Tool for estimating the probability of having COVID-19 with one or more negative RT-PCR results

Alejandro Jara^{1,2*}, Eduardo A. Undurraga^{3,4}, Rafael Araos^{4,5}

¹ *Department of Statistics, Pontificia Universidad Católica de Chile, Santiago, RM, Chile*

² *Millennium Nucleus Center for the Discovery of Structures in Complex Data (MIDaS) Santiago, RM, Chile*

³ *Escuela de Gobierno, Pontificia Universidad Católica de Chile, Santiago, RM, Chile*

⁴ *Millennium Initiative for Collaborative Research in Bacterial Resistance (MICROB-R), Santiago, RM, Chile*

⁵ *Instituto de Ciencias e Innovación en Medicina, Universidad del Desarrollo, Santiago, RM, Chile*

Abstract

Early case detection and isolation of infected individuals are critical to controlling COVID-19. RT-PCR is considered the diagnosis gold standard, but false-negatives occur. Based on previous work, we built a user-friendly online tool to estimate the probability of having COVID-19 with negative RT-PCR results and thus avoid preventable SARS-CoV-2 transmission.

Keywords: COVID-19, RT-PCR test, Bayesian, online tool

***Address for correspondence:** Alejandro Jara, Department of Statistics, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Macul CP 7820436, Santiago, Región Metropolitana, Chile. atjara@uc.cl, Tel +56-2-3544506, Fax: +56-2-3547729

The COVID-19 pandemic has imposed a catastrophic toll worldwide, with about 66 million reported cases and 1.5 million reported deaths as of December 4, 2020 [1]. Despite the historic approval of a COVID-19 vaccine tested in a large clinical trial in the United Kingdom [2], epidemic control still critically depends on non-pharmaceutical interventions, such as social distancing, and early detection and effective isolation of infected individuals [3].

Compared to other viruses, SARS-CoV-2, the virus that causes COVID-19 disease, has two characteristics that make early case detection and isolation of infected individuals particularly critical for epidemic control. First, the virus is an efficient spreader, with an average of about 2.5 secondary infections caused by a single infected individual in a susceptible population [4]. Second, the highest risk of transmission occurs very early in the disease, before or within the first days of symptom onset [5]. These characteristics hinder transmission control because early case detection and effective isolation are challenging. Thus far, COVID-19 surveillance has been primarily based on reverse transcriptase-polymerase chain reaction (RT-PCR) assays, considered the most reliable diagnostic test for COVID-19 [6]. However, RT-PCR positivity varies among infected patients depending on, for example, the timing of sample collection in relation to symptom onset or the sampling technique used (e.g., nasopharyngeal swabs, sputum) [7].

Accurate testing results are critical to prevent onward SARS-CoV-2 transmission in the community and hospital settings. As the second wave of infection continues to grow in Europe and the United States, straining health system capacity even further, critical workers with known exposures to the virus will continue to use RT-PCR tests to discard SARS-CoV-2 infection. False-negative RT-PCR results are particularly problematic among healthcare and other essential workers (e.g., firefighters, police), which may inadvertently become “superspreaders” [8]. As social distancing measures are relaxed, RT-PCR false-positives also affect community control measures if workers with suspected infection are cleared to return to work. If negative RT-PCR results are treated as evidence of no-infection, there is a non-negligible risk of preventable transmission of the virus.

We addressed this problem by designing a readily available, easy-to-use online tool to estimate the probability that an individual is infected with SARS-CoV-2 conditional on having one or more negative RT-PCR test results (<https://midas-uc.shinyapps.io/Calculadora-COVID19/>). Our tool, based mainly on the work by Kucirka et al. [9], requires users to first choose between one, two, or three consecutive negative RT-PCR tests. Second, to provide (i) the estimated probability of being positive before taking the RT-PCR diagnostic test (range: 0 to 1), and (ii) the number of days elapsed between the first symptoms and the first, second, or third RT-PCR test (range: -4 to 16 days). The tool is available in English and Spanish.

We build on Kucirka et al.'s [9] work and adjusted a hierarchical logistic model to estimate the rate of false negatives for different moments in time from the onset of symptoms. We implemented the model using JAGS and the RJAGS library in R [10]. We generated a Markov chain of 420,000 samples; the first 20,000 samples were discarded, and the rest were re-sampled to generate a sub-chain of size 20,000. In contrast to Kucirka et al. [9], we estimated a non-study-specific marginal rate of false-negative RT-PCR tests. Our online tool assumes a specificity of one of the RT-PCR test for detecting SARS-CoV-2 and independence of the results of different tests when considering more than one test (specific details are shown in the Technical Appendix).

To show the tool's utility, we input data from a healthcare worker in a hospital setting in Chile who was not working with COVID-19 patients. The worker was inadvertently exposed to SARS-CoV-2 by an asymptomatic patient who later developed symptoms and had confirmed COVID-19. The worker had a negative nasopharyngeal swab RT-PCR result three days following exposure. She reported symptoms the day after the first negative RT-PCR and had a second RT-PCR test that showed positive results (Technical Appendix). Because she was exposed for a relatively long period, we used a 50% pre-test probability of infection (Figure).

Our online tool enables users to estimate the probability of having COVID-19 with one or more negative RT-PCR results. The tool is available in English and Spanish. We hope this publicly available online tool will help decision-makers to avoid preventable transmission of SARS-CoV-2.

Acknowledgments

This work was supported by the ANID Millennium Science Initiative grants MIDAS NCN17_059 and MICROB-R NCN17_081.

Figure. Illustrative results from the tool to estimate the probability of having COVID-19 with one negative RT-PCR tool. (a) The worker reported symptoms the day after the first negative RT-PCR; because she was exposed for a relatively long period of time we used a 50% pre-test probability of infection (b) shows the posterior distribution of the sensitivity of the PCR test to the SARS-CoV-2 virus, and (c) shows the posterior distribution of the probability of being positive for the SARS-CoV-2 virus, given the negative result in a PCR test

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis.* 2020;20(5):P533-P4.
2. Ledford H, Cyranoski D, Noorden RV. The UK has approved a COVID vaccine — here's what scientists now want to know. *Nature.* 2020;588:205-6. doi: <https://doi.org/10.1038/d41586-020-03441-8>.
3. Li Y, Campbell H, Kulkarni D, Harpur A, Nundy M, Wang X, et al. The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries. *Lancet Infect Dis.* 2020;online October 22. doi: [https://doi.org/10.1016/S1473-3099\(20\)30785-4](https://doi.org/10.1016/S1473-3099(20)30785-4).
4. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet.* 2020;395(10225):689-97.
5. Cevik M, Tate M, Lloyd O, Maraolo AE, Schafers J, Ho A. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. *The Lancet Microbe.* 2020. doi: 10.1016/S2666-5247(20)30172-5.
6. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. *JAMA.* 2020;324(8):782-93. doi: 10.1001/jama.2020.12839.
7. Sethuraman N, Jeremiah SS, Ryo A. Interpreting Diagnostic Tests for SARS-CoV-2. *JAMA.* 2020;323(22):2249-51. doi: 10.1001/jama.2020.8259.
8. Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, et al. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature.* 2020. doi: 10.1038/s41586-020-2923-3.
9. 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.
10. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2019.
11. Danis K, Epaulard O, Bénet T, Gaymard A, Campoy S, Bothelo-Nevers E, et al. Cluster of coronavirus disease 2019 (Covid-19) in the French Alps, 2020. *Clin Infect Dis.* 2020.
12. Guo L, Ren L, Yang S, Xiao M, Chang D, Yang F, et al. Profiling early humoral response to diagnose novel coronavirus disease (COVID-19). *Clin Infect Dis.* 2020.
13. Kim ES, Chin BS, Kang CK, Kim NJ, Kang YM, Choi J-P, et al. Clinical course and outcomes of patients with severe acute respiratory syndrome coronavirus 2 infection: a preliminary report of the first 28 patients from the Korean cohort study on COVID-19. *J Korean Med Sci.* 2020;35(13).
14. Kujawski SA, Wong KK, Collins JP, Epstein L, Killerby ME, Midgley CM, et al. First 12 patients with coronavirus disease 2019 (COVID-19) in the United States. *MedRxiv.* 2020.

15. Liu L, Liu W, Zheng Y, Jiang X, Kou G, Ding J, et al. A preliminary study on serological assay for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in 238 admitted hospital patients. *Microbes Infect.* 2020.
16. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, et al. Virological assessment of hospitalized patients with COVID-2019. *Nature.* 2020;581(7809):465-9.
17. Zhao J, Yuan Q, Wang H, Liu W, Liao X, Su Y, et al. Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. *Clin Infect Dis.* 2020;71(16):2027-34.
18. Plummer M, editor JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd international workshop on distributed statistical computing*; 2003: Vienna, Austria.
19. Plummer M, Stukalov A, Denwood M. rJAGS: Bayesian Graphical Models Using MCMC 2019. Available from: <https://cran.r-project.org/web/packages/rjags/index.html>.

Technical appendix

Tool for estimating the probability of having COVID-19 with one or more negative RT-PCR results

Our online tool is based on Kucirka et al. [9]. In Kucirka et al.'s article, the data of samples from the upper respiratory tract of $n = 1330$ patients reported in seven studies on the performance of the PCR test for the detection of the SARS-CoV-2 virus at different moments of the disease development [11-17] were analyzed.

Kucirka et al. [9] considered only nasal samples in their analysis and adjusted a hierarchical logistic model to estimate the rate of false negatives for different moments in time from the onset of symptoms. Let Y_{ij} be the number of positive patients, out of a total of n_{ij} , in the study $i, i = 1, \dots, 7$, at day j since exposure to SARS-CoV-2. Kucirka et al. [9] considered the hierarchical model given by

$$Y_{ij} | \alpha_i, \beta_1, \beta_2, \beta_3 \sim \text{Binomial}(n_{ij}, \pi_{ij}), \quad i = 1, \dots, 7, \quad j = 1, \dots, n_i,$$

$$\text{logit}(\pi_{ij}) = \alpha_i + \sum_{k=1}^3 \beta_k B_k(\log(j)),$$

$$\alpha_i | \beta_0, \sigma^2 \sim N(\beta_0, \sigma^2), \quad i = 1, \dots, 7,$$

where B_k is the function that evaluates the orthogonal polynomial of degree $k, \beta_1, \dots, \beta_3$ are regression coefficients, α_i is the random effect of study $i, i = 1, \dots, 7, N(\mu, \tau)$ denotes the normal distribution with mean μ and variance τ . Kucirka et al. [9] implemented a Bayesian version of the model using the STAN library of the statistical program R [10]. They estimated the rate of false negatives on the day j , based on the expression

$$\text{logit}(\pi_j) = \beta_0 + \sum_{k=1}^3 \beta_k B_k(\log(j)),$$

which generates a conditional (study-specific) estimate of the false-negative rate.

We partially reproduced the analysis carried out by Kucirka et al. [9]. Specifically, we considered the same data and hierarchical model, which was implemented using the following prior distributions:

$$\beta_k \sim N(0, 10^4), \quad k = 0, 1, \dots, 4,$$

$$\sigma \sim U(0, 10^3)$$

We implemented the model using the JAGS program [18] and the RJAGS library [19] <https://cran.r-project.org/web/packages/rjags/index.html> of the R statistical program [10]. We generated a Markov chain of 420,000 samples; the first 20,000 samples were discarded, and the rest were re-sampled to generate a sub-chain of size 20,000.

Unlike Kucirka et al.'s [9] work, we estimated the false-negative rate using the expression

$$\pi_j = \int_{-\infty}^{\infty} \frac{\exp[\alpha + \sum_{k=1}^3 \beta_k B_k(\log(j))]}{1 + \exp[\alpha + \sum_{k=1}^3 \beta_k B_k(\log(j))]} d(\alpha | \beta_0, \sigma^2) d\alpha$$

where $d(\cdot | \mu, \sigma^2)$ denotes the density of the normal distribution with mean μ and variance σ^2 , which generates an estimate of the marginal rate of false negatives (not study-specific). Figure S1 shows the posterior median and the 95% credibility interval limits for the false-negative rate.

Figure S1. Posterior median and 95% credibility intervals for the false-negative rate of PCR tests by day since exposure to SARS-CoV-2

Our online tool assumes a specificity of one of the PCR tests for detecting SARS-CoV-2 virus and independence of the results of different tests when considering more than one test. The tool is publicly available in English and Spanish at <https://midas-uc.shinyapps.io/Covid19-calculator/> and <https://midas-uc.shinyapps.io/Calculadora-COVID19/>, respectively.