Supplementary Data

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

Alejandro Jara, Eduardo A. Undurraga, Rafael Araos*

S1. Estimation of the false negative rate

Our online tool is based on Kucirka et al.¹ In that 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 were analyzed.²⁻⁸

Kucirka et al.¹ 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 \( i, i = 1, \ldots, 7 \), at day \( j \) since exposure to SARS-COV-2. Kucirka et al.¹ considered the hierarchical model given by

\[
Y_{ij} | \alpha_i, \beta_1, \beta_2, \beta_3 \sim \text{Binomial} \left( n_{ij}, \pi_{ij} \right), \quad i = 1, \ldots, 7, \quad j = 1, \ldots, n_i,
\]

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

* Pontificia Universidad Católica de Chile, Santiago, Chile (A. Jara, E.A. Undurraga); Millenium Nucleus Center for the Discovery of Structures in Complex Data (MIDaS) Santiago, Chile (A. Jara); Universidad del Desarrollo, Santiago, Chile (R. Araos); Millennium Initiative for Collaborative Research in Bacterial Resistance (MICROB-R), Santiago, Chile (R. Araos, E.A. Undurraga); Emails: atjara@uc.cl, eundurra@uc.cl, rafaelaraos@udd.cl
\[ \alpha_i | \beta_0, \sigma^2 \sim N(\beta_0, \sigma^2), \ i = 1, \ldots, 7, \]

where \( B_k \) is the function that evaluates the orthogonal polynomial of degree \( k \), \( \beta_1, \ldots, \beta_3 \) are regression coefficients, \( \alpha_i \) is the random effect of study \( i \), \( i = 1, \ldots, 7 \), \( N(\mu, \tau) \) denotes the normal distribution with mean \( \mu \) and variance \( \tau \). Kucirka et al.\(^1\) implemented a Bayesian version of the model using the STAN library of the statistical program R.\(^9\) 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.\(^1\). Specifically, we considered the same data and hierarchical model, which was implemented using the following prior distributions:

\[
\beta_k \sim N(0, 10^4), \quad k = 0, 1, \ldots, 4,
\]

\[
\sigma \sim U(0, 10^3)
\]

We implemented the model using the JAGS program\(^10\) and the RJAGS library\(^11\) of the R statistical program.\(^9\) 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\(^1\) 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 \( \mu \) and variance \( \sigma^2 \), 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/](https://midas-uc.shinyapps.io/Covid19-calculator/) and [https://midas-uc.shinyapps.io/Calculadora-COVID19/](https://midas-uc.shinyapps.io/Calculadora-COVID19/), respectively.
S2. Estimation of the pre-test probability for Covid-19

We estimated the pre-test probability, as a function of relevant covariates, based on RT-PCR test results from a real-world sample of 926 individuals with suspected SARS-CoV-2 infection or close contact with a laboratory-confirmed COVID-19 patient from five medical centers in Santiago, Chile. Participants responded to a brief questionnaire of COVID-19 related signs and symptoms (e.g., fever, cough, rhinitis, breathing difficulty, muscular pain, cold-like symptoms) and had an RT-PCR test taken.

We estimate the pre-test probability by considering a Bayesian logistic regression model. The covariates and the way to be included into the model were selected by fitting a decision tree model to the data based on all available predictors. The results from the decision tree are shown in Figure S2.

Based on the previous analysis, we include fever, cold-like symptoms, cough, gender, and their interactions in the model. The corresponding dummy variables, were included in the 16th dimensional vector, $x_i$. The Bayesian model is given by

$$Y_i \mid \beta \sim Bernoulli(\pi_i), \ i = 1, \ldots, 926,$$

$$\text{logit}(\pi_i) = x_i^T \beta,$$

$$\beta_k \sim N(0, 10^4), \ k = 0, 1, \ldots, 16,$$

We implemented the model using the JAGS program and the RJAGS library of the R statistical program. 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.
Figure S2. Classification tree to assess the most relevant covariates to estimate the pre-test probability of disease in a random sample of our clinical dataset of 926 individuals with suspected COVID-19 infection from five medical centers in Santiago, Chile. Participants responded to a brief questionnaire of COVID-19 related signs and symptoms (e.g., fever, cough, rhinitis, breathing difficulty, muscular pain, cold-like symptoms) and had an RT-PCR test taken.
Figure S3. Illustration of pre-test probability of SARS-CoV-2 for a person who reported having fever and cough and no other symptoms, and had the RT-PCR taken 2 days following the onset of fever. The tool also shows the posterior distribution for the sensitivity of the RT-PCR test for the SARS-CoV-2 virus, and the posterior distribution for the probability of being positive to the SARS-CoV-2 virus, given the negative result in one RT-PCR test.
References

1. 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. *Annals of Internal Medicine.* 2020;173(4):262-267.

2. Danis K, Epaulard O, Bénet T, et al. Cluster of coronavirus disease 2019 (Covid-19) in the French Alps, 2020. *Clinical Infectious Diseases.* 2020;71(15):825-832.

3. Guo L, Ren L, Yang S, et al. Profiling early humoral response to diagnose novel coronavirus disease (COVID-19). *Clinical Infectious Diseases.* 2020;71(15):778-785.

4. Kim ES, Chin BS, Kang CK, 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. *Journal of Korean medical science.* 2020;35(13):e142.

5. Kujawski SA, Wong KK, Collins JP, et al. Clinical and virologic characteristics of the first 12 patients with coronavirus disease 2019 (COVID-19) in the United States. *Nature Medicine.* 2020;26:861-868.

6. Liu L, Liu W, Zheng Y, et al. A preliminary study on serological assay for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in 238 admitted hospital patients. *Microbes and Infection.* 2020;22(4-5):206-2011.

7. Wölfel R, Corman VM, Guggemos W, et al. Virological assessment of hospitalized patients with COVID-2019. *Nature.* 2020;581(7809):465-469.

8. Zhao J, Yuan Q, Wang H, et al. Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. *Clinical Infectious Diseases.* 2020;71(16):2027-2034.

9. *R: A language and environment for statistical computing* [computer program]. Vienna: R Foundation for Statistical Computing; 2019.
10. Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Paper presented at: Proceedings of the 3rd international workshop on distributed statistical computing 2003.

11. Plummer M, Stukalov A, Denwood M. rJAGS: Bayesian Graphical Models Using MCMC. 2019; https://cran.r-project.org/web/packages/rjags/index.html.