test described in sufficient detail to permit replication?

It is obviously important to be able to use the method if it appears useful and so its description not only allows that but will also present information on how practical the new test is.

Detailed information was provided about exactly when and how the two diagnostic tests were undertaken with information on computations and the need for training in order to interpret the D-dimer results.

What are the results?

5. Are likelihood ratios, or the data necessary to calculate likelihood ratios, provided?

Signs and symptoms displayed by a patient all provide information about how likely it is that the patient is suffering from a particular condition before any tests are done – this is **the pre-test probability**. Because people presenting to A&E, compared with those who go along to their GP, may have different signs and symptoms, or may express these at different levels, then the pre-test probability will be different in these differing clinical settings.

The pre-test probability is needed to:

- Interpret a diagnostic test result
- Select diagnostic tests for use in different settings
- Helping to decide whether to treat an individual without further tests
- Helping to decide whether to test or not

Even with a particular situation its actual estimation may vary considerably amongst different doctors although that can be improved with educational initiatives. Because the diagnostic test is used to modify or support that previous clinical impression it is important that it is as precise as possible.

In the context of reading a paper to evaluate a diagnostic test we need to know what the pre-test probability is in the various settings that they consider you may want to use the test. A test may be very useful for patients in A&E but may be useless in the GP setting.

It is calculated simply as the number of people with the disorder / number of people with the symptoms in a particular setting – this may also be referred to as the prevalence.

In the study undertaken the assessment of individual pre-test probability was made using a collection of signs, symptoms and risk factors to provide a clear scoring system to determine pre-test probability for each patient defined as no, low or high risk. This is helpful when assessing an individual patient but when looking at a diagnostic test we are more interested in the pre-test probability for the particular setting – this has been assessed using the gold standard diagnosis and is 16.8% for the described setting.

How useful a diagnostic test is will depend on how good it is at identifying the disorder. **The likelihood ratio** gives that information. The likelihood ratio is defined as:

Likelihood of the particular test result in someone with the disease / likelihood of the particular test result in someone without the disease.

The likelihood ratio (positive result) is the likelihood of getting a ***positive*** test result in someone with the disease / likelihood of the particular test result in someone without the disease.

The likelihood ratio (negative result) is the likelihood of getting a ***negative*** test result in someone with the disease / likelihood of the particular test result in someone without the disease.

(Note that the underlined words are the same in each case)

Many papers will talk about sensitivity and specificity of a test but the likelihood ratio has advantages because it is not as influenced by prevalence. To understand this you will find it useful to work though the accompanying evaluation of exercise testing for CAD. This is a worksheet that deals with the calculations. Each page asks you to fill our numbers in a table – don't worry the answers are given on the next page. These will help you understand the following terms:

Prevalence
Pre-test and post-test probability
Sensitivity
Specificity
Positive predictive value
Negative predictive value
Likeilhood of a positive test
Likelihood of a negative test

Likelihood ratios provide a number that can be combined with pre-test information to give an assessment that includes both. There is a simple calculation:

Post test odds of disease = pre test odds x likelihood ratio

This looks simple but note that I am now talking about odds, and not probability. If you are familiar with horse racing you probably won't have any difficulty but many people find the difference confusing.

For example if the pre-test probability of an MI was 50% then that would also be the same as saying that the chance was 50:50, or 1:1 – a pre test odds of 1. This is quite simple but it can get more difficult and the following formulae can be used to convert probability to odds:

Odds = probability / (1-probability) = 0.5 / (1-0.5) = 0.5 / 0.5 = 1 (from the example above)

You may also want to reconvert the odds to a probability.

Probability = odds / (odds + 1) = 1 / (1+1) = ½ = 0.5 (from the example above)

If you don't want to do these simple calculations you could use a nomogram – this is a chart which allows you just to work in probabilities – several are published and are usually available in evidence-based medicine books but the calculation are otherwise very straightforward.

Knowledge of the likelihood ratio alone can be useful.

- If LR >10 or <0.1 then the test alone is likely to give you a conclusive answer without worrying about pre-test probabilities
- If LR is between 5-10, or 0.1-0.2 then these will have a modertate influence on the pre-test probability
- If LR is between 2-5, or 0.2-0.5 then these will have only a small influence on the pre-test probability
- If LR is between 1-2, or 0.5-1 then the test is unlikely to be useful

Table 3 gives all the results, both for the individual tests and for the combined tests. The results are best interpreted using the LRs.

Comparing them with the table above we note that they are useful but are not likely to influence the pre-test probability greatly if the test is positive (LR 2.03 – combined test, 2.83 dead space and 2.85 – D-dimer).

In contrast, getting a negative combined test (LR 0.03), or negative D-dimer test (LR 0.09), is potentially very useful as it is very unlikely that these will be seen in someone with PE. On

the other hand the dead-space fraction test, if negative, will not be a particularly good pointer as to whether someone has PE or not (LR 0.43).

They have quoted some post test probabilites but we can calculate those we are interested in:

For a combined positive test:

Pre-test odds = 0.168 / (1-0.168) = 0.20

Post-test odds = 0.20 x 2.03 = 0.406

Post-test probability = 0.406 / (1+0.406) = 0.29 – in other words if the combined test is positive then the probability of PE increases to 29% from a pre-test probability of 17%

For a positive D – dimer test:

Pre-test odds = 0.168 / (1-0.168) = 0.20

Post-test odds = 0.20 x 2.85 = 0.57

Post-test probability = 0.57 / (1+0.57) = 0.36

A positive D- dimer test will therefore increase the probability of PE from 17% to 36%

For a combined negative test:

Pre-test odds = 0.168 / (1-0.168) = 0.20

Post-test odds = 0.20 x 0.03 = 0.006

Post-test probability = 0.006 / (1+0.006) = 0.006

If the combined test is negative then the probability of PE decreases to 0.6% from 17%
In the paper, page 766, they give this as decreasing to 0.75%.

Will the results help me in caring for my patients (or improve clinical effectiveness)?

6. Will the reproducibility of the test and its interpretation be satisfactory in my setting?

A test is valuable if it gives the same results on repeat testing when the patient remains stable. Of course we don't necessarily expect the quantitative result to be exactly the same – but certainly close. Differences are a result of error in the measurement scales but can also result from differences in clinical interpretation and that will depend on expertise and training.

If you have poor reproducibility but the LRs are very high then the test will still be useful. If you have very high reproducibility then either the test is very simple, or those doing it are very skilled. If the latter then those people doing the test in your setting may not be as skilled.

The study participants had a 90 minute training session, watched a training videotape and were given a booklet with photographs of assay interpretations. The test is obviously not that simple and it seems unlikely that most people would receive such intensive training for these tests – there is likely therefore to be more variability (error) in the results so probably the test will not prove as good as suggested in the paper.

7. Are the results applicable to my patient?

Here we are interested if our patient is likely to have a disease profile similar to the one described in the paper. If your patients tend to have less severe disease on average then the LR will not be as high as in the paper and so the test will not be as useful.

Their assessment was that 16.8% pre-test probability of symptomatic adults was consistent with an A&E in a teaching hospital). Reducing the prevalence (or pre-test probability) will lower the sensitivity and make the LR closer to 1 – making the test less useful. This is shown in the following example:

	Disease positive	Disease negative	TOTAL
Test positive	63	153	216
Test negative	1	163	164
TOTAL	64	316	380

Prevalence is 64/380 = 0.168

Sensitivity = 63/64 = 0.984

LR positive test = (63/64) / (153/316) = 0.984 / 0.484 = 2.03

LR negative test = (1/64) / (163/316) = 0.0156 / 0.516 = 0.03

This is the same as on page 766.

Reducing the prevalence (I've just made these up but kept the totals the same)

	Disease positive	Disease negative	TOTAL
Test positive	24	192	216
Test negative	1	163	164
TOTAL	25	355	380

Prevalence is 25/380 = 0.066

Sensitivity = 24/25 = 0.96

LR positive test = (24/25) / (192/355) = 0.96 / 0.541 = 1.77

LR negative test = (1/25) / (163/355) = 0.04 / 0.459 = 0.087

Therefore both LRs have moved towards 1, making the test less useful.

Therefore knowledge of the prevalence of disease in your own situation is important when valuing the usefulness of this test.

8. Will the results change my management?

There is normally a level at which you would decide that a person either does or does not have a disease, but in between is a grey area where another test would help. The actual probability of disease level you chose will depend on how important it is if you over treat some (the treatment could be toxic) and miss others (the implication may be that they will die). If most patients have test results with LRs near one then the test will not be very useful. The usefulness therefore depends on what proportion of your patients who are suspected of having the disease have very high of very low LRs.

Those people with high risk have very high LRs and those with low risk very low LR, leaving what often is a substantial group with a medium LR (closer to 1) – so if you have diagnostic tests that provide continuous data, rather than a yes-no answer, then if would be useful to examine that aspect.

Some information about levels of risk are provided in the clinical probability assessment but there is no grading within the two tests for us to calculate the actual LRs at the different risk levels. All you can do is note that they say that only 30% of their population was at low risk from their clinical probability assessment and see how that compares with your population.

9. Will patients be better off as a result of the test?

The important question is whether the test provides additional information that leads to a change in management and that is beneficial to the patient. Often there will be a change in management but whether it has a beneficial effect may be more doubtful.

Only a combined negative test result gives an LR that is a significant pointer to the patient not having a PE. Avoiding doing extra tests in these patients saves them from exposure to more dangerous procedures and cuts costs.