



# COVID-19: accurate interpretation of diagnostic tests—a statistical point of view

Takashi Asai<sup>1</sup>

Received: 19 October 2020 / Accepted: 24 October 2020 / Published online: 11 December 2020  
© Japanese Society of Anesthesiologists 2020

Severe acute respiratory syndrome-corona virus-2 (SARS-CoV-2), which causes coronavirus disease 2019 (COVID-19), is highly contagious, and is transmitted through human respiratory droplets and aerosols, and direct contact [1–3]. Among healthcare workers, we anesthesiologists, intensivists and emergency doctors, are particularly at increased risk of infection, as we frequently need to carry out procedures, which would increase the risk of spreading viral droplets and aerosol from the patient’s airway (so-called “aerosol-generating procedures”): tracheal intubation and extubation, mask ventilation, tracheostomy, high-flow oxygen delivery, bronchoscopy, and removal of oropharyngeal or tracheal secretion by suction [3].

Real-time reverse transcription-polymerase chain reaction (RT-PCR) targeting select genes of the SARS-CoV-2 RNA is the main diagnostic test for COVID-19 [4]. It is necessary to interpret accurately the results of a diagnostic test, to take all the measures to prevent spread of infection and to take care of infected people. Nevertheless, there are uncertainties in the validity of the diagnostic test, and the incidence of false positives and false negatives can be high [5, 6]. In addition, the results are frequently misinterpreted [7, 8]. Here, I describe accurate interpretations of diagnostic tests, mainly from a statistical point of view.

## Validity of diagnostic tests

The validity of diagnostic test varies considerably between different test kits. Therefore, we first should know the validity of a diagnostic test, which can be determined by sensitivity and specificity.

Sensitivity is the proportion of people with a disease in whom a diagnostic test correctly indicated a positive result [9], as shown below:

$$\begin{aligned} \text{Sensitivity} &= \frac{\text{Number of people with a disease who tested positive}}{\text{Total number of people with a disease}} \\ &\times 100 (\%) \end{aligned}$$

For example, if there are 100 people infected with COVID-19, and a RT-PCR indicated positive in 80 of them, the sensitivity is 80% ((80/100) × 100 = 80 (%)). The remaining 20% is the proportion of people with a disease in whom a diagnostic test falsely indicated a negative result (false negative).

Specificity is the proportion of people without a disease in whom a diagnostic test correctly indicated a negative result, as shown below:

$$\begin{aligned} \text{Specificity} &= \frac{\text{Number of people without a disease who tested negative}}{\text{Total number of people without a disease}} \\ &\times 100 (\%) \end{aligned}$$

For example, if there are 100 people who are not infected with COVID-19, and an RT-PCR indicated negative result in 90 of them, the specificity is 90% ((90/100) × 100 = 90 (%)). The remaining 10% is the proportion of people without a disease in whom a diagnostic test falsely indicated a positive result (false positive).

The test can be regarded useful when both the sensitivity and specificity are high (or when both false-negative rate and false-positive rate are low). Reported sensitivity of currently available RT-PCR kits is in the range of 70–98% [10, 11], and sensitivity 95–99.7%. The World Health Organization (WHO) has indicated that the validity of the test would be ‘acceptable’ when the sensitivity is ≥ 80% and the specificity ≥ 90%, and ‘desirable’ when the sensitivity is ≥ 90% and the specificity ≥ 99% [4].

✉ Takashi Asai  
asaita@dokkyomed.ac.jp

<sup>1</sup> Department of Anesthesiology, Dokkyo Medical University  
Saitama Medical Center, 2-1-50 Minamikoshigaya,  
Koshigaya, Saitama 343-8555, Japan

### Predictive values

Some seem to interpret that, when both the sensitivity and specificity are high, the test results are trustworthy enough for clinical decision-making. From a statistical viewpoint, this interpretation is insufficient.

When we perform a diagnostic test, we want to know what the probability is of the test giving correct diagnosis, whether it is positive or negative. The sensitivity and specificity do *not* give this information.

Sensitivity is calculated based on people who are *known* to have a specific disease, and specificity is based on people who are *known* to be without the disease. We perform a diagnostic test, because we do *not know* whether or not the person having the test has the disease (if we knew, we would not need a diagnostic test!). Therefore, neither sensitivity nor specificity provides the accuracy of the test in a clinical setting. This information can be obtained by calculating the positive predictive value (PPV) and negative predictive value (NPV).

The PPV is the proportion of people with positive test results who are correctly diagnosed (or who are turned out to be truly infected with the disease), whereas the NPV is the proportion of people with negative test results who are correctly diagnosed (or who are truly not infected with the disease):

$$PPV = \frac{\text{Number of people with the disease who have positive test results}}{\text{Number of people with positive test results}} \times 100 (\%)$$

$$NPV = \frac{\text{Number of people without the disease who have negative test results}}{\text{Number of people with negative test results}} \times 100 (\%)$$

For example, suppose that the results of 430 people (who were tested for COVID-19, with a diagnostic test with the sensitivity of 0.7, and specificity of 0.9) are shown in Table 1. Table 1 indicates that 100 of 430 people were confirmed to be truly infected with disease after the testing.

**Table 1** Hypothetical infectious state and test results (sensitivity of the test: 0.7; specificity: 0.9)

Test results	Disease	
	Infected (100)	Not infected (330)
Positive (103)	70	33
Negative (327)	30	297

PPV (70/103) × 100 = 68 (%), NPV (297/327) × 100 = 91 (%)

As the sensitivity is 0.7, the number of patients with the disease who have positive test results is 70 (100 × 0.7 = 70). Remaining 330 people were confirmed to be not infected after the testing. The specificity is 0.9, and thus, the number of people having false-positive results is calculated to be 33 (330 × 0.1 = 33). From this, the PPV is calculated to be 68% ((70/(70 + 33)) × 100 = 68(%)). The NPV is calculated to be 91%. If the same people are tested with a diagnostic test of a better validity (sensitivity of 0.95, and specificity of 0.95) (Table 2), both the PPV and NPV become higher (86 and 98%).

### Influencing factor on predictive values

The PPV and NPV give a direct assessment of the accuracy of the test, but there is another essential aspect of the analysis to consider—the prevalence of disease. The PPV and NPV are strongly influenced by the prevalence of the disease.

The prevalence of COVID-19 is difficult to estimate, partly because the prevalence may be constantly changing, and partly because a considerable number of infected people are asymptomatic. All we know is the infection status of those who have been tested. For example, in Japan (a population of approximately 126 million), the average “current”

number of active cases in September 2020 is 5,000–6,000, and the total number of diagnosed cases during the last 9 months (January–September, 2020) is approximately 83,000. Roughly 10–20% of people with COVID-19 are said to be asymptomatic [12], and thus, simple calculation

**Table 2** Hypothetical infectious state and test results (sensitivity of the test: 0.95; specificity: 0.95)

Test results	Disease	
	Infected (100)	Not infected (330)
Positive (111)	95	16
Negative (319)	5	314

PPV (95/111) × 100 = 86 (%), NPV (314/319) × 100 = 98 (%)

indicates that the number of currently active cases would be 6,000–7,200 cases, and the total cases of 100,000. Therefore, the active case at this moment (in October 2000) may be estimated to be somewhere between 6,000 and 100,000.

If we assume that 10,000 people in Japan (prevalence of approximately 0.008%) are currently infected with the disease, and if the entire population is tested for COVID-19 using RT-PCR (with a relatively high sensitivity of 0.9 and specificity of 0.95), the theoretical distributions would be summarized, as shown in Table 3. The PPV is calculated to be 0.14%. This means that almost all people with positive test results are *not* infected with COVID-19.

Even if we assume that 10 times greater number of people (or 100,000 people) are currently infected with COVID-19, and if the test is performed using the RT-PCR with a much higher sensitivity of 0.99 and specificity of 0.99 (Table 4), the PPV is calculated to be 7.3%. This means that, even if a test with a very high sensitivity and specificity is used, the positive test results would likely to be false-positive results when the prevalence of the disease is low (in this example, 0.08%).

If the prevalence of the disease is higher, the incidence of false-positive results will reduce. To take an example: Diamond Princess cruise ship, which saw an outbreak of coronavirus disease in January–February 2020, and Japanese Ministry of Health, Labour and Welfare [13] reports that 3,711 people were on board, and 712 (approximately 19%) were tested positive. If we assume that all of those 712 people were truly infected with the virus, and if a diagnostic test with a high sensitivity (0.99) and a high specificity (0.99) had been used (Table 5), the PPV would have been quite high (96%).

It is apparent from these simulations that the reliability of test (judged by the PPV) is strongly affected by the prevalence of the disease. Table 6 indicates the PPV and NPV for different prevalence of a disease. It can be seen that, when the prevalence of the disease is low, the rate of false positives is high, and the PPV can be increased when a diagnostic test with a high specificity is used.

**Table 3** Hypothetical infectious state and test results in Japan (hypothetical population of 126 million people, sensitivity of the test: 0.9, and specificity: 0.95)

Test results	Disease	
	Infected (10,000)	Not infected (125,990,000)
Positive (6,308,500)	9,000	6,299,500
Negative (11,9691,500)	1,000	11,9690,500

$PPV (9,000/6,308,500) \times 100 = 0.1427 (\%)$ ,  $NPV (11,9690,500 / 11,9691,500) \times 100 = 99.999 (\%)$

**Table 4** Hypothetical infectious state and test results in Japan (hypothetical population of 126 million people, sensitivity of the test: 0.99, and specificity: 0.99)

Test results	Disease	
	Infected (100,000)	Not infected (125,900,000)
Positive (1,358,000)	99,000	1,259,000
Negative (124,642,000)	1,000	124,641,000

$PPV (99,000/1,358,000) \times 100 = 7.290 (\%)$ ,  $NPV (124,641,000/124,642,000) \times 100 = 99.999 (\%)$

### Clinical implications

As the RT-PCR is the only recommended diagnostic test for the COVID-19 [4], we healthcare workers need to carry out necessary measures to treat patients and to prevent transmission of COVID-19, based on testing results. Nevertheless, as false positives or false-negative rates can be unacceptably high, the WHO states that “test results should always be considered in combination with other elements of the patient history, physical examination and the epidemiological context” [4].

Tedros Adhanom Ghebreyesus, the director-general of the WHO, stated at a news conference in Geneva on 16<sup>th</sup> March 2020, that “We have a simple message to all countries – test, test, test”. Some experts from a country where testing is not routinely performed claim that screening tests should be performed as the other countries are doing. Nevertheless, as confirmed above, the reliability of the diagnostic test can be considerably different between different countries with different prevalence of the disease. In the country where the prevalence is low (such as in Japan), the false-positive rate would be extremely high if screening test is performed. Therefore, in the area of low prevalence of COVID-19, diagnostic testing should be limited to people who are judged to be at increased risk of infection, or to people who have symptoms.

**Table 5** Hypothetical infectious state and test results for Diamond Princess cruise ship (3,711 people, hypothetical sensitivity of the test: 0.99, and specificity: 0.99)

Test results	Disease	
	Infected (712)	Not infected (2999)
Positive (735)	705	30
Negative (2976)	7	2969

$PPV (705/735) \times 100 = 95.9 (\%)$ ,  $NPV (2969/2976) \times 100 = 99.8 (\%)$

**Table 6** Hypothetical PPV and NPV of a test for different prevalence of disease

Sensitivity of 0.8; specificity of 0.9		
Prevalence (%)	PPV (%)	NPV (%)
0.01	0.1	99.99
0.1	0.8	99.98
1	7.5	99.8
10	47.1	97.6
20	66.7	94.7
30	77.4	91.3
50	88.9	81.8
Sensitivity of 0.99; specificity of 0.99		
Prevalence (%)	PPV (%)	NPV (%)
0.01	1.0	99.99
0.1	9.0	99.99
1	50.0	99.98
10	91.7	99.9
20	96.1	99.7
30	97.7	99.6
50	99.0	99.0
Sensitivity of 0.99; specificity of 0.997		
Prevalence (%)	PPV (%)	NPV (%)
0.01	3.2	99.99
0.1	24.8	99.99
1	76.9	99.99
10	97.3	99.9
20	98.8	99.7
30	99.3	99.6
50	99.7	99.0

If isolation of people with positive test result is the policy, each person should truly be isolated in a hospital room or a hotel room, and those people with positive test results should not be allowed to share a space even temporarily (for example, an elevator, or a dining area). This is because, the majority of people with positive test results are not infected with the disease, but if they contact people who are truly infected, those not infected can easily be infected, because of highly contagious nature of the disease.

In a statistical sense, the NPV of the currently available diagnostic tests is generally high (Table 6), and thus, the rate of false negatives is low. Nevertheless, in reality, diagnostic tests may frequently fail to detect positive cases, particularly 3–4 days after infection, and after 10–12 days of infection [5, 6]. Therefore, we should be aware that any asymptomatic people with negative test results can be infectious.

## Conclusions

RT-PCR is a useful diagnostic test for COVID-19, but we should adjust testing policy based on the prevalence of the disease, and should correctly interpret the test results, to establish preventive and treatment strategies, and to end the pandemic of COVID-19.

## References

1. Yamakage M. Anesthesia in the times of COVID-19. *J Anesth.* 2020. <https://doi.org/10.1007/s00540-020-02798-4>.
2. Asai T, O'Sullivan EP, Hemmings HC Jr. A Special issue on respiration and the airway: critical topics at a challenging time. *Br J Anaesth.* 2020;125:1–4.

3. Peng PWH, Ho P-L, Hota SS. Outbreak of a new coronavirus: what anaesthetists should know. *Br J Anaesth*. 2020;124:497–501.
4. World Health Organization (2020) COVID-19 Target product profiles for priority diagnostics to support response to the COVID-19 pandemic v.1.0. <https://www.who.int/publications/m/item/covid-19-target-product-profiles-for-priority-diagnostics-to-support-response-to-the-covid-19-pandemic-v.0.1>. Accessed 5 Oct 2020
5. Kucirka LM, Lauer SA, Laeyendecker O, Boon D, Lessler J. Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based SARS-CoV-2 tests by time since exposure. *Ann Intern Med*. 2020;173:262–7.
6. Woloshin S, Patel N, Kesselheim AS. False negative tests for sars-CoV-2 infection—challenges and implications. *N Engl J Med*. 2020;383:e38.
7. Casscells W, Schoenberger A, Graboys TB. Interpretation by physicians of clinical laboratory results. *N Engl J Med*. 1978;299:999–1001.
8. Watson J, Whiting PF, Brush JE. Interpreting a covid-19 test result. *Br Med J*. 2020. <https://doi.org/10.1136/bmj.m1808>.
9. Yerushalmy J. Statistical problems in assessing methods of medical diagnosis with special reference to X-ray techniques. *Public Health Rep*. 1947;62:1432–9.
10. Alcoba-Florez J, Gil-Campesino H, Artola DG, González-Montelongo R, Valenzuela-Fernández A, Ciuffreda L, Flores C. Sensitivity of different RT-qPCR solutions for SARS-CoV-2 detection. *Int J Infect Dis*. 2020;99:190–2.
11. Arevalo-Rodriguez I, Buitrago-Garcia D, Simancas-Racines D, et al. False-negative results of initial RT-PCR assays for covid-19: a systematic review. medRxiv. 2020. <https://doi.org/10.1101/2020.04.16.20066787>.
12. He J, Guo Y, Mao R, Zhang J. Proportion of asymptomatic coronavirus disease 2019: a systematic review and meta-analysis. *J Med Virol*. 2020. <https://doi.org/10.1002/jmv.26326>. Advanced access published on July 21.
13. Japanese Ministry of Health, Labour and Welfare. Response to the Diamond Princess cruise ship. In: Preventing the spread of infection and developing medical service systems. [https://www.mhlw.go.jp/stf/covid-19/kansenkakudaiboushi-iryouteikyoku\\_00005.html](https://www.mhlw.go.jp/stf/covid-19/kansenkakudaiboushi-iryouteikyoku_00005.html). Accessed 5 Oct 2020

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.