The risk of over-diagnosis in serological testing. Implications for communications strategies

Il rischio di sovradiagnosi nei test sierologici. Implicazioni per le strategie di comunicazione

Annibale Biggeri,[1] Silvia Forni,[2] Mario Braga[2]

[1] Department of Statistics, Computer Science, Applications G. Parenti, University of Florence, Florence (Italy)
[2] Regional Health Agency of Tuscany, Florence (Italy)

Corresponding author: Silvia Forni; silvia.forni@ars.toscana.it

WHAT IS ALREADY KNOWN

- **Accuracy of serological tests in the context of COVID-19 pandemic is limited.**

WHAT THIS STUDY ADDS

- **Empirical importance of tests depends on the context under study, in particular on expected prevalence of infection in the study population.**
- **Estimates of uncertainty inherent these tests must be considered reporting confidence intervals of sensitivity and specificity.**
- **This uncertainty must be taken into account when data on serological surveys are considered both by individuals and policy decision makers.**

ABSTRACT

BACKGROUND: since the beginning of the COVID-19 pandemic, the importance of developing a serological test has emerged and a debate on test accuracy and reliability become an issue widely discussed in the media. The importance of communication during this pandemic has been strongly underlined by public health experts, epidemiologists, media expert, psychologists, sociologists. In the case of serological tests, there are several aspects that have to be considered: why we perform the test, what population is tested, which are the parameters conditioning the results and their interpretation.
OBJECTIVES: to show how to quantify the uncertainty related to the validity of the serological test with respect to its predictive value and in particular the positive predictive value.
METHODS: the evaluation of a qualitative diagnostic test includes four distinct assessments: accuracy, empirical evidence, practical importance, and prevalence of the pathology. Accuracy is measured by the sensitivity and specificity of the test; empirical evidence is quantified by the likelihood ratio, respectively for a positive and negative test result; the practical importance of the result of a diagnostic test is assessed by the positive or negative predictive value. Prevalence of COVID-19 is substantial uncertainty and it is possible to estimate the apparent prevalence starting from the results obtained with a diagnostic test.
RESULTS: at the moment, the knowledge about the accuracy of serological tests is limited and little attention is paid to confidence interval on point estimates. In terms of practical importance of testing at individual level, while negative predictive values are high whatever the level of sensitivity of the test, the interpretation of a positive results is very cumbersome. Positive predictive values above 90% can be reached only by tests with specificity above 99% at the expected prevalence rate of 5%. There is a linear relationship between apparent – testing positive – prevalence and real prevalence. The apparent prevalence in the context of serological test for COVID-19 is always larger than real prevalence. The level of specificity is crucial.
CONCLUSIONS: the main applications of the serological test in the epidemic contest are: to study the seroprevalence of the virus antibodies in the general population; to screen the healthcare workers for the early identification of contagious subjects' health care settings and to screen the general population in order to identify new incident cases. In the first two cases, seroprevalence study and screening of a high-risk population, the consequences of the uncertainty associated to the statistics are already accounted for in the first situation, or are overcome by repeating the screening on the healthcare workers, and using the molecular test to verify the presence of the virus in those tested positive. The case of screening of general population is more complex and of major interest for the implication it may have on individual behaviours and on the implementation of public health interventions by the political decision makers. A positive result has, per se, no practical value for individuals since the probability of being really infected by the virus is low. The uncertainty associated with the different estimates (sensitivity, specificity and disease prevalence) play a double role: it is a key factor in defining the informative content of the test result and it might guide the individual actions and the public policy decisions.

Keywords: serological tests, uncertainty, communication, SARS-CoV-2 prevalence

RIASSUNTO

INTRODUZIONE: la pandemia di COVID-19 ha, fin dall'inizio, fatto emergere l'importanza di avere a disposizione test diagnostici sierologici accurati e affidabili. Quest'ultimo aspetto è diventato oggetto di studio a livello globale e ha portato a un acceso dibattito anche sui media italiani. Un ulteriore elemento di attenzione è stato l'aspetto comunicativo, le cui caratteristiche di complessità e criticità sono state sottolineate da esperti di sanità pubblica, epidemiologici, giornalisti, psicologici, sociologici. In particolare, in questo specifico contesto, gli aspetti da considerare sono molteplici: il motivo dell'esecuzione del test, quale è la popolazione sottoposta a test, quali aspetti condizionano il risultato e come interpretarlo.
OBIETTIVI: fornire strumenti per la quantificazione dell'incertezza, particolarmente in riferimento al valore predittivo dei test sierologici.
METODI: la valutazione di un test diagnostico riguarda quattro diversi aspetti: accuratezza, evidenza empirica, importanza nella pratica e prevalenza della patologia. L'accuratezza del test è data dalla sua sensibilità e specificità; l'evidenza empirica viene misurata dal *likelihood ratio*, rispettivamente, per un risultato positivo e negativo del test; l'importanza pratica è quantificata in termini di valore predittivo positivo e ne-

gativo del test. L'ultimo elemento, la prevalenza dell'infezione, in caso di mancanza di una stima affidabile, può essere ricavata partendo dai risultati osservati del test diagnostico.
RISULTATI: al momento, le stime circa l'accuratezza dei test sierologici sono estremamente limitate e, nei pochi studi disponibili, viene posta poca attenzione agli intervalli di confidenza delle stesse. In termini di rilevanza pratica per il singolo cittadino, questo si traduce in valori predittivi negativi sostanzialmente elevati, mentre l'interpretazione di un risultato positivo è alquanto controversa. Valori predittivi positivo sopra al 90% possono essere raggiunti solo con test la cui specificità supera il 99% e la prevalenza attesa si attesa oltre il 5%. Vi è una relazione lineare tra la prevalenza apparente, come misurata dal test sierologico, e quella reale, con quella apparente sempre maggiore. In queste condizioni, diventa cruciale disporre di una buona stima della specificità del test.
CONCLUSIONI: le principali applicazioni dei test sierologici nel contesto epidemico sono state: lo studio di sieroprevalenza degli anticorpi al virus nella popolazione; lo screening di popolazioni a rischio (per esempio, i professionisti della sanità) per identificare precocemente i soggetti contagiosi; lo screening della popolazione generale per identificare nuovi casi. Per quanto riguarda gli studi epidemiologici di sieroprevalenza, le conseguenze dell'incertezza legata al test sono incorporate nelle stime in fase di analisi dei risultati; nel caso delle attività di screening su popolazioni a rischio, l'incertezza viene minimizzata ripetendo i test sierologici e/o sottoponendo i soggetti positivi al test molecolare da tampone. Il caso dello screening della popolazione generale è più complesso, anche per le ricadute che l'esito del test può avere sui comportamenti degli individui e sulla promozione di interventi di politiche di sanità pubblica. Infatti, un risultato positivo al test sierologico non ha di per sé alcun valore pratico per i singoli individui, dato che la probabilità di essere davvero infetto è molto bassa. Per questo motivo, l'incertezza delle stime di accuratezza dei diversi test deve essere considerata perché la sua entità può avere ricadute sia sul contenuto informativo del test che sulle azioni individuali e collettive che ne conseguono.

Parole chiave: test sierologici, incertezza, comunicazione, prevalenza di SARS-CoV-2

INTRODUCTION

Since the beginning of the COVID-19 pandemic, the importance of developing a serological test has emerged in order to identify people who had contracted the infection and developed specific immunity. The debate on test accuracy and reliability become an issue widely discussed in the media.[1] Serological tests for this new virus were developed during the initial phase of the first epidemic wave by different research institutions. Nowadays 200 scientific articles on this subject are available in the WHO global research database.[2]

At the moment, the COVID-19 infection in Italy is in the final epidemic phase with few new cases, mostly asymptomatic or with very mild symptoms and a low viral load.[3] Under these circumstances, it is essential to have a serological test that promptly evaluate the immune status of people, in a short time and at a low cost in order to immediately identify potential new cases. Furthermore, for reasons related to public health interventions (strengthening or not the measures of social distancing), it seems useful to assess the spread of the virus in the general population.

The declared quality of the commercial tests has to be verified in real conditions and it is consequently necessary to evaluate the validity of these tests (sensitivity and specificity) and their relevance in practice (positive and negative predictive value). This latter aspect reflects the importance and the usefulness of the diagnostic results for the subjects who took the test. It is a typical example of unknown a priori information which scientific research addressed in the first phase of the pandemic.

Duca[4] underlines the fundamental importance of test validation in order to decide which one should be routinely used and for what purposes to use it. One of the key aspects is the number of infected people on which the test will be performed and evaluated, since this element is crucial to correctly interpret the test sensitivity and the meaning of the positive predictive value.

Given a defined level of sensitivity, a test used in population with different disease prevalence can result in positive predictive values of limited or null meaning. The author stresses the need for considering this aspect when comparing commercial serological tests and suggests to provide confidence intervals together with the point sensitivity estimates.

A recent systematic review on studies analysing serological test validity has been published by Lisboa Bastos et al.[5] A high risk of selection bias was measured in 98% of the 49 studies included in the review. Sensitivity and specificity of different tests have been estimated. Sensitivity for ELISA tests resulted in a 95% confidence interval of 76%-91%, LFIA 49%-79%, chemoluminescence CLIA 46%-100%. For specificity, the 95% confidence interval values of 93-99% (ELISA), 94%-98% (LFIA) are reported. For CLIA test the 96% confidence interval has been reported separately for the two immunoglobulins measured (IgG and IgM). If we limit ourselves to IgG, the specificity reported is very good for ELISA (97%-100%) and LFIA (96%-99%), less for CLIA (63%-100%). The authors reported an important heterogeneity in predictive values when the prevalence of the infection is between 5% to 20%.

However, these scenarios of high prevalence appear unrealistic compared to the findings of the first national study of serological prevalence[6] and can be, at most, referred to the peak epidemic phase in few specific areas (i.e., Bergamo) but are certainly inapplicable to the present situation. In Spain, a sample of 61,075 subjects representative of general population has been selected and tested: the result suggests a prevalence of positivity ranging from 4.6 (CLIA) to 5% (LFIA).

In Italy, the Ministry of Health on 25 May, with the Italian National Institute of Statistics (Istat) and Red Cross, launched a survey on a sample of 150,000 citizens aimed at estimating the serum prevalence of SARS-CoV-2 virus infection in the Italian population.[7] The aim of this investigation reported in the study protocol is to determine the proportion of population who developed an antibody response to antigenic determinants of SARS-CoV-2. A tender has been performed to select a proper test for serum IgG, whose constraints were to have a sensitivity of not less than 90% and a specificity of not less than 95% and based on ELISA or LFIA technology. The tender has been won by Abbott with the Architect system. In a recent study,[8] as reported by the authors: "the sensitivity was 72% vs IFA and 66.7% vs a realtime PCR, the specificity was 100%".
In order to ensure adequate adherence to the survey, the Ministry of Health launched an information campaign, with massive advertising through the media.[9] In particular, an information booklet has been prepared where the aim of the investigation is described and indications are given on how the individual participant must interpret the result of the serological test. The aim declared is to draw a precise picture of the spread of the virus in all Italian regions. The information leaflet suggests that if an individual results positive to a serological test this implies that she/he has been in contact with the Coronavirus.
The preliminary results of this survey published at the beginning of august,[10] which involved 64,660 individuals (less than half of the planned sample size), shows that there is a wide geographical heterogeneity and a relevant North-South gradient. The Lombardy region is confirmed as the one with the highest number of positive cases (7.5%), with local peak close to 25% (Bergamo), while the positive rate in South of Italy was below 1%. The positive rate of SARS-CoV-2 in Italy was estimated to be 2.5%.
The most affected category was represented by the Healthcare workers and the estimated proportion of asymptomatic patients was high (27.3%).[10]
The importance of communication during this pandemic has been strongly underlined by public health experts, epidemiologists, media expert, psychologists, sociologists, etc. In the case of serological tests, there are several aspects that have to be considered: why we perform the test, what population is tested, which are the parameters conditioning the results and their interpretation.
In this work, we will show how to quantify the uncertainty related to the validity of the serological test with respect to its predictive value and in particular the positive predictive value. Moreover, we will discuss the uncertainty with respect to the measure of the prevalence of subjects with antibodies. In fact, all studies currently report results in term of prevalence of test positives, but there are no results on the prevalence adjusted for the imprecision of the diagnostic tests.

METHODS

The evaluation of a diagnostic test is a classic topic of medical statistics and clinical epidemiology. Limiting our attention to the case of qualitative tests that return a result in terms of positive/negative, there are four distinct assessments to be made: accuracy, empirical evidence, practical importance and prevalence of the pathology.

ACCURACY

Accuracy is measured by the sensitivity and specificity of the test. Sensitivity is the probability of being positive on the test being ill while specificity is the probability of being negative on the test being healthy. These measures quantify the technological quality of the diagnostic test, which cannot be improved except by changing the type of test (in the present case ELISA rather than LFIA or CLIA). To estimate these probabilities, it is necessary to have a sample of subjects whose disease or non-disease status is known and to submit them to the diagnostic test whose quality we want to measure. The main distortions in this type of study are related to the selection of the sample of subjects and the misclassification of the disease status. The sample must be representative of the test candidate population: a common mistake that leads to an overestimation of the accuracy of the test is to select clearly ill and healthy subjects. Furthermore, post hoc construction of the sample in a differential way for sick and healthy subjects is potentially a source of distortion. In fact, the administration of the test must be blind with respect to the state of the subject. Differential misclassification is always possible when a gold standard is available on which to evaluate the disease status but the evaluation of cases differs from the evaluation of non-cases (as when the sample is built post hoc). For instance, a sample of non-cases is reconstructed starting from blood samples stored in blood banks in the pre-COVID-19 period as long as they meet certain inclusion/exclusion criteria, while cases are a sample of COVID-19 swabs also these with particular inclusion/exclusion criteria.
Having made these design considerations, there is always the need to quantify the uncertainty of the estimate. Being proportions, the correct confidence interval should satisfy the following inequality:

$$P\left\{\left(p - z_{1-\alpha/2}\sqrt{\frac{\pi\,(1-\pi)}{n}}\right) \leq \pi \leq \left(p + z_{1-\alpha/2}\sqrt{\frac{\pi\,(1-\pi)}{n}}\right)\right\}$$

whose solution requires the roots of the second-degree equation (method also known as Wilson). So-called exact ranges are not to be used.[11]
Without this clarification, the case in which the point estimate is at the extreme of the possible values (zero or 100%) remains to be considered. In this case, a "one-tailed" confidence interval is appropriate, that is, spending the probability of first-type error α only for the calculation of one end of the confidence interval. In this situation, the exact interval (Clopper-Pearson) is a good solution.[12]

EMPIRICAL EVIDENCE

Empirical evidence is quantified by the likelihood ratio, respectively for a positive and negative test result. In case of positive test, it is given by the relationship between sensitivity and the complement to one of the specificity (i.e. ratio between the probability of being positive since I am sick and the probability of being positive since I am not):

$$LR_{+} = \frac{P(+|M)}{P(+|S)} = \frac{P(+|M)}{1 - P(-|S)}$$

In case of negative test by the relationship between specificity and the complement to one of sensitivity (ie relationship between the probability of being negative since I am healthy and the probability of being negative since I am not healthy):

$$LR_{-} = \frac{P(-|S)}{P(-|M)} = \frac{P(-|S)}{1 - P(+|M)}$$

These can be particularly useful if the test provides quantitative results to establish areas of values for almost certain results or for inconclusive results. In the case of the tests for the diagnosis of COVID-19 we have only positive/negative qualitative values and therefore the use of the likelihood relationship is of less interest.

PRACTICAL IMPORTANCE

The practical importance of the result of a diagnostic test is assessed by the positive predictive value (in the case of a positive test result) or by the negative predictive value (in the case of a negative test). A diagnostic test is clinically important if the doctor's diagnostic-therapeutic strategy changes following the result. In more formal words, if the probability of being sick (or healthy) following the test result (positive/negative) is substantially different from the probability that the doctor assigned to the patient before performing the test. Not only substantially different but absolutely such as to be sufficient to make different diagnostic-therapeutic decisions. Obviously, an inconclusive test has no practical importance. But even a test with a high likelihood ratio may not have any practical importance, because in absolute terms the probability of being ill in the positive can remain low. This is because this probability (the positive and negative predictive values) depends on the prevalence of the disease. It is Bayes' well-known formula:

$$P(M|+) = \frac{P(+|M)P(M)}{P(+)}$$

for the positive predictive value, and

$$P(S|-) = \frac{P(-|S)P(S)}{P(-)}$$

for the negative predictive value.

High values for the VPP place the diagnosis, high values of the VPN reassure the absence of the disease.

If the prevalence of the disease is low, we could have predictive values of a practical importance only in the case of very high sensitivity or specificity. It is not obvious to understand but, as we will illustrate below, exploding the denominator of Bayes' formula as a mixture of probabilities, we note how specificity plays a critical role with respect to the positive predictive value and sensitivity with respect to the negative predictive value:

$$P(M|+) = \frac{P(+|M)P(M)}{P(+)} = \frac{P(+|M)P(M)}{\{P(+|M)P(M) + P(+|S)P(S)\}}$$

and, therefore, the positive predictive value depends on how many false positives $P(+|S)P(S)$ there are where $P(+|S)$ is the complement to one of specificity, and

$$P(S|-) = \frac{P(-|S)P(S)}{P(-)} = \frac{P(-|S)P(S)}{\{P(-|S)P(S) + P(-|M)P(M)\}}$$

and, therefore, the negative predictive value depends on how many false negatives $P(-|M)P(M)$ where $P(-|M)$ is where the complement to one of the sensitivity.

In order to maximize specificity and positive predictive value a usual strategy would be test repetition.[13] Given that posterior odds are equal to likelihood ratio multiplied prior odds the serial testing strategy will give:

$$Odds_{prior} = \frac{P(M)}{1 - P(M)}$$

$$Odds^{1}_{post} = LR_{+} \times Odds_{prior}$$

$$Odds^{2}_{post} = LR_{+} \times Odds^{1}_{post} = (LR_{+})^{2} \times Odds_{prior}$$

$$P(M|+) = \frac{Odds^{2}_{post}}{1 + Odds^{2}_{post}}$$

This will result in a posteriori positive predictive value bigger than using a single test.

PREVALENCE

Usually the prevalence of the disease is a priori information known by doctors, or in any case for which there is a priori information (in Bayes' formula the prevalence is indicated as a priori probability). There are cases such as in the context of the COVID pandemic19 where there is substantial uncertainty about the prevalence of the disease.

In this case, one might ask how to estimate the prevalence starting from the results obtained with a diagnostic test (imperfect by definition!). Usually the so-called "apparent" prevalence, i.e., the percentage of test positive, are reported as results in the literature and by the mass media.

The relationship between true and apparent prevalence is

expressed by the following formula (Diggle) (valid for the expected value of the apparent prevalence, not for a sample value):

$$E(PA) = P(+|M) \times P(M) + P(+|S) \times P(S)$$

where it is clear the role of false positives. In the case of COVID-19 we expect an apparent prevalence always greater than the true prevalence.
For prevalence:

$$\hat{P}(M) = \frac{PA - P(+|S)}{\{P(+|M) + P(-|S) - 1\}}$$

and for the confidence interval, where (a, b) are the extremes of the confidence interval of the apparent prevalence:

$$c = max\left\{0, \frac{a - P(+|S)}{\{P(+|M) + P(-|S) - 1\}}\right\}$$

$$d = min\left\{1, \frac{b - P(+|S)}{\{P(+|M) + P(-|S) - 1\}}\right\}$$

These formulas will be applied to the diagnostic tests for COVID-19 in the Italian context. However, it should be clear that, in order to use the apparent prevalence, the key assumption is that the study is unbiased.
When sensitivity and specificity are unknown, different formulas should be used.[14]
The main threat to the validity of the prevalence estimates is the potential bias in the selection of subjects. This distortion is almost certain in all cases based on opportunistic nasopharyngeal swabs where a representative sample of population is subject to serological tests.[6,15]

RESULTS

ACCURACY

At the moment, the knowledge about the accuracy of serological tests is limited. As reported in the review mentioned in the introduction section pooled sensitivity for ELISA tests resulted in a 95% confidence interval of 76-91%, LFIA 49-79%, chemoluminescence CLIA 46-100%. For pooled specificity, the 95% confidence interval values of 93-99% (ELISA), 94-98% (LFIA) are reported. For CLIA test the 95% confidence interval has been reported for IgG as 63-100%.[6] These pooled analyses are based on few thousands of subjects and there is a substantial uncertainty as testified by the width of the confidence intervals.
In the recent study on the Italian serological test selected for the national survey, the authors report 95 positives among 140 COVID-19 patients (table 2).[8] This finding translate in a 95% confidence interval for sensitivity (Wilson score method) of 60%-75%. For specificity, the authors report 0 positive over 37 subjects tested (table 3) for a 95% one-sided exact lower confidence limit of 92%.
At the current level of evidence, point-of care serological tests have lower bound for specificity around 93% and an upper bound for sensitivity around 90% (ELISA), 80% (LFIA) and 75% (Abbott Italian serological test).
Surprisingly, very little attention is paid to confidence intervals and, in the literature, most of the conclusions are based on point estimates.

EMPIRICAL EVIDENCE

This aspect has a marginal interest in our context since we are focusing on qualitative testing. However, using data from the systematic review and the Italian paper on Abbott serological test we notice that likelihood ratios vary depending on several covariates.
In the table 1 we report the likelihood ratios as function of days after symptoms onset. For the positive LR we use the lower bound of specificity and the point and 95% interval estimate of sensitivity (in brackets). It is a worst case scenario. The positive LR values support good evidence of the test in some cases also in the first phase on the infection. For the negative LR we use the upper bound of sensitivity and the point and 95% one-sided confidence interval of specificity – which according to the paper is constant and not function of days after onset. It is a best case scenario. A negative test is informative only after 14 days after symptom onset.
To appraise the empirical evidence quantified by the likelihood ratios the figure 1 reports the posterior odds by positive LR for different prior odds. The posterior odd is dominated by the likelihood ratio only when the prior odd is based on a prevalence greater than 15% - and even in this case we need positive LR in the order of 10.

PRACTICAL IMPORTANCE

This substantial uncertainty of accuracy and the strength of the empirical evidence have a consequence on the predictive values of the test. To illustrate this, we will fix the prevalence to 2.5%, a reasonable level according to the Italian national survey,[10] and plot the Positive Predictive Value – the probability of having the disease testing positive – by specificity and sensitivity.
It is important to notice that the relationship between accuracy and predictive values is not linear. What is relevant here, is to concentrate on critical points in the range of values of specificity and sensitivity included within the confidence interval. Even if a point estimate seems to be reassuring, taking the lower bound of the confidence interval for specificity gives a very poor positive predictive value (fixing sensitivity at 80%) (figure 2).
For sensitivity, fixing specificity to 95%, we found a good performance in term of negative predictive value even with low values of sensitivity (figure 3).
In terms of practical importance of testing at individual level, while negative predictive values are high whatever the level of sensitivity of the test, the interpretation of a posi-

METHOD	IgM				IgG			
TIME POST-ONSET	LR + LOWER BOUND OF SPECIFICITY (95% CI)		LR – UPPER BOUND OF SENSITIVITY (95% ONE-SIDED CI)		LR + LOWER BOUND OF SPECIFICITY (95% CI)		LR – UPPER BOUND OF SENSITIVITY (95% ONE-SIDED CI)	
ELISA								
1-7 days	26.7	(15.6-35.6)	1.5	(1.5)	7.9	(4.2-12.7)	1.6	(1.6)
7-14 days	57.6	(15.9-88.2)	8.4	(8.4)	21.8	(15.4-26.5)	4.8	(4.7)
>14 days	78.4	(54.1_91.9)	12.2	(12.2)	27.4	(25.5-29.7)	9.0	(8.8)
LFIA								
1-7 days	4.2	(2.7-5.2)	1.4	(1.4)	3.35	(1.2-0.4)	1.4	(1.4)
7-14 days	8.6	(5.0-11.6)	3.2	(3.1)	12.5	(6.2-2.5)	4.3	(4.2)
>14 days	11.6	(9.7-13.3)	4.8	(4.7)	19.9	(17.8-4.3)	7.5	(7.3)
CLIA								
1-7 days	3.3	(0.7-5.4)	5.2	(4.5)	1.4	(0.8-1.8)	3.0	(1.9)
7-14 days	5.0	(1.1-6.6)	161.7	(141.7)	2.3	(1.3-2.6)	51.6	(33.2)
>14 days	6.0	(3.4-6.6)	161.7	(141.7)	2.7	(2.3-2.7)		

Ig: immunoglobulin / *immunoglobulina*; **ELISA:** enzyme linked immunosorbent assay / *saggio immuno-assorbente legato a un enzima*; **LFIA:** lateral flow immunoassay / *dosaggio immunologico a flusso laterale*; **CLIA:** chemiluminescent immunoassay / *dosaggio immunologico chemiluminescente*; **LR+:** positive likehood ratio / *rapporto di verosimiglianza positivo*; **LR-:** negative likehood ratio / *rapporto di verosimiglianza negativo*

Table 1. Positive and negative likehood ratios by serology test method and timing of symptoms onset.[5]
Tabella 1. Rapporti di verosimiglianza positivi e negativi dal momento di comparsa dei sintomi, per metodo di test sierologico.[5]

TIME POST-ONSET	LR + LOWER BOUND OF SPECIFICITY (95% CI)		LR – UPPER BOUND OF SENSITIVITY (95% ONE-SIDED CI)	
1-7 days	5.0	(3.0-6.0)	1.9	(1.8)
7-14 days	7.2	(6.2-8.2)	2.9	(2.7)
>14 days	8.5	(7.5-9.4)	4.0	(3.7)

LR+: positive likehood ratio / *rapporto di verosimiglianza positivo;* **LR-:** negative likehood ratio / *rapporto di verosimiglianza negativo*

Table 2: Positive and negative likehood ratios by timing of symptoms onset for Abbott ARCHITECT.[8]
Tabella 2. Rapporti di verosimiglianza positivi e negativi dal momento di comparsa dei sintomi per Abbott ARCHITECT

tive results is very cumbersome. Positive predictive values above 90% can be reached only by tests with specificity above 99% at the expected prevalence rate of 5%. A specificity which cannot be expected by the current available tests. This opens the question of how to communicate this and what are the implications in terms of public health and individual health.

Assuming a serological test with sensitivity of 80% and specificity varying between 90% to 99%, it is possible to calculate the positive predictive values for different levels of disease prevalence (figure 4).

Figure 4 represents the curve of positive predictive values for different levels of COVID-19 prevalence. The range of disease prevalence is coherent with that observed in the different Italian regions.

The observed average probability of being tested positive to IgG, IgM antibodies in Italy is around 2.5%.[10] This probability has a relevant geographical variability between the Italian regions, depending on the intensity of the epidemic experience by that territory. For example, the most severely hit areas of the north regions have a test positive prevalence of 15%, while in many regions of the south of Italy the probability of being test positive is around 1%.

Considering this range of values (1%-15%), the positive predictive value even for a person living in the Lombardy region (the highest test positive prevalence) could be between 58% (specificity 90%) and 93% (specificity 99%). Only a serial testing strategy on positive would result in a bigger posterior positive predictive value. Considering a prior prevalence of 2.5% a sensitivity of 80% and specificity of 99%, repeating the test on positive will increase the positive predictive value from 67% to 99%.

APPARENT PREVALENCE

At population level an estimate of the true prevalence can be obtained from imperfect test with known sensitivity and specificity, as we reported in the methods section. There is a linear relationship between apparent – testing positive – prevalence and real prevalence. The apparent prevalence in the context of serological test for COVID-19 is always larger than real prevalence. The level of specificity is crucial (figure 5).

Sensitivity has a minor paradoxical role. In fact, a less sensitive test will compensate the number of false positives – which inflates the prevalence estimate – by the false negatives (figure 6)

At population level, serological surveys can be useful even in presence of an imperfect test. The problem is in the interpretation of the survey results and their value if projected at the individual level.

DISCUSSION

WHAT IS THE RELEVANCE OF BEING TEST POSITIVE FOR AN INDIVIDUAL? FOR A PHYSICIAN? FOR A COMMUNITY?

A simpler and more understandable way to communicate the concepts described in the previous paragraph is repre-

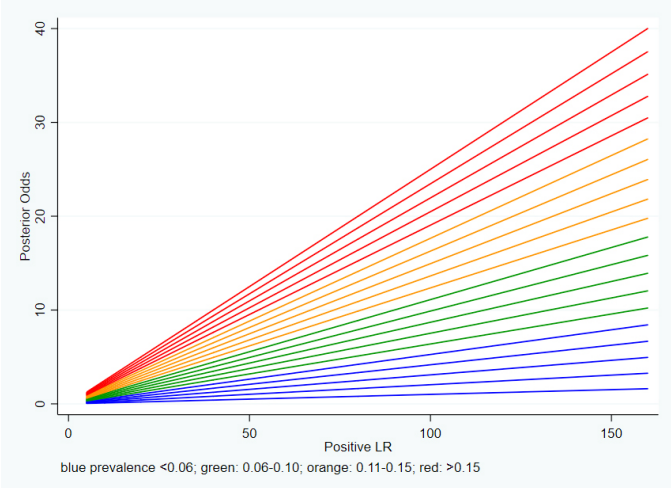

Figure 1. Positive likelihood ratio and Posterior Odds, by prevalence.
Figura 1. Rapporto di verosimiglianza positivo e odds a posteriori, per prevalenza.

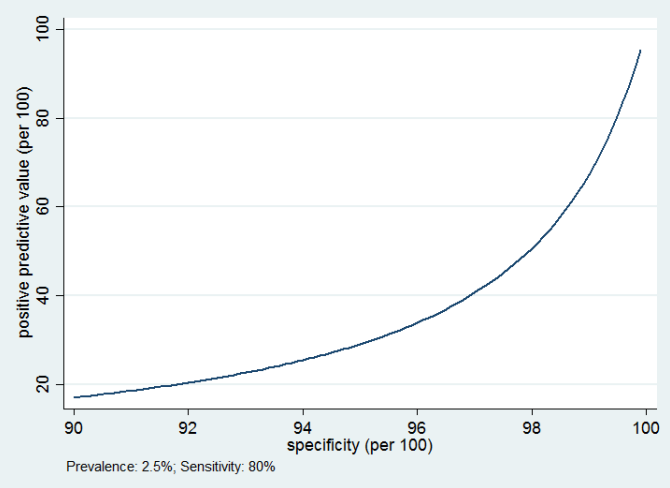

Figure 2. Specificity and Positive Predictive Values, fixing prevalence at 2.5% and sensitivity at 80%
Figura 2. Specificità e valori predittivi positivi, mantenendo fissa la prevalenza a 2,5% e la sensibilità a 80%.

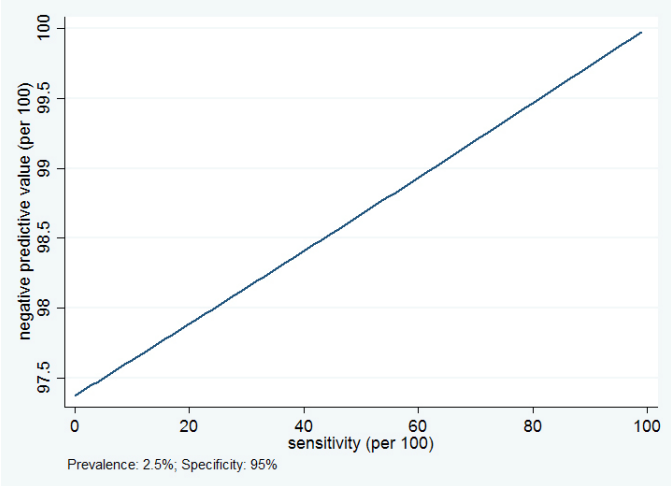

Figure 3. Sensitivity and negative predictive values, fixing prevalence at 2.5% and specificity at 95%.
Figura 3. Specificità e valori predittivi negativi, mantenendo fissa la prevalenza a 2,5% e la sensibilità a 95%.

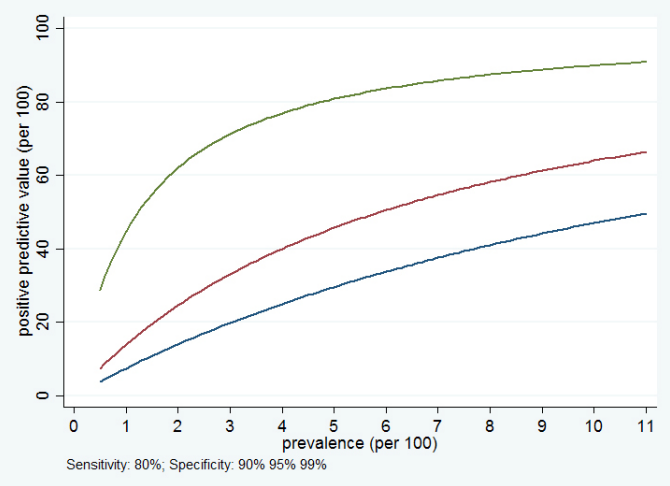

Figure 4. Prevalence and positive predictive values, by different levels of specificity.
Figura 4. Prevalenza e valori predittivi positivi, per diversi livelli di specificità.

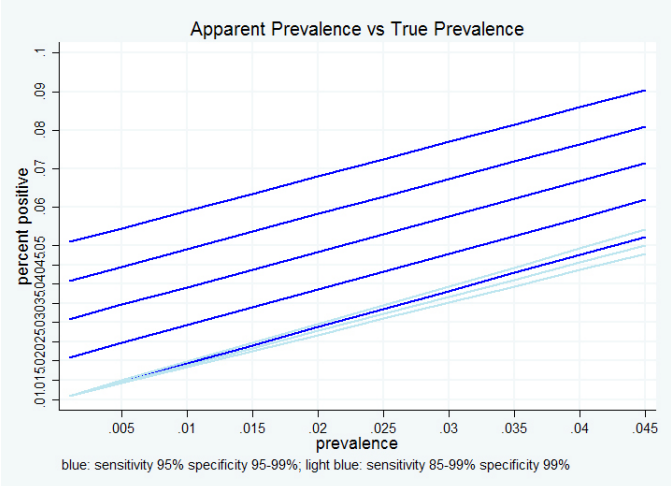

Figure 5. Prevalence and percent positive (apparent prevalence), by different levels of sensitivity and specificity.
Figura 5. Prevalenza e percentuali di casi positivi (prevalenza apparente), per diversi livelli di sensibilità e specificità.

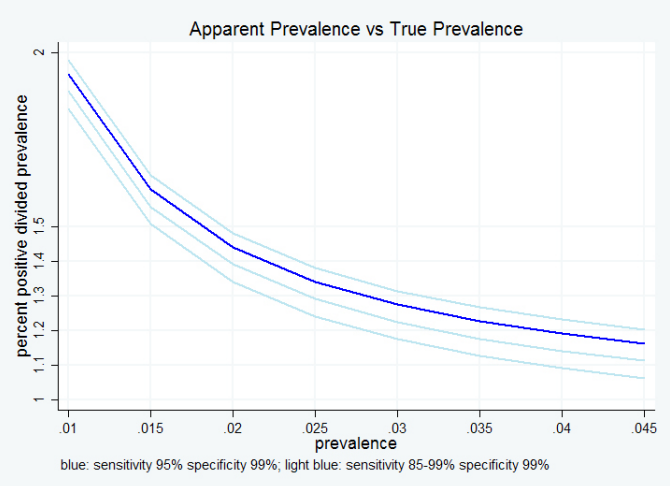

Figure 6. Prevalence and relative bias, by different levels of sensitivity and specificity.
Figura 6. Prevelenza e bias realtiva, per diversi livelli di sensibilità e specificità.

sented in figure 7. In the case of a diagnostic test with 95% sensitivity and 95% specificity, as it was required for the commercial test to be used in the National Survey sponsored by the Italian Ministry of Health, and with disease prevalence assumed to be equal to 2.5%, among the 1000 subjects who undergone the serological test, 73 will result positive, but only 24 of them would have been infected by the virus. On the other side, only one infected person will result test negative.

These concepts are well known inside the scientific community of medical statisticians and epidemiologists who are familiar with these topics and they have different meaning for the single person, for the clinician and for public health decision maker.

From the subject viewpoint, a negative result of the serological test should be interpreted as the absence of antibodies and is consequently an indication that there was no contact with the virus. In this situation, that person should maintain high level of adherence to the well know protection behaviours (physical distance, frequent hand washing, protective facial mask, etc.).

On the contrary, a positive result has, per se, no practical value, since the probability of being really infected by the virus is low. Only if the serological test is followed by a viral RNA test, pointing to the presence of the SARS-CoV-2 virus it would be possible to confirm the infection. But this is impossible because the timing of infection is different from the timing of serological response. This is clearly illustrated by the Italian paper on the Abbott test selected for the national survey. At 14 days from onset of symptoms only 81% of patients were still positive to viral RNA testing.

In general, the presence of antibodies should be interpreted as a real infection only in 24 out of 73 people which tested positive. As shown to maximize positive predicted value a usual strategy would be test repetition.

Analogously, from a clinical point of view a positive serological test has low informative content, particularly in case of no or mild symptoms, with difficulties in defining the stage of the disease and the opportunity to reduce or to remove the public health measures used to contrast the epidemic.

The case of the health workers' community follow different logics. In this case, the hospital staff could benefit from a serological test campaign that can be used to identify individuals with positive results. These people have to stop working, stay in self isolation and get a molecular test in order to verify if they get the infection and be contagious. This strategy is fundamental to stop the contagious chain and to avoid the spread of the infection in a very fragile environment and has to be replicated regularly to control the spread of the infection in these settings.

From a public health point of view, a seroprevalence study provides a real-world estimate of the virus diffusion in the general population. This is particularly useful in those circumstances of widespread virus circulation, such as hospitals or healthcare homes, where there is a high concentration of high risk subjects and where the high disease prevalence will lead to a highly informative positive predictive value.

HOW TO CORRECTLY COMMUNICATE THE RESULTS OF A SEROLOGICAL TEST? WHICH IS THE BEST STRATEGY TO INVITE THE POPULATION TO TAKE PART OF A SURVEY?

Serological tests have been used in different setting and both, the characteristics of the test and the epidemiological profile of the disease, have consequences in terms of meaningfulness of the results, behaviours to be adopted and operational/organizational fallout.

The main applications of the serological test in the epidemic contest are:

- to study the seroprevalence of the virus antibodies in the general population;
- to screen the healthcare workers for the early identification of contagious subjects' health care settings;
- to screen the general population in order to identify new incident cases.

In the first two cases, seroprevalence study and screening of a high-risk population, the consequences of the uncertainty associated to the statistics are already accounted for in the first situation, or are overcome by repeating the screening on the healthcare workers, and using the molecular test to verify the presence of the virus in those tested positive.

Studies on seroprevalence of SARS-CoV-2 antibodies provide estimates on the infectious rate and are useful for monitoring the progression of the epidemic. Particularly on the early phases of the COVID-19 epidemic, the lack of pharyngeal swabs for organizing a screening campaign for those at risk of infection lead to a large underestimation of the number of COVID-19 cases, mostly with mild symptoms. Under these circumstances, seroprevalence surveys are essential to provide a better estimate of the proportion of the general population with antibodies against SARS-CoV-2 and, potentially, immune to a subsequent infection.

The replication of these surveys over time, as suggested by WHO, might be useful to monitor the changes of seroprevalence and to implement public health interventions in advance.

Repeated screening campaign with serological test in hospitals or other healthcare settings have been used to protect healthcare workers and patients through the early identification of subjects that might be infected and contagious. The delay of implementing this intervention at the beginning of the epidemic has caused the diffusion of the infection among the population working and living in these environments with dramatic consequences in terms of number of COVID-19 cases and deaths.

The last case is more complex and of major interest for the implication it may have on individual behaviours and on the implementation of public health interventions by the political decision makers.

In this context, the uncertainty associated with the differ-

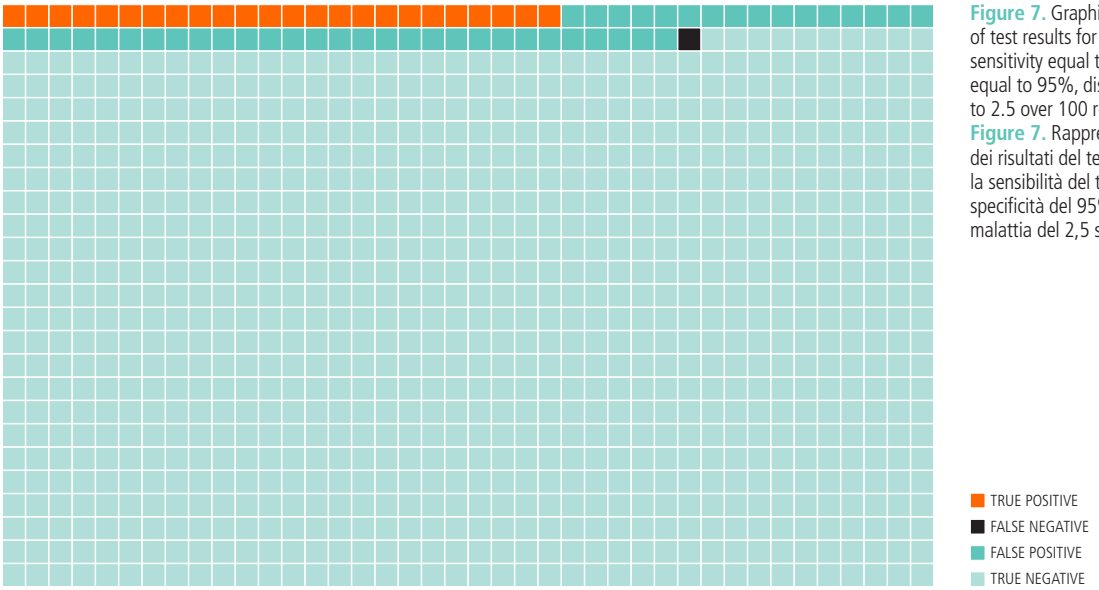

Figure 7. Graphical representation of test results for 1,000 people: test sensitivity equal to 95%, test specificity equal to 95%, disease prevalence equal to 2.5 over 100 residents.
Figure 7. Rappresentazione grafica dei risultati del test per 1.000 persone: la sensibilità del testo è del 95%, la specificità del 95%, la prevalenza di malattia del 2,5 su 100 residenti.

ent estimates (sensitivity, specificity and disease prevalence) play a double role: it is a key factor in defining the informative content of the test result and it might guide the individual actions and the public policy decisions.
The choice of providing a dichotomous information (positive – negative) to the tested individual overshadows the underling uncertainty of this result and give the simple message that there was no infection, which has no relevant implications since these individuals will continue to maintain all the precautionary measures (facial mask, hand washing and social distancing), or that the individuals positive to the test are, or were, infected. In this case, the individual behaviour could change in the assumption that there is no further risk and that the protective measures could be relaxed.
Changes in adherence to the public health measures could jeopardize their efficacy and may lead to the development of new outbreaks.
From the policy decision maker's side, not considering the uncertainty of the test validity will cause a waste of economic resources, because the test is useless, and will divert the attention of the public from the correct adoption of all the non-pharmacological interventions.

Conflict of interest: none declared.

Submission date: 24.07.2020
Accepted on: 04.12.2020

REFERENCES

1. Menietti E. La complicata questione dei test sierologici. Il Post, 2020. Available from: https://www.ilpost.it/2020/04/21/test-sierologici-coronavirus/?fbclid=IwAR07SmYYdg9GyGqpfEAhI82sJdHa419mR9S2ToMDPNj3wWEZn2wrRr_8x3s
2. World Health Organization. Global research on coronavirus disease (COVID-19). Geneva, WHO, 2020. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov (last accessed 20.07.2020).
3. Ministero della salute. Covid-19 – Situazione in Italia. Roma: Ministero della salute; 2020. Available from: http://www.salute.gov.it/portale/nuovocoronavirus/dettaglioContenutiNuovoCoronavirus.jsp?area=nuovoCoronavirus&id=5351&lingua=italiano&menu=vuoto (last accessed 20.07.2020).
4. Duca P. Sensitivity, specificity, predictive values in serological Covid-19 tests. Epidemiol Prev 2020;44(2-3):189-90.
5. Lisboa Bastos M, Tavaziva G, Abidi SK, et al. Diagnostic accuracy of serological tests for covid-19: systematic review and meta-analysis BMJ 2020;370:m2516.
6. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. Lancet 2020;396(10250):535-44.
7. Ministero della salute. COVID-19, indagine di sieroprevalenza. Roma: Ministero della salute; 2020. Available from: http://www.salute.gov.it/portale/nuovocoronavirus/dettaglioNotizieNuovoCoronavirus.jsp?lingua=italiano&id=4805 (last accessed 20.07.2020).
8. Meschi S, Colavita F, Bordi L, et al. Performance evaluation of Abbott ARCHITECT SARS-CoV-2 IgG immunoassay in comparison with indirect immunofluorescence and virus microneutralization test. J Clin Virol 2020;129:104539.
9. Ministero della salute. Opuscolo Informativo, indagine sierologica nuovo Coronavirus. Roma: Ministero della salute-Istat; 2020. Available from: http://www.salute.gov.it/imgs/C_17_opuscoliPoster_470_allegato.pdf
10. Sabbadini LL, Romano MC, Luzi O. Primi risultati dell'indagine di sieroprevalenza sul SARS-CoV-2. Roma: Istat- Ministero della salute, 2020. Available from: https://www.istat.it/it/files/2020/08/ReportPrimiRisultatiIndagineSiero.pdf
11. Clayton D, Hill M. Statistical Models in Epidemiology. Oxford: Oxford University Press; 2013.
12. Bland M. An Introduction to Medical Statistics. 4th Edition Oxford: Oxford University Press; 2015.
13. Fletcher RH, Fletcher SW, Fletcher GS. Clinical epidemiology: the essentials. 5th edition. Philadelphia: Lippincott Williams & Wilkins; 2013.
14. Diggle PJ. Estimating prevalence using an imperfect test. Epidemiology Research International 2011;2011: 608719.
15. Pearce N, Vandenbroucke JP, VanderWeele TJ, Greenland S. Accurate statistics on COVID-19 are essential for policy guidance and decisions. Am J Public Health 2020;110(7):949-51.