15. Interpretation of diagnostic results in aquatic animal health

De Blas I., Muniesa A., Ruiz-Zarzuela I.

in

Zrncic S. (ed.).
Diagnostic Manual for the main pathogens in European seabass and Gilthead seabream aquaculture

Zaragoza : CIHEAM
Options Méditerranéennes : Série B. Etudes et Recherches; n. 75

2020
pages 137-142

15. Interpretation of diagnostic results in aquatic animal health

I. de Blas, A. Muniesa and I. Ruiz-Zarzuela

University of Zaragoza, Faculty of Veterinary, Laboratory of Fish Diseases, Zaragoza, Spain

Table of contents

15.1. Assessment of the accuracy of a diagnostic test

15.2. Setting up a diagnostic test

15.3. Diagnostic accuracy vs laboratorial accuracy

15.4. A possible solution to overcome low accuracy of diagnostic test

References

15.1. Assessment of the accuracy of a diagnostic test

Both pathologists and clinicians must always keep in mind that diagnostic tests are not perfect (even Real-Time qPCR test), although we sometimes find it difficult to accept. Once we have assumed this fact, we must reflect on the consequences of possible diagnostic errors in our interpretation of the results. For further information about diagnostic testing, we recommend consulting Dohoo *et al*. (2003), Thrusfield (2007), Gordis (2014), and the OIE manual "*Principles and methods of validation of diagnostic assays for infectious diseases*".

Classically, the reliability of a diagnostic test has been characterized by assuming the existence of a perfect reference test, known as a **gold standard**, which indicates the true health status of the individuals analysed. So, all the positive individuals for this test are diseased and all the negative results correspond to healthy animals. Actually, this test does not exist, but it serves to understand the concepts that we are going to explain next. An alternative to the gold test is to use two groups of animals with known health status, i.e., a group of animals that have been experimentally infected and have manifested the disease to be diagnosed and another group of disease-free animals (from historically free territories or using Specific Pathogen Free (SPF) animals).

If we were to diagnose a sample with two diagnostic tests: the gold standard and the test to be evaluated, we would obtain a contingency table (Fig. 26), where we would find four possible combinations of diagnostic results. There are the concordant results: true positives, TP (diseased animals diagnosed as positive by the assessed test) and true negatives, TN (healthy animals diagnosed as negative by the assessed test). The problem arises in the discordant results since they are errors committed by the evaluated test: false positives, FP (healthy animals diagnosed as positive by the assessed test) and false negatives, FN (sick animals diagnosed as negative by the assessed test). Therefore, the purpose of designing and interpreting a diagnostic test is to minimize the number of false positives and/or negatives that may appear.

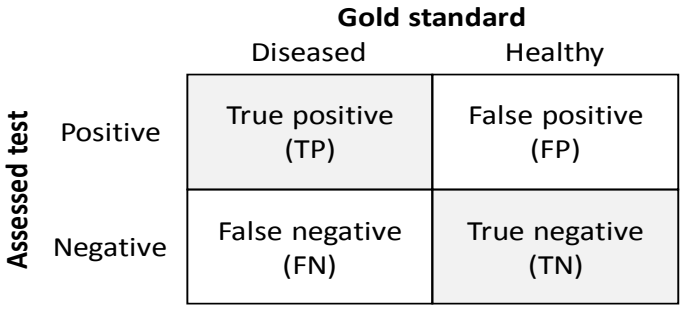

Fig. 15.1. Contingency table for the assessment of a diagnostic test

Considering this contingency table, it is possible to estimate the success rates of a diagnostic test through measures of sensitivity and specificity, respectively; they are independent of the prevalence.

Sensitivity (Se) is defined as the probability of obtaining a positive result when an animal is diseased and corresponds to the following conditional probability formula:

$$Se = p(+|\, D) = TP/(TP + FN)$$

Similarly, **specificity** (Sp) is defined as the probability of obtaining a negative result when an animal is healthy and corresponds to this other conditional probability formula:

$$Sp = p(-|\, H) = TN/(TN + FP)$$

In order to facilitate the understanding of these concepts, we are going to propose an example where 622 trout have been selected, of which 227 are surely diseased and 395 are SPF trout. We wish to assess a diagnostic test which is applied to all trout, obtaining 190 positive results in the group of diseased trout (and therefore 37 negatives) and 303 positive results in the group of healthy trout (and therefore 92 negatives).

If we represent the data of the example to scale, considering the groups created by the gold standard (Fig. 27), we can see that in the group of diseased fish the proportion of successes (TP) is quite high. Something similar happens with the proportion of successes in the group of healthy fish (TN). Therefore, in this example, the sensitivity is 83.70% (=190/227) and the specificity 76.71% (=303/395).

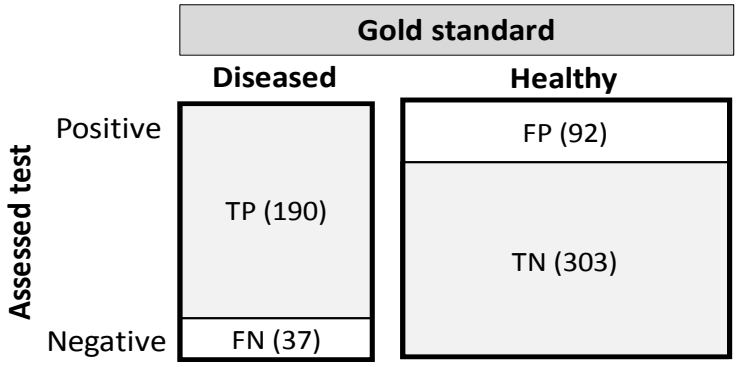

Fig. 15.2. Contingency table for calculation of sensitivity and specificity

As can be seen in (Fig. 26) and in the formulas, the existence of false negatives reduces the sensitivity while false positives affect the specificity. That is to say, when a test has a perfect sensitivity (100%) there will be no false negatives and therefore all negative results will correspond to healthy animals, which would allow us to safely rule out that a negative animal is diseased. While a test with perfect specificity will not have false positives, and all the results obtained will be diseased animals, therefore a positive result will confirm that it is a diseased animal.

15.2. Setting up a diagnostic test

The problem that arises when a diagnostic test is developed is that when we try to improve the sensitivity, the specificity will worsen and vice versa. To explain this fact, we will use the example of an analytical diagnostic test based on quantitative results (for example, an ELISA test). We will apply the test to a group of diseased animals and to another group of healthy animals obtaining a distribution of results similar to the one shown in (Fig. 28). As can be seen, there is an overlap of both distributions which forces us to look for the most suitable cut-off value (or threshold value) to optimize diagnostic reliability, i.e. minimizing false positives and false negatives. Once the cut-off value has been established, all results with values lower than the cut-off value are considered negative and those with higher values are considered positive.

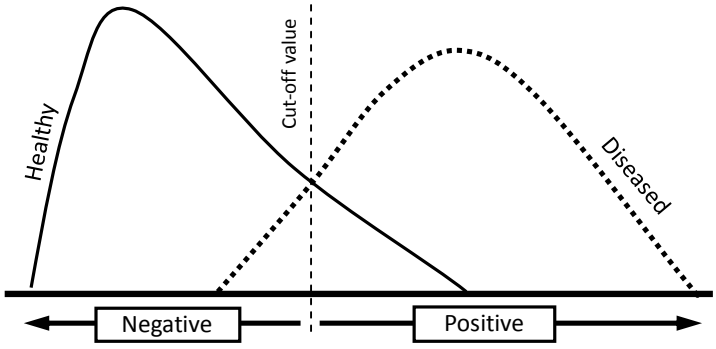

Fig. 15.3. Distribution of results of a diagnostic test stratified by health status

Once the cut-off value is set, we can build the contingency table used to assess a diagnostic test. Diseased animals with values above the cut-off value will be the true positives and healthy animals with values below the cut-off value will be the true negatives. Unfortunately, there will be false positives (the right tail of the distribution of healthy animals) and false negatives (the left tail of the distribution of diseased animals).

If we choose to minimize false negatives (by reducing the cut-off value), we will be able to increase the sensitivity until it becomes perfect (100%) (Fig. 29 a), although at the cost of significantly increasing the number of false positives and therefore worsening the specificity. If, on the other hand, we decide to minimize false positives (by increasing the cut-off value), the specificity will increase until it becomes perfect (Fig. 29 b), but with the corresponding increase in false negatives that leads to a worsening of sensitivity. Normally there is a tendency to maximize sensitivity (minimizing false negatives) at the cost of worsening specificity (increasing false positives). Although there are statistical techniques to identify the value that maximizes both values (using ROC curves).

In some cases, the test results which are under the overlapping area of the curves are considered doubtful, meaning that the test cannot discriminate between false positive and false-negative results.

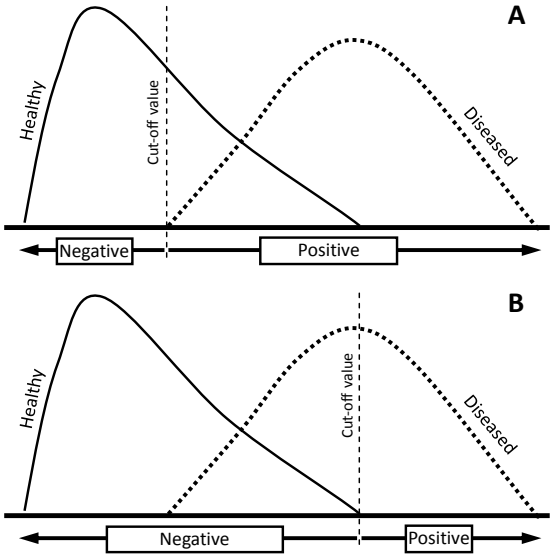

Fig. 15.4. Effect of cut-off value in the diagnostic results and diagnostic accuracy.

15.3. Diagnostic accuracy vs laboratorial accuracy

An interesting aspect that needs to be reflected upon is why these these types of diagnostic errors are made.

In the case of **false negatives**, we must first consider the latency period of the disease (for example, in the case of serological tests, antibodies take a period of time to develop from the moment of infection and during that period they will not be detectable) or the existence of substances that may inhibit the analytical reaction performed.

At the laboratory level, the term “laboratory sensitivity” is used to refer to the detection limit of a diagnostic test. In (Fig. 30) we can see how the pathogen load of an infected fish begins with a low number of pathogens and the amount increments (it usually coincides with the symptomatic phase) and afterwards in the survival animals (convalescent fish) the pathogens disappear or their number decreases significantly.

The performance of a diagnostic test can vary. This means that a simple PCR (Fig. 30 a) can detect, for example, a concentration of at least 10^4 pathogens/g of tissue; however, a qPCR (Fig. 30 b) have a lower detection limit and can give a positive result with concentration of 10^2 pathogens/g. Therefore, the laboratory sensitivity (detection limit) is correlated with the diagnostic sensitivity, since the probability of a negative result is greater with the simple PCR (because the latency period and convalescence periods are longer) and in consequence the sensitivity is lower than in the second case (qPCR) where the probability of negative results in an infected fish is lower, in which case the sensitivity will be higher. This means that low detection limits produce high diagnostic sensitivities.

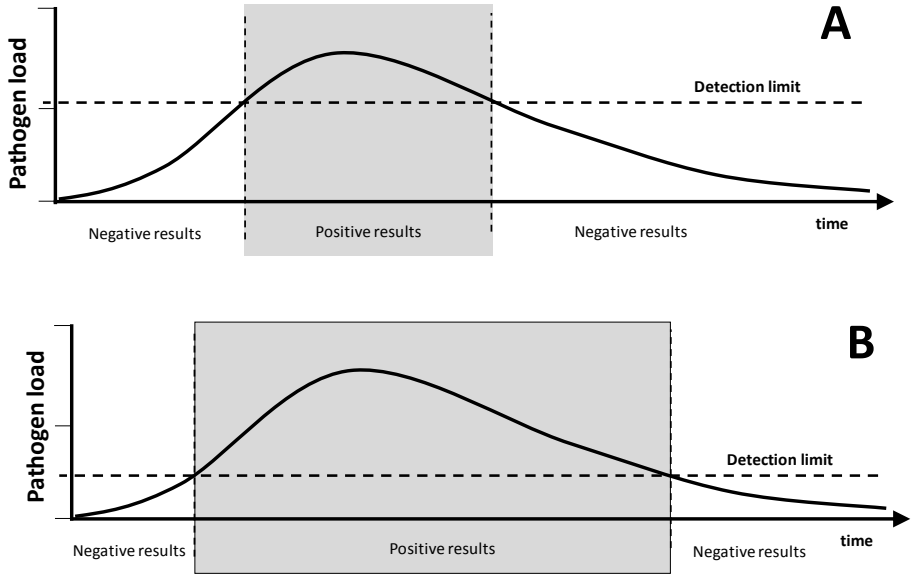

Fig. 15.5. Laboratorial sensitivity vs diagnostic sensitivity

In the case of **false positives** for direct diagnostic tests such as PCR (where the aim is to detect the pathogen) the existence of other organisms with genome fragments identical to the pathogen to be detected should be considered. In the case of indirect diagnosis (where the aim is to find antibodies against the pathogen, such as ELISA) the phenomena of natural immunization (animals recovered from infection) or acquired immunization (vaccinated animals) or serological cross-reactions should be taken into account.

To assess this laboratory specificity, the diagnostic test is tested against several pathogens that we can usually find in the same fish species. For example, the laboratory specificity of a PCR assay for *Aeromonas salmonicida* can be tested with *Aeromonas hydrophila, Vibrio salmonicida, Pseudomonas fluorescens, Shewanella putrefaciens, Yersinia ruckeri*, etc. If these samples do not give positive results, the laboratory specificity of the PCR assay for Aeromonas salmonicida is proven.

15.4. A possible solution to overcome low accuracy of diagnostic test

It is important to keep in mind that a laboratory result gives a punctual indication of the infectious status of an animal linked to a specific time and tissue sampled (i.e blood or organ), as biological activities are highly influenced by internal and external conditions (i.e. water temperature, age, hormonal balance, etc.). It is therefore important to know if the test used is validated for the actual situation, animal and sample tested.

To reduce the problem correlated with false positive or false negative results, the following actions can be undertaken:

- Use a second independent laboratory test, if available, with different sensibility/specificity characteristics in order to confirm the previous results. This should be done according to a specific procedure (described in 1,2,3).

- Change the kind of sample tested (i.e. collect different organs). The optimal organ for testing may vary in time through-out the infection process and between infectious agents. This may be especially important in the early phase of infection.
- Increase the number of samples submitted. This will not influence the result of the individual animals tested, but increase the certainty of the target population being infected or not.
- Repeat the sample a few weeks later. This will allow the disease to progress or the fish to be clear of the cause that resulted in the false positive/doubtful reaction. This is particularly useful when the laboratory test targets antibody level/presence, which evolves quite rapidly.

In conclusion, the correct interpretation of laboratory results requires the knowledge of fish biology and farm history, as well as test performance and pathogen characteristics and epidemiology.

References

Dohoo I., Martin W., Stryhn H., 2003. Chapter 5. Screening and diagnostic test. In: Veterinary Epidemiologic Research. Charlottetown, AVC Inc, 95-120.

Gordis L., 2014. Chapter 5**.** Assessing the validity and reliability of diagnostic and screening tests. In: Epidemiology (5th edition). Lippincott Williams and Wilkins, 88-115.

OIE, 2019. Manual of diagnostic test for aquatic animals. Chapter 1.1.2. Principles and methods of validation of diagnostic assays for infectious diseases.
https://www.oie.int/index.php?id=2439&L=0&htmfile=chapitre_validation_diagnostics_assays.htm

Thrusfield M., 2007. Chapter 17. Diagnostic testing. In: Veterinary Epidemiology (3rd edition). Philadelphia: Elsevier Saunders, 305-330.