

Merit of Test: Perspective of Information Economics*

Andrea Galeotti
LBS

Jakub Steiner
Cerge-Ei and Zurich U

Paolo Surico
LBS

August 24, 2020

Abstract

This article assesses the merit of a test through the lenses of economics, with applications to SARS-CoV-2. This allows us to rank distinct tests and to show that this ranking is not universal; it depends on the pre-test information available to the decision-maker and the losses stemming from incorrect actions. We provide a method to select, from multiple tests with different sensitivity and specificity, the test that helps the decision-maker the most to achieve her objective.

KEYWORDS: Testing, Covid-19, Portfolio of tests

1 Introduction

This article assesses the merit of a test through the lenses of economics, with applications to SARS-CoV-2.¹ We define the merit of a test to be the extent its outcome improves the choice of the decision-maker. This definition allows to rank distinct tests. We show that this ranking is not universal; it depends on the pre-test information available to the decision-maker and the losses stemming from incorrect actions. We provide a method to select, from multiple tests with different sensitivity and specificity, the test that helps the decision-maker the most to achieve her objective.

The method is suitable to describe the clinical as well as the epidemiological merit of a test. For the attending physician, the merit of the test is high if it helps providing optimal care.

*This work was supported by the ERC grant 770652 and Experientia Foundation (Steiner), and ERC grant 724356 (Andrea Galeotti).

¹There is a large literature in economics on the role of information for decision-making. We refer to (1), (2) and (3) for seminal papers on this topic. This approach has been used and extended recently by (4). It has also been incorporated in medical research, see for example (5).

From an epidemiological perspective, a test should identify infected individuals before they go on to infect others. In both cases there is a decision maker that can test to achieve an objective. A portfolio of tests with distinct sensitivity and specificity parameters is useful, both for the physician and the epidemiologist, since one can adjust the choice of the test to their particular circumstance. Returns of such sophistication in the test choice will be large if careful statistical work will allow us to reliably assess the pre-test infection probabilities and when the employed tests will be carefully validated.

2 Framework

An individual is or is not infected by SARS-CoV-2; that is her state is C or N . If the decision-maker (DM) knew this state, she would choose action a_C when the individual is infected and action a_N when the individual is not. When the DM does not know the true state of the individual, she could choose the wrong action thereby incurring a loss. In particular, the loss is ℓ_C when the individual is infected but DM chooses action a_n ; analogously, the loss is ℓ_N when the individual is not infected but DM chooses a_C . The values of the two losses are free parameters of the framework and can be adjusted to specific applications.² Under incomplete information, the DM follows four steps: (i) she formulates a pre-test probability that the individual is infected,³ (ii) applies a test to the individual, (iii) uses the test result and the Bayes rule to formulate a post-test probability that the individual is infected, (iv) selects the action that minimises the expected loss. This motivates the following definition.

Definition 1. The merit of a test is the difference between the expected loss that the DM incurs if she acts solely based on the pre-test information and the expected loss that the DM incurs if she acts based on the post-test information.

We use Definition 1 to compare performances of distinct tests. We endow the DM with a portfolio of tests that differ in their sensitivities and specificities. The test sensitivity describes the probability that its result is correct when the individual is infected. The test specificity describes the probability that its outcome is correct when the individual does not have SARS-CoV-2. For example, Figure 1 plots the merit of a test with a sensitivity and a specificity of 90%; it also assumes that both types of the decision error lead to a loss of one unit. When the pre-test probability of infection is either low or high, the test never improves

²Losses can be measured in monetary terms, quality-of-life adjusted years or any other units as long as both losses are expressed consistently in same units.

³This probability is evaluated based on all information the DM has before the test is applied; the pre-test probability is informed by the individual's symptoms, contact history, etc. The DM can learn it from past data by setting it equal to the share of people similar to the considered individual who have had SARS-CoV-2.

the choice of the decision maker, and so it has no merit. This is because for an extreme enough pre-test probability, the result of imprecise test does not affect the DM’s choice. For the intermediate pre-test probabilities the outcome of the test informs the decision, and thus the test has a positive merit. The merit is maximal when the DM, in absence of the test, cannot do better than deciding by a coin flip (e.g. 1/2 pre-test probability for equal losses). In the formal analysis below we derive the merit of a test for arbitrary specificity, sensitivity and pre-test probability. In what follows, we develop two applications.

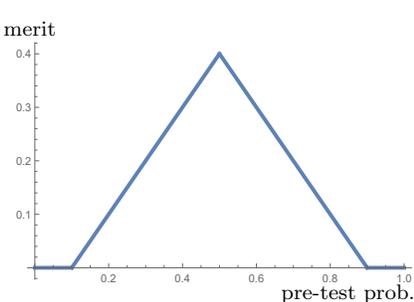


Figure 1: A test with 90% sensitivity and specificity

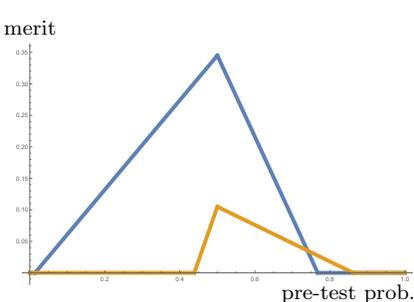


Figure 2: PCR test (blue) vs. CT diagnostic (yellow)

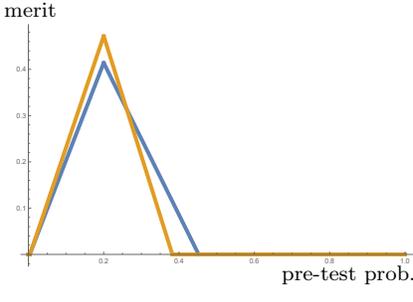


Figure 3: PCR NP swab (blue) vs. PCR OP swab (yellow)

Example 1. PCR or CT. (6) compare diagnostic accuracy of the PCR testing with that of radiologic CT-based diagnostics. Based on this study we set the sensitivity and specificity of the PCR test to be 70% and 99%, respectively. The CT-based diagnostics is highly sensitive but it has low specificity due to many false-positive errors; we set its sensitivity to 96% and its specificity to 25%. In this example, we assume that both types errors lead to a same loss of one unit. Figure 2 plots the merit of each of the two tests for different pre-test probabilities. Because of its high rate of the false-positive errors, CT-based diagnostics is dominated by the PCR testing for most of the values of the pre-test probability P . However, when the pre-test probability is between 0.74 and 0.86, CT-based diagnostics dominates because the high false-positive error rate of the CT diagnostics is less detrimental than the intermediate false-negative error rate of the PCR testing. If the health professional can choose only one of these two diagnostic approaches, perhaps due to the time constraints, then CT scan is optimal for this group of patients. If both diagnostic procedures are done and they return contradictory results, then under these circumstances, it is optimal to follow the CT result.

Example 2. PCR analysis: oropharyngeal (OP) vs. nasopharyngeal (NP) swabs. (7) compare accuracy of the PCR analysis of the oropharyngeal (OP) and nasopharyngeal (NP) swabs. Based on this study, we assume a 70% sensitivity of NP, 60% sensitivity of OP, and a high specificity of 99% for both. Without further considerations, the NP method has unambiguously higher merit. In practice, however, the NP test is more unpleasant for the

patient. When the test is being done for an epidemiological purpose and only asymptomatic individuals are targeted for the test, this could lead to a lower compliance when a NP method is used. For concreteness, we assume that only 3 out of 4 targeted individuals agree to be tested with the NP swab, whereas the OP swab has full compliance. In this epidemiological application, we assume that the loss stemming from the false-negative error is four times bigger than the loss from the false-positive error. Figure 3 plots the merit of the two tests where the test merit for NP swab is scaled by factor $\frac{3}{4}$. For low pre-test probabilities the OP swab method dominates since the minor increase in the false-positive rate is dominated by the benefit of the larger compliance.

3 Conclusion

This article provides a methodology that ranks available diagnostic tests as a function of their sensitivities, specificities, the losses of both types of the decision error, and of the pre-test probability of infection.

Additional information: formal analysis

The *state* of the individual is $\theta \in \{N, C\}$ where N stands for not having SARS-CoV-2, and C for the infection. The decision-maker (DM) chooses between two available actions, a_N and a_C , where the first is optimal when $\theta = N$, and the latter is optimal when $\theta = C$. If the DM does not choose the optimal action, then she suffers a loss ℓ_N or ℓ_C for the state of the individual N or C , respectively. No loss is incurred if the action is chosen optimally.

The DM does not know the individual's state. Let $Q \in [0, 1]$ denote the probability that the DM attaches to the individual having SARS-CoV-2 at the point of decision (and let $1 - Q$ be the probability the DM attaches to the individual not having SARS-CoV-2). As explained in the main text the DM, who chooses whichever action leads to a smaller expected loss, incurs the expected loss given her belief Q equal to

$$L(Q) = \min \{Q\ell_C, (1 - Q)\ell_N\}.$$

Let $P \in [0, 1]$ be the pre-test probability that the DM attaches to the individual having SARS-CoV-2 before she tests her. In absence of the test, the DM cannot learn further information, and her expected loss is therefore $L(P)$.

We endow the DM with a test with sensitivity $Se \in [0, 1]$ and specificity $Sp \in [0, 1]$. The DM applies the test to the individual, observes its result, and updates her belief about the

individual's SARS-CoV-2 status according to the Bayes law. If the DM observes the positive test result, she will assign positive post-test probability

$$\bar{Q} = \frac{P\text{Se}}{P\text{Se} + (1 - P)(1 - \text{Sp})}$$

to the individual having SARS-CoV-2. Similarly, if the DM observes the negative test result, she assigns negative post-test probability

$$\underline{Q} = \frac{P(1 - \text{Se})}{P(1 - \text{Se}) + (1 - P)\text{Sp}}$$

to the individual having SARS-CoV-2.

The imprecise test returns positive result with probability $P\text{Se} + (1 - P)(1 - \text{Sp})$ which accounts for both the true positives and for the false-positives. Similarly, the test returns negative result with probability $P(1 - \text{Se}) + (1 - P)\text{Sp}$. Overall, the expected loss of the DM who has access to the test is

$$(P\text{Se} + (1 - P)(1 - \text{Sp}))L(\bar{Q}) + (P(1 - \text{Se}) + (1 - P)\text{Sp})L(\underline{Q}). \quad (1)$$

The *merit of the test*,

$$V(P) = L(P) - ((P\text{Se} + (1 - P)(1 - \text{Sp}))L(\bar{Q}) + (P(1 - \text{Se}) + (1 - P)\text{Sp})L(\underline{Q})),$$

is the difference between the expected loss $L(P)$ that the DM would have incurred in absence of the test, and the expected loss in equation (1) obtained with testing.⁴

References

- [1] Aumann RJ, Maschler M, Stearns RE. Repeated games with incomplete information. MIT press; 1995.
- [2] Kamenica E, Gentzkow M. Bayesian persuasion. American Economic Review. 2011;101(6):2590–2615.
- [3] Radner R, Stiglitz J. A Nonconcavity in the Value of Information. Bayesian models in economic theory. 1984;5:33–52.

⁴This definition of the test merit is formulated for known model parameters (the two loss values, sensitivity, specificity, and the prior). If values of these parameters are not known but are confined to intervals, then the above approach can be used to deliver an interval of the plausible test merits.

- [4] Ely J, Galeotti A, Steiner J. Optimal Test Allocation. Mimeo; 2020.
- [5] Genders TS, Ferket BS, Hunink MM. The quantitative science of evaluating imaging evidence. *JACC: Cardiovascular Imaging*. 2017;10(3):264–275.
- [6] Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology*. 2020:200642.
- [7] Carver C, Jones N. Comparative accuracy of oropharyngeal and nasopharyngeal swabs for diagnosis of COVID-19. Centre for Evidence-Based Medicine, Nuffield Department of Primary Care Health Sciences, University of Oxford. 2020.