INTERPRETATION OF DIAGNOSTIC TESTS

sensitivity and specificity

In medicine, we rely on diagnostic tests to identify or exclude diseases. Tests are particularly relevant when a treatment is to be implemented. Table 1 shows the **case of a test carried out on 100 children to find the presence of a disease.** The results of the test are compared with those of the reference test (or *gold standard*) for that diseases, that correctly identifies the sick children (in this example 34) and the healthy ones, i.e. that do not have the disease (in this example 66). Four different situations can arise:

1. the test correctly detects the presence of the disease: cases are defined **true positives (TP)**, in this example 29 children;
2. the test correctly does not detect the presence of the disease: cases are defined **true negatives (TN)**, in this example 64 children;
3. the test wrongly detects the presence of the disease: cases are defined **false positives (FP)**, in this example 2 children;
4. the test wrongly does not detect the presence of the disease: cases are defined **false negatives (FN)**, in this example, 5 children.

Obviously, the less frequent the false positives and false negatives are, the more reliable the test is. However, errors always occur, to a variable extent, and limit the capacity of the tests to distinguish the individuals with and without the disease.

Table 1. *Example of results of a test to detect the presence of a disease, compared with the results of a reference (gold standard) test.*

	Sick *(gold standard)*	**Healthy** *(gold standard)*	
Positive test	29 TP	2 FP	***31***
Negative test	5 FN	64 TN	***69***
	34	***66***	***100***

How to describe the efficacy of a test?

Sensitivity

It is the capacity of a test to detect the disease in someone who is really sick (see figure 1). In the example given in table 1, the test detects the presence of the disease in 29 out of 34 children, i.e. in 85% of cases (29/34 = 0.85).

Sensitivity is crucial when the objective of the test is not to lose a single case (for example, in the event of a serious disease, when an early surgery may be crucial).

If a test has a high sensitivity and the result is negative, probably there is no disease[1].

Figura 1. *Example of a test with high sensitivity (1 false negative) but low specificity (various false positives)*

Sick *Healthy*

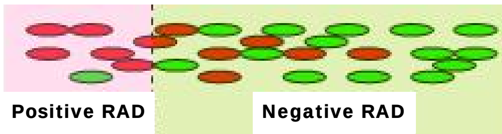

Figura 2. *Example of test with high specificity (1 false positive) but low sensitivity (various false negatives)*

Sick *Healthy*

Specificity

It is the capacity of the test to identify the patients who do not have the disease (see figure 2). In the example in table 1, the test result is correctly negative in 64 out of 66 children without the disease, i.e. in 97% of cases (64/66 = 0.97).

Specificity is crucial when it is necessary to be precise in making a diagnosis (that, for example, could lead to a demolitive surgery).

If a test has a high specificity and the result is positive, probably the disease is present and it is correct to proceed with the planned treatment[1].

NB In general, **tests with high sensitivity usually do not show high specificity** (they can give more easily false positive results). On the other hand, **very specific tests usually do not show high sensitivity** (they can give more easily false negatives results).

BIBLIOGRAPHY

1. Altman DG et al. Diagnostic tests 1: sensitivity and specificity. BMJ 1994;308:1552
2. MHRA. Group A streptococcus rapid antigen detection test kits: a review of evaluation literature. Department of Health. Jan 2005
3. Altman DG, Bland JM Predictive values. Statistics Notes: Diagnostic tests 2: BMJ 1994;309:102
4. Models for calculation of post-test likelihood. Available on-line at http://www.robertobuzzetti.it

IN PRACTICE: RADs in pharyngotonsillitis

- The **sensitivity** of RAD of II generation (currently the most used) is generally lower than 95%. Therefore, in case of a strong clinical suspicion, **if the RAD has a negative result** a confirmatory throat swab culture would be necessary.
- The **specificity** of RAD is generally higher than 95%, therefore **in case of a positive RAD** the probability that is a false positive is lower than 5% and an antibiotic treatment is indicated, without further confirmatory tests (*see page 7 of the information pack).*

INTERPRETATION OF DIAGNOSTIC TESTS

predictive value and likelihood ratio

Predictive value: it depends on the frequency of the disease

The information on sensitivity and specificity of the test are important, but does not directly answer the two questions that principally concern the doctor:

- **If a test has a positive result, how likely is that my patient really has the disease?** This information is given by the Positive Predictive Value (PPV), that is the percentage of sick patients who test positive. In the example in table 1, 29 out of 31 children with a positive test have the disease, i.e. the PPV is 94% (29/31 = 0.94).
- **If a test has a negative result, how likely is that my patient does not have the disease?** This information is given by the Negative Predictive Value (NPV), i.e. the percentage of healthy patients with a negative test. In the example in table 1, 64 out of 69 patients with a negative test do not have the disease, i.e. the NPV is 93% (64/69 = 0.93).

The predictive values (positive and negative) depend on the frequency of the disease in the population under study. For example, if in a given setting the disease is rare, the "true positives" will be few; in this situation even detecting few "false positive" will considerably reduce the PPV of the test[3].

Table 2. *Example of results of a test to detect the presence of a disease, compared with the results of a reference (gold standard) test.*

	Sick *(gold standard)*	**Healthy** *(gold standard)*	
Positive test	29	2	***31***
Negative test	5	64	***69***
	34	***66***	***100***

Likelihood ratio: useful but ... complex

The doctor may be interested to know how likely the result of a test is. This information is given by the **likelihood ratio (LR)**, which is a characteristic of the test itself which value does not change with the frequency of the disease in the population under study[3].

The LR (likelihood ratio) is a ratio among likelihoods. In particular:

- The **positive LR (LR+)** tells us how much it increases the probability of having a positive result in individual with the disease compared to the healthy ones. In the example in table 2, the likelihood of a positive test in children with a disease (i.e. sensitivity) is 29 out of 34 (85%) - thus fairly high - whereas the likelihood of a positive test in children without a disease is 2 out of 66 (3%) - thus fairly low. The LR+ tells us that it is **28 times more likely** to find a positive result in a sick child than in a healthy one (85% compared to 3%). The higher the LR+, the higher the probability to have the disease if the test is positive.
- The **negative LR (LR-)**, on the other hand, tells us how it decreases the probability of having a negative result in individual with the disease compared to those without it. In the example in table 2, the likelihood of a negative test in sick children is 5 out of 34 (15%) - thus fairly low - whereas the likelihood of a negative test in healthy children (i.e. specificity) is 64 out of 66 (97%) - thus fairly high. The LR- tells us that it is **6-7 times less likely** to find a negative result in a sick child than in a healthy one (15% is 6-7 times smaller than 97%). The lower the LR- the smaller the probability to have the disease if the test is negative.

The LRs (+ or -) are useful measures because they merge the information given by sensitivity and specificity of the test in a single parameter which value does not change with the frequency of the disease in the population under study.

IN PRACTICE: RAD in pharyngotonsillitis

- The **positive** and **negative predictive value** varies with the prevalence of the disease studied. In a population of children with an estimated prevalence of *S. pyogenes* of 35%, the PV+ and the PV- for RAD of second generation are on average 97% and 90% respectively[2].
- The **likelihood ratio** is an intrinsic characteristic of a test. On average, for RAD of second generation the LR+ is 40 and the LR- is 0.20. Thus, for example, if a child has a probability of having an infection due to *S. pyogenes* of 35% before the test, the likelihood will increase to 96% in case the RAD is positive, whereas if the RAD is negative, the likelihood will drop to 10%[2, 4].