Report 4: COVID-19: analysis of two serological tests

Francesco Conforte, s277683,
ICT for Health attended in A.Y. 2020/21

December 20th, 2020

1 Introduction

Spreading of COVID-19 (COrona VIrus Disease 19) infection can be reduced with early detection of ill people, so that they can start quarantine as soon as possible. The nasopharyngeal swab test is reliable but it requires time and is expensive, serological tests are faster and cheaper, but less reliable. Serological tests find the presence of IgG (Immunoglobulin G) and a high level of this antibody in blood means that the person is or has been affected by COVID-19.

This work reports the results of the analysis of two serological tests, discussing the setting of the thresholds to declare a positive result.

2 Method

A group of 879 people was subjected to nasopharyngeal swab test and two serological tests (Test 1 and Test 2 in the following), recording the amount of IgG; 17 cases were removed from the dataset due to an uncertain swab test result.

Swab test result was considered correct, and ROC curve (sensitivity versus false alarm, see Fig. 1) was measured for the two serological tests.

For convenience, sensitivity and specificity versus threshold are also plotted in Fig. 2.

The general approach in case of a positive serological test is to check again the person using the nasopharyngeal swab. This makes acceptable a relatively large false positive probability, with the only drawback that healthy people stay for a couple of days at home, maybe in an anxious state.

What cannot be accepted is instead a large false negative probability: in this case nasopharyngeal swab is not tested, and the person can spread the virus to many others. Thus, it is important to have a large sensitivity $P(T_p|D)$ (probability that the test is positive given that the person has the disease), but even more important is the probability $P(D|T_n)$ that the person has the disease given the test is negative. This last probability should be kept as small as possible. The following notation will be used: D means that the patient is really

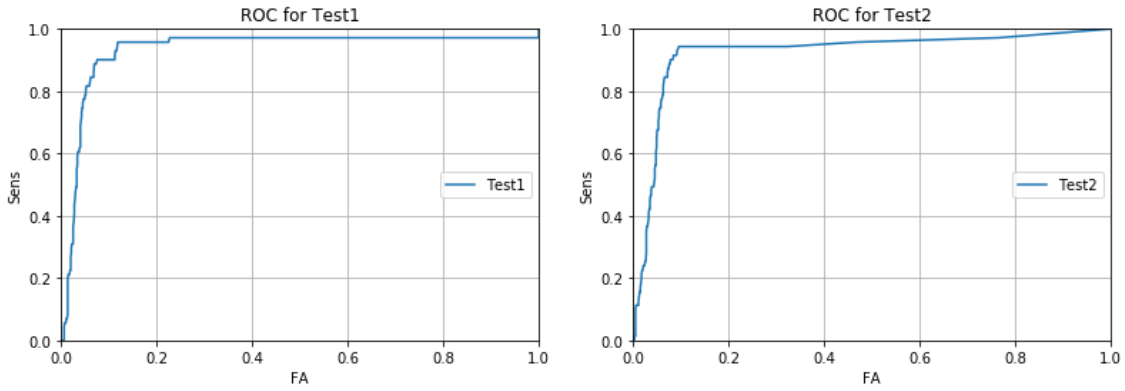

Figure 1: ROC curve for Test 1 (left) and Test 2 (right).

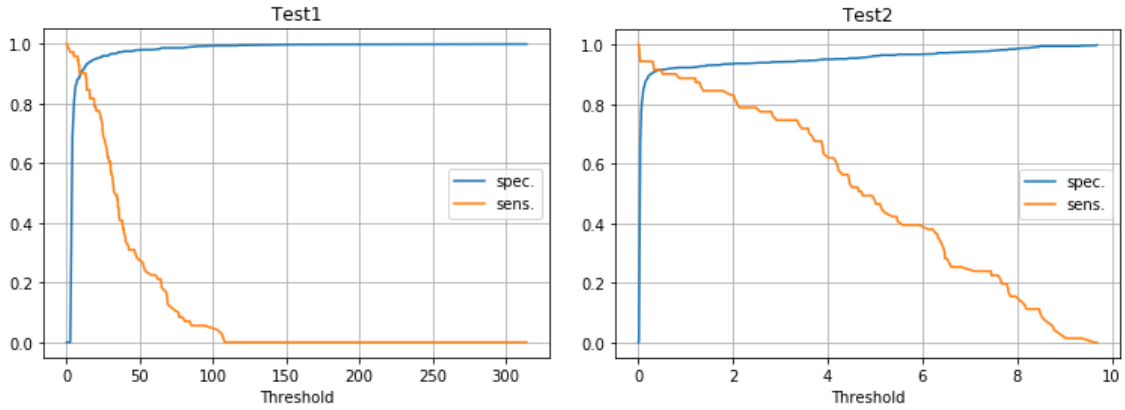

Figure 2: Sensitivity and specificity versus threshold for Test 1 (left) and Test 2 (right).

ill, H means that the patient is healthy, T_p means that the test is positive, T_n means that the test is negative.

Having assumed COVID-19 prevalence ($P(D)$) equal to 2.5 %, Fig. 3 shows:

- The probability $P(D|T_p)$ that the patient is truly ill given that the test is positive.
- The probability $P(D|T_n)$ that the patient is ill given that the test is negative.

versus the threshold. Moreover, the chosen thresholds are also highlighted.

3 Choice of the threshold

For Test 1:

- Sensitivity and specificity are both equal to 0.9 when the threshold is equal to 10, in which case $P(D|T_n) = 3 \cdot 10^{-3}$ and $P(D|T_p) = 0.2$.

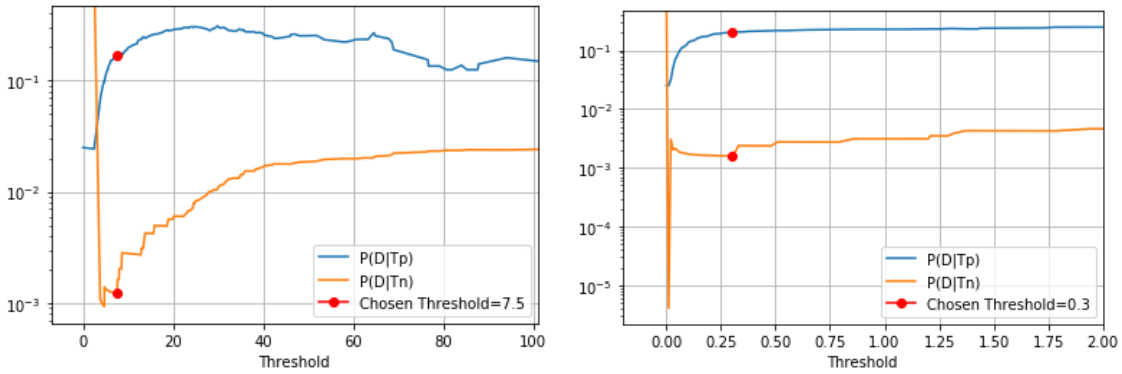

Figure 3: $P(D|T_p)$ and $P(D|T_n)$ versus threshold for Test 1 (left) and Test 2 (right).

- Since sensitivity $P(T_p|D)$ equal to 0.9 cannot be considered sufficient, it is convenient to decrease the threshold below 10; we suggest a threshold equal to **7.5** for which:
  - Sensitivity $P(T_p|D) = 0.96$, false negative probability $P(T_n|D) = 0.04$.
  - Specificity $P(T_n|H) = 0.88$, false positive probability $P(T_p|H) = 0.12$.
  - $P(D|T_n) = 1 - P(H|T_n) = 1.2 \cdot 10^{-3}$, $P(H|T_n) = \frac{P(T_n|H)P(H)}{P(T_n|D)P(D)+P(T_n|H)P(H)} = 0.9988$[1]
  - $P(D|T_p) = \frac{P(T_p|D)P(D)}{P(T_p|D)P(D)+P(T_p|H)P(H)} = 0.17$, $P(H|T_p) = 0.83$.

For Test 2:

- Sensitivity and specificity are both equal to 0.91 when the threshold is equal to 0.48, in which case $P(D|T_n) = 2.4 \cdot 10^{-3}$ and $P(D|T_p) = 0.22$.
- Since sensitivity $P(T_p|D)$ equal to 0.91 cannot be considered sufficient, it is convenient to decrease the threshold below 0.48; we suggest a threshold equal to **0.3** for which:
  - Sensitivity $P(T_p|D) = 0.94$, false negative probability $P(T_n|D) = 0.06$.
  - Specificity $P(T_n|H) = 0.90$, false positive probability $P(T_p|H) = 0.10$.
  - $P(D|T_n) = 1.6 \cdot 10^{-3}$, $P(H|T_n) = 0.9984$
  - $P(D|T_p) = 0.20$, $P(H|T_p) = 0.80$.

4 Conclusions

Both analysed tests present a good accuracy. This is claimed by the fact that the Area Under ROC Curve (AUC) is greater than 0.9 for both of them[2]. In particular, for Test 1 AUC is 0.93 and for Test 2 it is 0.92.

[1] P(H)=1 - P(D)

[2] Conclusion made according to scale reported in [1].

Values of antibodies IgG detected by the two tests in the blood of patients stored in the dataset vary between two different ranges. Test 1 ranges between 2.5 and 314, whereas Test 2 ranges between 0 and 9.69. This evident difference can be due to the different test formats as well as to the different antigens chosen by the manufacturers. Because of this, the two chosen thresholds are not two close values.

Nevertheless, by observing results reported in Sec. 3, both tests have good performance in terms of sensitivity and specificity with their respective thresholds set according to targets reported in Sec. 2. Indeed, the two reasonable thresholds were chosen by opting for a higher prevalence of sensitivity with respect to specificity and, in turn, by accepting a high false positive probability.

Furthermore, $P(H|T_p)$ is very high and $P(D|T_p)$ is very low for both tests. This result may seem unsatisfactory, but if the test is positive, then the person undergoes a swab test. After all, at this time of the pandemic, it is more important that people who might be positive do not go around to spread the virus. Therefore, tests should not be used as the sole basis to diagnose COVID-19 but could help healthcare professionals in identifying individuals who may have developed an immune response to SARS-CoV-2. In addition, these test results can aid in determining who may donate a part of their blood called convalescent plasma, which may serve as a possible treatment for those who are seriously ill from COVID-19.

It is also important to recall that probabilities reported in Sec. 3 were computed considering a hypothesized prevalence and not a real one. Moreover, they are based on a relatively small number of tested people, which should be much greater to get very reliable conclusions. Therefore, results cannot be considered as ground truth.

Finally, according to the observed performance characteristics, the serological test that may be mainly used between the two analyzed is Test 1. In fact, Test 1 has slightly better accuracy, it has slightly higher sensitivity and a smaller $P(D|T_n)$.

References

[1] David W. Hosmer and Stanley Lemeshow. *Assessing the Fit of the Model*, chapter 5, pages 177–178. John Wiley & Sons, Ltd, 2000.