Journal of Data Science 19 (2), 243–252
April 2021
DOI: 10.6339/21-JDS993
Data Science in Action

Testing for COVID-19: Some Statistical Issues

Grace Y. Yi[1,2,*], Wenqing He[1], Dennis K. J. Lin[3], and Chun Ming Yu[4]

[1]*Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario, Canada*
[2]*Department of Computer Science, University of Western Ontario, London, Ontario, Canada*
[3]*Department of Statistics, Pennsylvania State University, University Part, PA, USA*
[4]*Northfield FHO @ Boardwalk Medical Center, Waterloo, Ontario, Canada*

Abstract

The swift spread of the novel coronavirus is largely attributed to its stealthy transmissions in which infected patients may be asymptomatic or exhibit only flu-like symptoms in the early stage. Undetected transmissions present a remarkable challenge for the containment of the virus and pose an appalling threat to the public. An urgent question is on testing of the coronavirus. In this paper, we evaluate the situation from the statistical viewpoint by discussing the accuracy of test procedures and stress the importance of rationally interpreting test results.

Keywords *COVID-19; false negative; false positive; pandemic; repeatedly testing*

1 Introduction

The coronavirus disease 2019 (COVID-19) was found to be caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Wong et al., 2020). While certain epidemiological and clinical characteristics of patients with COVID-19 have been reported (Xiao and Torok, 2020), the manifestation of the disease has not been fully understood (Zhou et al., 2020). The early presentation of COVID-19 infection may be nonspecific. Infected individuals often show flu-like symptoms such as dry cough, sore throat, low-grade fever, or malaise in the first few days (Wong et al., 2020), whereas other infected cases may be asymptomatic.

The mystery of the virus and the lack of effective treatment for COVID-19 have presented a striking threat to the public. The early diagnosis of infected patients is essential to manage the situation. While the testing capacity is ramping up, it is imperative to enhance our understanding of testing for COVID-19 by factoring in the assessment of the imperfectness associated with the test procedures; otherwise, misleading and erroneous outcomes can be produced. In this article, we examine test procedures in terms of their sensitivity and specificity, and quantify the degrees of false test results under several scenarios. We offer the assessment as to how likely we may miss identifying COVID-19 carriers based on consecutive negative results and how many times we should test a suspected COVID-19 patient to reduce the chance of errors.

2 Notation and Framework

In the following discussion, the term *population* is used to describe a group of subjects of interest; it may represent the collection of all people in a country, a city, or a region; it may also refer to a cohort of individuals, a ward of patients, or a community of people.

*Corresponding author. Email: gyi5@uwo.ca.

Received June 1, 2020; Accepted September 1, 2020

We let $t = 0$ represent the last day before *patient zero* (i.e., the first person who has COVID-19) appears, and we are interested in the status change of the population for the following days. Suppose that on day $t = 1, 2, \ldots$, the target population has $N(t)$ people in total in which $N_h(t)$ people have no COVID-19 and $N_s(t)$ people have COVID-19. Let $P(t) = N_s(t)/N(t)$ be the *prevalence* on day t. Then $P(1) > 0$, and $0 \leqslant P(t) \leqslant 1$ for $t = 2, 3, \ldots$. Due to the dynamic feature and the spread of the virus, the relative size of $N_h(t)$ to $N_s(t)$ varies with time t.

For *any* individual in the population, we are interested in the COVID-19 status for this individual. Let Y be the binary variable showing the *true* status for an individual to have COVID-19, with $Y = 1$ if having COVID-19 and 0 otherwise. In reality, the true value of Y is *unknown* for any individual, and we can only apply diagnostic tests to find out an individual's disease status. To feature this, let Y^* represent the test result for an individual who is tested: $Y^* = 1$ if the test result is positive, and $Y^* = 0$ if the test result is negative.

However, no medical test is 100% accurate in practice. There is a chance that a medical test can give us an incorrect result. To describe the accuracy of a test, we consider the *sensitivity* and the *specificity*, respectively, defined as

$$p_{sen} = P(Y^* = 1|Y = 1) \text{ and } p_{spe} = P(Y^* = 0|Y = 0).$$

Basically, the sensitivity p_{sen} reports the probability that a test successfully confirms the true status for an individual having the disease. In other words, the value of p_{sen} indicates the success rate of the test when applied to the *subpopulation* of diseased people, so the sensitivity p_{sen} is also called the *true positive rate*. On the other hand, the specificity p_{spe} measures the probability that the test successfully reveals the disease-free status for any individual who has no disease. The value of p_{spe} indicates the proportion of the time for obtaining the correct result when the test is applied to the *subpopulation* of healthy people. Consequently, the specificity p_{spe} is also called the *true negative rate*.

Alternatively, as commented by a referee, the *positive predictive value* (PPV) and the *negative predictive value* (NPV) can be used to describe the performance of a diagnostic test. Let TP represent the number of true positives, FP be the number of false positives, TN denote the number of true negatives, and FN stand for the number of false negatives. The PPV and NPV are, respectively, defined as

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \text{and} \quad \text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}}.$$

Those measures are connected via the equalities

$$\text{PPV} = \frac{p_{sen} \times \text{prevalence}}{p_{sen} \times \text{prevalence} + (1 - p_{spe}) \times (1 - \text{prevalence})},$$

and

$$\text{NPV} = \frac{p_{spe} \times (1 - \text{prevalence})}{(1 - p_{sen}) \times \text{prevalence} + p_{spe} \times (1 - \text{prevalence})}.$$

In the following sections, we use p_{spe} and p_{spe}, in combination with the prevalence, to carry out the discussion, bearing in mind that PPV and NPV can be equivalently employed.

3 The Numbers of False Results at the Population Level

To assess the impact of false test results, we consider a hypothetical situation where everyone in the population is to be tested on day $t = 1, 2, \ldots$. We are interested in evaluating two numbers

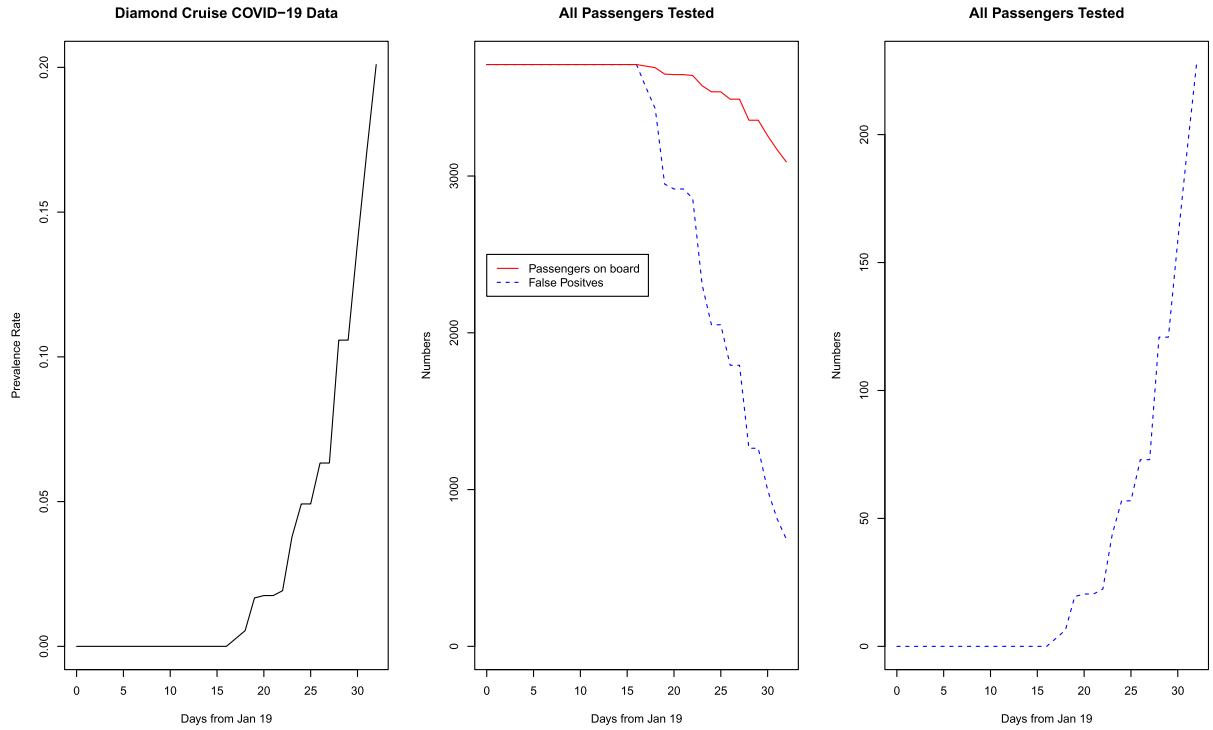

Figure 1: Analysis of the Diamond Princess data if everyone would be tested everyday from day 17 to day 44: the left panel records the prevalence versus the date, the right panel shows the false negative number versus the date, and the middle panel displays the false positive number versus the date (in blue) and the population size versus the date (in red).

at the population level:

$$\#_{\text{fp}}(t) = \text{the number of false positives on day } t$$

and

$$\#_{\text{fn}}(t) = \text{the number of false negatives on day } t,$$

which can be respectively approximated by

$$\#_{\text{fp}}(t) \approx N_h(t) \times (1 - p_{spe}) \quad \text{and} \quad \#_{\text{fn}}(t) \approx N_s(t) \times (1 - p_{sen}). \tag{1}$$

For illustrations, we analyze the data of the Diamond Princess cruise for the period of January 19, 2020 (the day before *patient zero* embarked on the cruise) to February 20, 2020 when all passengers were disembarked (Diamond Princess, 2020). There were 2,666 guests and 1,045 crew on board. *Patient zero* stayed on Diamond Princess for five days and disembarked in Hong Kong on January 25, 2020. During those five days, patient zero did not report being ill; he was tested positive for COVID-19 six days after leaving the ship. On February 4, 2020, 10 people were tested positive for COVID-19 among the first batch of tested passengers. In subsequent days, more guests were tested positive. People with positive test results were transported to local hospitals for medical care. Table 1 displays the number of people whose test results were positive on different days. Using the notation in Section 2, January 19 of 2020 is taken as $t = 0$ on which $N(0) = N_h(0) = 3711$ and $N_s(0) = 0$; January 20 of 2020 is taken as $t = 1$ on which $N(1) = N(0)$

Table 1: COIVD-19 Data from Diamond Princess Cruise.

Day	1-16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33-44
# Cases	0	10	10	41	3	0	6	65	39	0	47	0	134	0	99	88	79	84

and $N_s(1) = 1$; on February 4 of 2020 (i.e., $t = 17$), $N_s(17) = 10$ and $N(17) = N(0) - 10$. The left panel of Figure 1 displays the day-dependent prevalence, and in the middle panel we report the changing population size using the red curve. The accuracy of the current COVID-19 tests is not precisely known. Here we use the estimated sensitivity and specificity of COVID-19 tests, respectively, 70% and 95%, suggested by Watson et al. (2020) and Woloshin et al. (2020). Assuming that *all* passengers would be tested *every day* starting $t = 17$, we approximate the daily numbers of false positives and negatives results using (1), and display them in the middle panel (blue dashed curve) and the right panel in Figure 1. Clearly, when $P(t)$ is very small, $\#_{fp}(t)$ is close to $N(t)$ and $\#_{fn}(t)$ is near 0; when $P(t)$ becomes larger, $\#_{fp}(t)$ becomes smaller but $\#_{fn}(t)$ gets larger.

4 Chance of Error at the Individual Level

We are interested in interpreting test results if the same test is repeatedly applied to an individual. To this end, let Y_k^* represent the kth result of applying the test to an individual, where $k = 1, \ldots, K$ with K greater than 1. We assume that the test is independently applied to an individual a number of times and that the Y_k^* have the same distribution as that of Y.

4.1 What is the chance of being a virus carrier if a positive test result appears after a sequence of consecutive negative results?

We want to find the value of K so that the conditional probability $P(Y = 1|Y_1^* = 0, \ldots, Y_{K-1}^* = 0, Y_K^* = 1)$ is smaller than a tolerance value. Using Bayes' theorem gives

$$\begin{aligned}
&P(Y = 1|Y_1^* = 0, \ldots, Y_{K-1}^* = 0, Y_K^* = 1) \\
=&\frac{P(Y_1^* = 0, \ldots, Y_{K-1}^* = 0, Y_K^* = 1|Y = 1)P(Y = 1)}{\sum_{r=0,1} P(Y_1^* = 0, \ldots, Y_{K-1}^* = 0, Y_K^* = 1|Y = r)P(Y = r)} \\
=&\frac{(1 - p_{sen})^{K-1} p_{sen} P(t)}{(1 - p_{sen})^{K-1} p_{sen} P(t) + p_{spe}^{K-1}(1 - p_{spe})(1 - P(t))} \\
=&\frac{1}{1 + \left(\frac{p_{spe}}{1-p_{sen}}\right)^{K-1} \left(\frac{1-p_{spe}}{p_{sen}}\right) \left(\frac{1-P(t)}{P(t)}\right)},
\end{aligned}$$

where we assume that $P(Y = 1)$ is identical to $P(t)$ for $t = 1, 2, \ldots$.

Figure 2 reports the conditional probability $P(Y = 1|Y_1^* = 0, \ldots, Y_{K-1}^* = 0, Y_K^* = 1)$ versus the sensitivity and specificity of the test for populations with different prevalence for $K = 3$. For a test with a low sensitivity, there is a high chance that a patient is infected even when the first positive result appears at the third test if we test individuals from the population with a high prevalence (such as 85%). However, if the population has a small prevalence (such as 1%), it is unlikely that the patient with the first positive result occurring at the third try is infected.

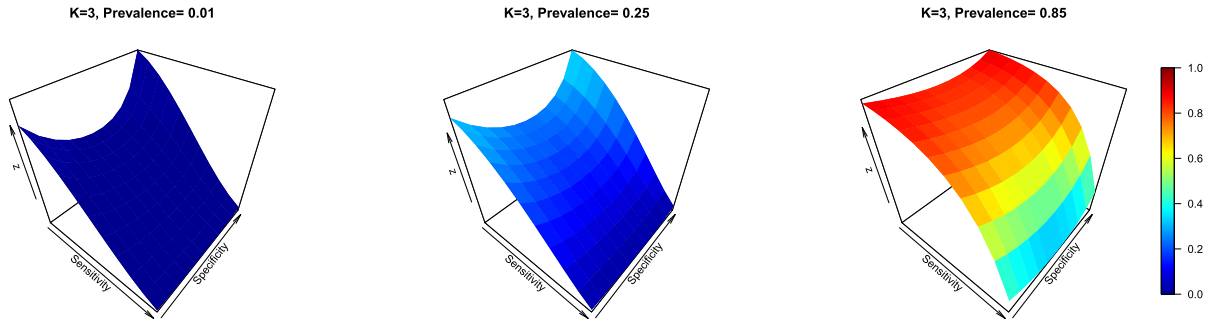

Figure 2: The conditional probability $P(Y = 1|Y_1^* = 0, \ldots, Y_{K-1}^* = 0, Y_K^* = 1)$ versus the sensitivity and specificity of the test: $K = 3$ and the prevalence $P(t) = 0.01, 0.25, 0.85$; The color shows the magnitude of the probability.

Message 1. *With a given test, when receiving a positive result after consecutive negative results, the interpretation is different for patients coming from different cohorts with different prevalences.*

To further visualize how the cohort prevalence affects the probability of identifying an infected case, we examine the COVID-19 tests described by Watson et al. (2020) and Woloshin et al. (2020) as opposed to the *COVID-19 IgM/IgG Rapid Test*, a test newly released by the ISO13485 registered company *BioMedomics* (BioMedomics, 2020). On March 8, 2020, the company announced that it has received CE Mark-IVD certification for its new test to help diagnose COVID-19. This test, available only for research use at this stage, takes 15 minutes to obtain the result and can be used for rapid screening of COVID-19 carriers. The sensitivity and specificity of the test were estimated to be 88.66% and 90.63%, respectively, based on the test results for 525 infected cases and 128 non-SARS-CoV-2 infection patients (BioMedomics, 2020).

We graph the results of $P(Y = 1|Y_1^* = 0, ..., Y_{K-1}^* = 0, Y_K^* = 1)$ in Figure 3. There is a good chance that the tested person contracts COVID-19 if the first positive result of applying the COVID-19 IgM/IgG Rapid Test occurs at the second test, unless the population prevalence is very small; the chance is higher for testing patients coming from a cohort with a higher prevalence. For the COVID-19 IgM/IgG Rapid Test, if the positive result occurs only at the 4th test, then the chance of the tested subject is infected is very slim unless the patient comes from a cohort with a high prevalence (such as 60% or higher); if the first positive result occurs at the 6th test, the chance of the tested subject is infected is almost 0 no matter which cohort this patient comes from. However, for the COVID-19 test described by Watson et al. (2020) and Woloshin et al. (2020), even the first five test results are negative, there is still a good chance that the tested subject is infected with COVID-19.

Message 2. *With the same cohort of patients, different interpretations should be given for the positive result after the same number of consecutive negative results that are obtained from different tests with different sensitivities and specificities.*

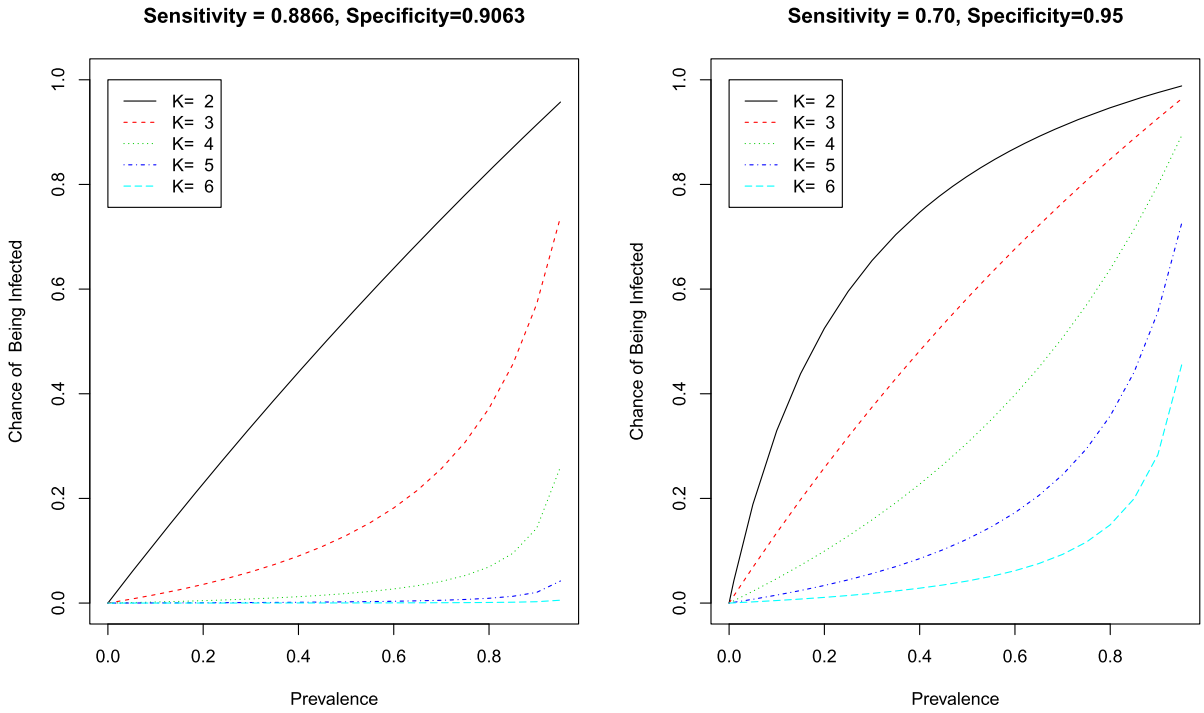

Figure 3: The probability of identifying an infected case after $K-1$ consecutive negative results followed by a positive result: the plots of the probability versus the prevalence for $K = 2, \ldots, 6$; The left panel is for the COVID-19 IgM/IgG Rapid Test and the right panel is for the COVID-19 test described by Watson et al. (2020) and Woloshin et al. (2020).

4.2 What is the chance of being a virus carrier if receiving a sequence of consecutive negative test results?

We are interested in assessing the conditional probability $P(Y = 1|Y_1^* = 0, \ldots, Y_{K-1}^* = 0, Y_K^* = 0)$ for $K = 2, 3, \ldots$. Using Bayes' theorem, we obtain that

$$\begin{aligned} & P(Y = 1|Y_1^* = 0, \ldots, Y_{K-1}^* = 0, Y_K^* = 0) \\ = & \frac{P(Y_1^* = 0, \ldots, Y_{K-1}^* = 0, Y_K^* = 0|Y = 1)P(Y = 1)}{\sum_{r=0,1} P(Y_1^* = 0, \ldots, Y_{K-1}^* = 0, Y_K^* = 0|Y = r)P(Y = r)} \\ = & \frac{(1 - p_{sen})^K P(t)}{(1 - p_{sen})^K P(t) + p_{spe}^K \{1 - P(t)\}} \\ = & \frac{1}{1 + \left(\frac{p_{spe}}{1-p_{sen}}\right)^K \times \left(\frac{1-P(t)}{P(t)}\right)}. \end{aligned} \tag{2}$$

Given the prevalence $P(t)$, the conditional probability (2) decreases as K increases *if and only if*

$$\frac{p_{spe}}{1 - p_{sen}} > 1,$$

which is satisfied by a test with the sensitivity larger than the false positive rate (i.e., $1 - p_{spe}$) or the specificity higher than the false negative rate (i.e., $1 - p_{sen}$). This condition must be met

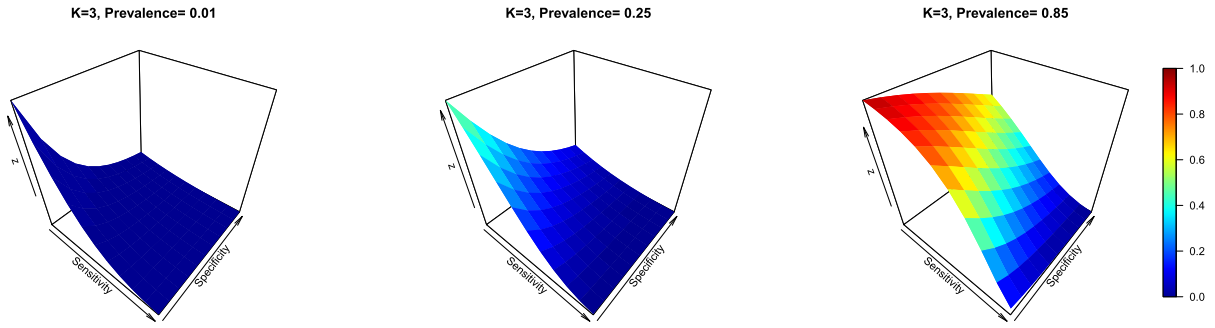

Figure 4: The chance of missing a case with K consecutive negative results versus the sensitivity and specificity of the test: $K = 3$ and the prevalence $P(t) = 0.01, 0.25, 0.85$; The color shows the magnitude of the chance.

by any test in use; otherwise, there is no point of using a test that is even worse than a random guess.

Message 3. *To reduce the chance of missing a COVID-19 carrier based on negative test results, it is important to increase the number of tests.*

Figure 4 shows 3-dimensional graphs of the conditional probability

$$P(Y = 1|Y_1^* = 0, \ldots, Y_{K-1}^* = 0, Y_K^* = 0)$$

for $K = 3$ and populations with different prevalences in contrast to the sensitivity and specificity of the test. If the test has a low sensitivity, the chance of missing an infected case is high if the test is done only once for populations with a large prevalence, even if the specificity of the test is high; the larger the prevalence, the higher the chance of missing. For a test with a reasonably large sensitivity, the more we test, the smaller the chance of missing an infected case.

To see how the prevalence may affect the conditional probability (2), we consider a situation of discharging inpatients. Again we compare the COVID-19 tests described by Watson et al. (2020) and Woloshin et al. (2020) to the *COVID-19 IgM/IgG Rapid Test*. Figure 5 displays how the probability of missing an infected case after receiving K consecutive negative test results depends on the population prevalence of COVID-19 for $K = 2, \ldots, 6$. With the COVID-19 IgM/IgG Rapid Test, having two consecutive negative test results ensures a nearly zero chance of missing infected cases if the tested inpatients come from a ward with the prevalence lower than 40%; for the inpatient ward with about 80% COVID-19 carriers, obtaining 3 consecutive negative test results warrants a slim chance of missing infected cases; receiving 4 consecutive negative results is enough for discharging any inpatients.

In comparison, the COVID-19 test described by Watson et al. (2020) and Woloshin et al. (2020) has about 18.66% smaller sensitivity than the COVID-19 IgM/IgG Rapid Test. It requires more consecutive negative results than the COVID-19 IgM/IgG Rapid Test for retaining a slim chance of missing infected cases, and different numbers of consecutive negative results produced by this test yield more different probabilities of missing infected cases than those obtained from the COVID-19 IgM/IgG Rapid Test. This comparison illustrates how a high sensitivity of a test can make a difference in reducing the chance of missing infected cases based on consecutive negative results.

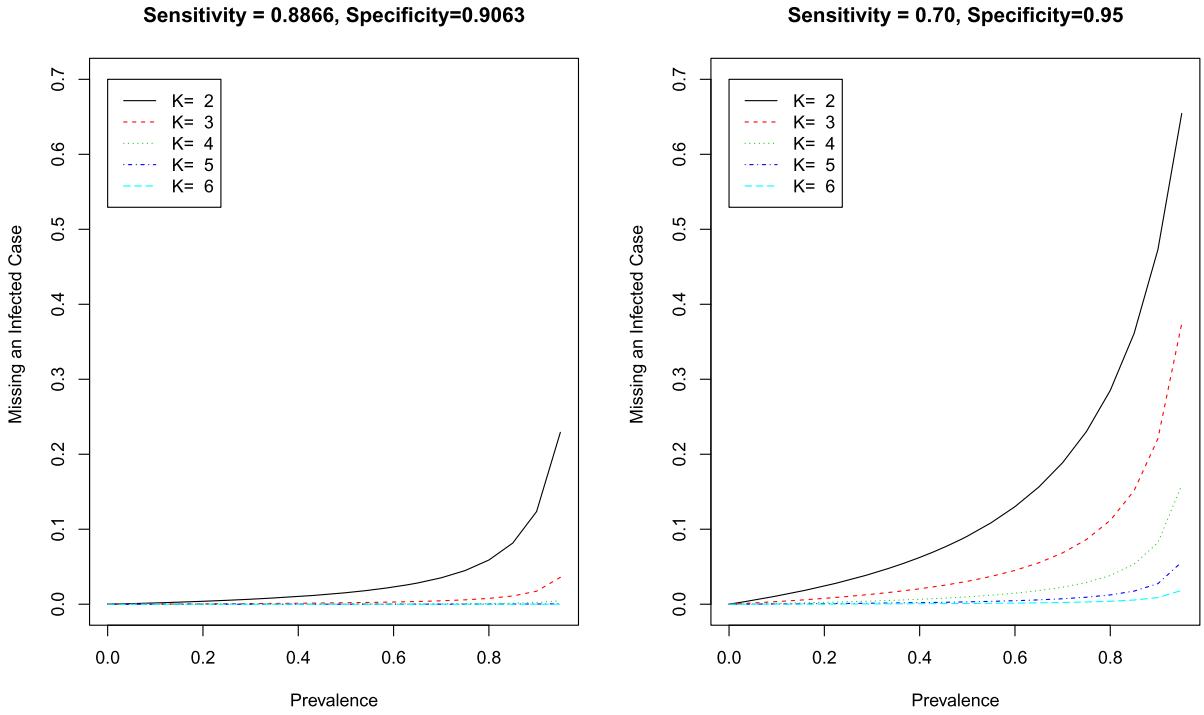

Figure 5: The conditional probability for missing an infected case based on K consecutive negative results: the plots of the conditional probability versus the prevalence for $K = 2, \ldots, 6$; The left panel is for the COVID-19 IgM/IgG Rapid Test and the right panel is for the COVID-19 test described by Watson et al. (2020) and Woloshin et al. (2020).

5 Conclusion and Discussion

In this article we take the statistical standpoint and examine several aspects of COVID-19 testing. We evaluate the uncertainty induced from the imperfectness of COVID-19 tests and highlight the importance of rationally interpreting test results. This article provides the assessment as to how likely we may miss identifying COVID-19 carriers based on consecutive negative results and how many times we should test a suspected COVID-19 patient to reduce the chance of errors. On equal footing, our discussion provides the guidelines for discharging inpatients who are treated for COVID-19. Several aspects warrant our further attention. In the discussion of the Diamond Princess cruise data in Section 3, the dynamic number of COVID-19 carriers, $N_s(t)$, is taken as the confirmed cases for each day t, which is very likely to be smaller than the *true* number of infected cases due to the undetected asymptomatic infections and the time lag related to the disease incubation period. While we consider a dynamic framework in Section 2 to characterize the change in the population size and the relationship between the number of infected people and the number of infection-free people, our discussion on the COVID-19 status for individuals focuses on a *static* state to highlight the ideas. A time-dependent status $Y(t)$, instead of Y, should be used to better reflect the dynamic status for individuals as well as the incubation period for infected cases.

The test performance may depend on what is being measured (e.g., nucleic acids or antibodies) and likely varies with application over time, as commented by a referee. The discussion

here, however, does not explicitly factor in these aspects nor patients' characteristics such as age, severity of COVID-19-like symptoms, and the history of health conditions. Due to the lack of information on the sensitivity and specificity of available test kits estimated from different cohorts of patients with varying medical conditions, we are unable to conduct more refined discussion by accommodating these features. Motivated by this, here we highlight

Message 4. *Instead of being assessed by an* overall *sensitivity and specificity, the performance of COVID-19 tests should be evaluated in a more refined measure by reporting their sensitivities and specificities obtained from the stratified population by the patient's medical conditions.*

In the article, we compare two types of COVID-19 tests: the COVID-19 tests described by Watson et al. (2020) and Woloshin et al. (2020) and the COVID-19 IgM-IgG Rapid Test. While the calculations show that the latter test outperforms the former tests, we are not ready to recommend to replace the former tests by the latter one. While the estimated sensitivity and specificity of the COVID-19 IgM-IgG Rapid Test are higher than those described by Watson et al. (2020) and Woloshin et al. (2020), those estimates are obtained from different groups of patients whose conditions may differ considerably and the sizes may not be comparable either. Accurately assessing the sensitivity and specificity of COVID-19 tests is challenging due to a number or reasons, including the lack of a clear-cut "gold-standar", the dependence on the site and quality of sampling, the stage of disease, the degree of viral multiplication or clearance, and the nature of test procedures (Watson et al., 2020). In addition, asymptomatic manifestation of COVID-19 adds extra difficulty in gaining an accurate estimate of the sensitivity and specificity for a COVID-19 test.

In our discussion of repetitions of a test procedure in Section 4, we assume that the test is *independently* repeated. This assumption, as pointed out by the Associate Editor, is not realistic, and consequently, the discussion here can only be taken as an approximation to describing the *underlying unknown* truth. With multiple test kits becoming available, we face the decision of choosing suitable test kits to reach a balance among several key factors. These includes, but are not limited to, the time of acquiring results, the associated cost, the test accuracy, and the suitability for patients with different conditions. We may apply a fast but less accurate test kit to do screening, and then apply a more accurate but time-consuming and costly test to do further checks. It is useful to ponder the questions: How do we use them effectively? How to combine two test procedures with different sensitivities and specificities to complement their benefits? In addition, in our discussion of repeating the test for COVID-19, we have not looked into the issue of the gap time between two consecutive tests.

Supplementary Material

The Diamond Princess data and the R code for the figures in the paper are available online at the *Journal of Data Science* website.

Acknowledgements

The authors thank the Editor Jun Yan and the review team for their helpful comments on the initial version. The research of Yi and He is partially supported by fundings from the Natural Sciences and Engineering Research Council of Canada (NSERC). Yi is Canada Research Chair in Data Science (Tier 1). Her research was undertaken, in part, thanks to funding from the Canada Research Chairs Program.

References

BioMedomics (2020). COVID-19 rapid test. https://www.biomedomics.com/products/infectious-disease/covid-19-rt/.

Diamond Princess (2020). Diamond princess updates. https://www.princess.com/news/notices%20_and_advisories/notices/diamond-princess-update.html.

Watson J, Whiting PF, Brush JE (2020). Interpreting a Covid-19 test result. *BMJ*, 369.

Woloshin S, Patel N, Kesselheim AS (2020). False negative tests for SARS-CoV-2 infection—challenges and implications. *New England Journal of Medicine*, 383(6): e38.

Wong JE, Leo YS, Tan CC (2020). COVID-19 in Singapore—current experience: Critical global issues that require attention and action. *JAMA*, 323(13): 1243–1244.

Xiao Y, Torok ME (2020). Taking the right measures to control COVID-19. *The Lancet Infectious Diseases*, 20(5): 523–524.

Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. *The lancet*, 395(10229).