# Variation in the Probability of SARS-CoV-2 Detection Using RT-qPCR Testing Prior to and After Symptom Onset

Scott Sherrill-Mix

## Abstract

Effective RT-qPCR testing for SARS-CoV-2 is essential for treatment, surveillance and control. A recent meta-analysis[1] suggested that testing prior to the onset of symptoms is likely to miss the majority of infected individuals. Since many community transmissions are likely to occur prior to or near the onset of symptoms[2,3], these findings cast doubt on screening efforts intended to enable early SARS-CoV-2 detection. However, additional data and alternative analyses refine these estimates and suggest that the hope of detecting SARS-CoV-2 prior to symptom onset is not as grim as suggested.

## Main

In order to more adequately account for uncertainty in time periods with little data, we developed a Bayesian autoregressive moving average state space model (B-ARMA-SSM) and used Markov chain Monte Carlo sampling to estimate posterior probabilities for parameters of interest from the data collected by Kucirka et al.[1] The detection rates estimated for the day of and days following symptom onset by the B-ARMA-SSM and the polynomial model of Kucirka et al.[1] closely resembled each other (Figure 1). Both models estimate a peak in detection rates 3–4 days after symptom onset followed by a progressive decline in the probability of a positive detection.

However, the models markedly differ in their estimates of the probability of detection prior to symptoms. Unfortunately, only three tests from a single patient were available prior to the onset of symptoms in this data hindering any inferences in this critical presymptomatic period (Figure 1). In spite of this sparseness, the polynomial model produces precise estimates prior to the onset of symptoms with little inferred potential for error. In contrast, the B-ARMA-SSM estimates higher probabilities of presymptomatic detection while displaying a much larger amount of uncertainty in its estimates (Figure 1). For example, at 4 days prior to symptom onset the B-ARMA-SSM estimates a detection rate of 35% (95% credible interval: 6–76%) while the polynomial model estimates a detection rate of 0% (95% CrI: 0–0%). Of course, without further data it is unclear which of these estimates is more likely to reflect reality.

Luckily, additional data has become available allowing a test of the predictive powers of these models. We identified 7 additional studies[2–8] by following the PubMed search strategy described previously[1]. These new data contained 381 results from 124 patients, including 21 patients measured prior to symptom onset. Combined with the previous seven studies, these data contain the results of 1619 RT-qPCR tests.

Fitting the B-ARMA-SSM to this combined data produces an updated estimate with remarkable similarity to that predicted by it without the additional presymptomatic data and well within the uncertainty estimated by the original model (Figure 2). In contrast, the new data does not seem to agree with the predictions made by the polynomial model. For example, the polynomial model predicts a 0% probability of detection earlier than 3 days before symptom onset yet in the new data 8 positives were detected out of 13 tests administered between 4–7 days prior to symptom onset.

Overall, the Bayesian B-ARMA-SSM with the combined data estimates RT-qPCR detection rates average 80% (95% CrI: 65–89%) on the day of symptom onset, increase to 86% (95% CrI: 76-92%) at 3 days after symptoms and then fall steadily as time passes from symptom onset. Similarly, detection rate drops prior to the onset of symptoms. For example, at 4 days prior to symptom onset, the detection rate is estimated at 56% (95% CrI: 30-79%).

Note that these estimates reflect the mean detection rate estimated over the various studies. Some studies have significantly higher or lower rates, potentially reflecting variability in techniques and patient populations. For example, the highest predicted detection rate at day 4 in any study was predicted at 97% (95% CrI: 94–95%) in van Kampen et al. [7]

The rapid development of the COVID-19 pandemic and research into SARS-CoV-2 means that we must rapidly update our assumptions and understanding with new data. Here we show that RT-qPCR detection of SARS-CoV-2 prior to the onset of symptoms is more likely than previously estimated.

The relatively high false negative rates in SARS-CoV-2 RT-qPCR testing predicted here and previously [1] is worrisome. Of course, what is a "false negative" depends greatly on the purpose of testing. If tests are intended to provide documentation of previous SARS-CoV-2 infection then these do indeed represent false negatives. However with the rapid development of serological testing, the more likely use for RT-qPCR testing is to detect positive cases prior to the onset of symptoms and to monitor viral clearance after symptom onset. In these cases, many "false negatives" are likely to be due to the reduction viral load to undetectable, and potentially less transmissible [2,7,9,10], levels rather than a failure of PCR testing. Thus at later time points, the false negative rate is likely overestimated. A similar problem arises in the interpretation of rates prior to symptom onset. Are patients testing negative because of a missed test or because they have not yet been infected or reached significant levels of viral load?

In interpreting this data, it is important to note that these analyses ignore censoring in the patient data. If a patient tests negative several times in a row then their doctor will often not continue to test them. This means that these data are likely to be biased towards patients with prolonged disease. Thus the estimates of detection rate are more likely to be overestimates further from symptom onset.

The detection of SARS-CoV-2 infection prior to the onset of symptoms is critical to public health and patient care. This preliminary analysis suggests that detection prior to the onset of symptoms is indeed possible although the potential for false negatives is certainly concerning. Further data from prospective sampling of at risk patients is essential to further characterize this critical period.

# Methods

All data was collected from patients who developed symptoms, were tested multiple times and tested positive at least once. Where raw data were unavailable, data were digitized from published plots. Data from Arons et al. [2] was provided by personal communication. Only nasopharyngeal swabs were used from Xiao et al. [4] and tests detecting $\geq 1$ PCR target were counted as positives. Only upper respiratory tract samples were used from van Kampen et al. [7] The data from a preprint by Kujawski et al. [11] was replaced by data from the final publication by COVID-19 Investigation Team [12]

The Bayesian autoregressive moving average state space model (B-ARMA-SSM) estimates the underlying probabilities of RT-PCR detection as a binomial probability $p_{t,j}$ at each time point $t$ days after symptom onset for each study $j$. This probability is not directly observed but inferred from the counts of positive and negative tests. The model assumes the detection probability on a given day $t$ should be similar to the probability on the previous day $t-1$ and similarly change between day $t$ and day $t-1$ should be similar to the change between day $t-1$ and $t-2$. To anchor the model, detection at 20 days prior to symptom onset is given a low prior probability. Differential detection rates in the individual studies are modeled as a constant multiplicative factor drawn from a normal distribution of potential study offsets.

$$x_{j,t} \sim \text{Binomial}(n_{j,t}, p_{j,t})$$

$$\text{logit}(p_{j,t}) = \beta_j + \mu_t$$

$$\mu_{-20} \sim \text{Normal}(-8, 4)$$

$$\mu_t = \mu_{t-1} + \epsilon_t \text{ if } t > -20$$

$$\epsilon_t \sim \begin{cases} \text{Normal}(0, \sigma) & \text{if } t = -19 \\ \text{Normal}(\epsilon_{t-1}, \sigma) & \text{if } t > -19 \end{cases}$$

$$\beta_j \sim \text{Normal}(0, \psi)$$

$$\sigma \sim \text{Gamma}(1, 1)$$

$$\psi \sim \text{Gamma}(1, 1)$$

where $x_{j,t}$ is the number of positives out $n_{j,t}$ total tests from study $j$ at $t$ days after symptom onset. The posterior probabilities of the Bayesian model were estimated using Markov chain Monte Carlo sampling as implemented in Stan[13] and analysis and plotting were performed in R[14].

## Data availability

All data and code is available at https://github.com/sherrillmix/covidRTPCR and will be archived at Zenodo prior to publication.

## References


[1] Lauren M. Kucirka, Stephen A. Lauer, Oliver Laeyendecker, Denali Boon, and Justin Lessler. Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based SARS-CoV-2 tests by time since exposure. *Ann Intern Med*, May 2020. ISSN 1539-3704. doi: 10.7326/M20-1495.

[2] Melissa M. Arons, Kelly M. Hatfield, Sujan C. Reddy, Anne Kimball, Allison James, Jesica R. Jacobs, Joanne Taylor, Kevin Spicer, Ana C. Bardossy, Lisa P. Oakley, Sukarma Tanwar, Jonathan W. Dyal, Josh Harney, Zeshan Chisty, Jeneita M. Bell, Mark Methner, Prabasaj Paul, Christina M. Carlson, Heather P. McLaughlin, Natalie Thornburg, Suxiang Tong, Azaibi Tamin, Ying Tao, Anna Uehara, Jennifer Harcourt, Shauna Clark, Claire Brostrom-Smith, Libby C. Page, Meagan Kay, James Lewis, Patty Montgomery, Nimalie D. Stone, Thomas A. Clark, Margaret A. Honein, Jeffrey S. Duchin, and John A. Jernigan. Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. *N Engl J Med*, April 2020. ISSN 1533-4406. doi: 10.1056/NEJMoa2008457.

[3] Seong Eun Kim, Hae Seong Jeong, Yohan Yu, Sung Un Shin, Soosung Kim, Tae Hoon Oh, Uh Jin Kim, Seung-Ji Kang, Hee-Chang Jang, Sook-In Jung, and Kyung-Hwa Park. Viral kinetics of SARS-CoV-2 in asymptomatic carriers and presymptomatic patients. *Int J Infect Dis*, 95:441–443, May 2020. ISSN 1878-3511. doi: 10.1016/j.ijid.2020.04.083.

[4] Ai Tang Xiao, Yi Xin Tong, and Sheng Zhang. Profile of RT-PCR for SARS-CoV-2: a preliminary study from 56 COVID-19 patients. *Clin Infect Dis*, April 2020. ISSN 1537-6591. doi: 10.1093/cid/ciaa460.

[5] Jonathan Altamirano, Prasanthi Govindarajan, Andra L. Blomkalns, Lauren E. Kushner, Bryan Andrew Stevens, Benjamin A. Pinsky, and Yvonne Maldonado. Assessment of sensitivity and specificity of patient-collected lower nasal specimens for sudden acute respiratory syndrome coronavirus 2 testing. *JAMA network open*, 3:e2012005, June 2020. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2020.12005.

[6] Ivan Fan-Ngai Hung, Vincent Chi-Chung Cheng, Xin Li, Anthony Raymond Tam, Derek Ling-Lung Hung, Kelvin Hei-Yeung Chiu, Cyril Chik-Yan Yip, Jian-Piao Cai, Deborah Tip-Yin Ho, Shuk-Ching Wong, Sally Sau-Man Leung, Man-Yee Chu, Milky Oi-Yan Tang, Jonathan Hon-Kwan Chen, Rosana Wing-Shan Poon, Agnes Yim-Fong Fung, Ricky Ruiqi Zhang, Erica Yuen-Wing Yan, Lin-Lei Chen, Charlotte Yee-Ki Choi, Kit-Hang Leung, Tom Wai-Hin Chung, Sonia Hiu-Yin Lam, Tina Poy-Wing Lam, Jasper Fuk-Woo Chan, Kwok-Hung Chan, Tak-Chiu Wu, Pak-Leung Ho, Johnny Wai-Man Chan, Chak-Sing Lau, Kelvin Kai-Wang To, and Kwok-Yung Yuen. Sars-cov-2 shedding and seroconversion among passengers quarantined after disembarking a cruise ship: a case series. *The Lancet. Infectious diseases*, June 2020. ISSN 1474-4457. doi: 10.1016/S1473-3099(20)30364-9.

[7] Jeroen JA van Kampen, David AMC van de Vijver, Pieter LA Fraaij, Bart L Haagmans, Mart M Lamers, Nisreen Okba, Johannes PC van den Akker, Henrik Endeman, Diederik AMPJ Gommers, Jan J Cornelissen, et al. Shedding of infectious virus in hospitalized patients with coronavirus disease-2019 (covid-19): duration and key determinants. *medRxiv*, 2020. doi: 10.1101/2020.06.08.20125310.

[8] Wang-Da Liu, Sui-Yuan Chang, Jann-Tay Wang, Ming-Jui Tsai, Chien-Ching Hung, Chia-Lin Hsu, and Shan-Chwen Chang. Prolonged virus shedding even after seroconversion in a patient with covid-19. *J Infect*, 2020. doi: 10.1016/j.jinf.2020.03.063.

[9] Jared Bullard, Kerry Dust, Duane Funk, James E. Strong, David Alexander, Lauren Garnett, Carl Boodman, Alexander Bello, Adam Hedley, Zachary Schiffman, Kaylie Doan, Nathalie Bastien, Yan Li, Paul G. Van Caeseele, and Guillaume Poliquin. Predicting infectious sars-cov-2 from diagnostic samples. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, May 2020. ISSN 1537-6591. doi: 10.1093/cid/ciaa638.

[10] Sin Fun Sia, Li-Meng Yan, Alex WH Chin, Kevin Fung, Ka-Tim Choy, Alvina YL Wong, Prathanporn Kaewpreedee, Ranawaka APM Perera, Leo LM Poon, John M Nicholls, et al. Pathogenesis and transmission of sars-cov-2 in golden hamsters. *Nature*, pages 1–7, 2020. doi: 10.1038/s41586-020-2342-5.

[11] Stephanie A Kujawski, Karen K Wong, Jennifer P Collins, Lauren Epstein, Marie E Killerby, Claire M Midgley, Glen R Abedi, N Seema Ahmed, Olivia Almendares, Francisco N Alvarez, et al. First 12 patients with coronavirus disease 2019 (covid-19) in the united states. *MedRxiv*, 2020.

[12] COVID-19 Investigation Team. Clinical and virologic characteristics of the first 12 patients with coronavirus disease 2019 (COVID-19) in the United States. *Nat Med*, April 2020. ISSN 1546-170X. doi: 10.1038/s41591-020-0877-5.

[13] Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *J of Stat Softw*, 76(1), 2017. doi: 10.18637/jss.v076.i01.

[14] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL `https://www.R-project.org/`.

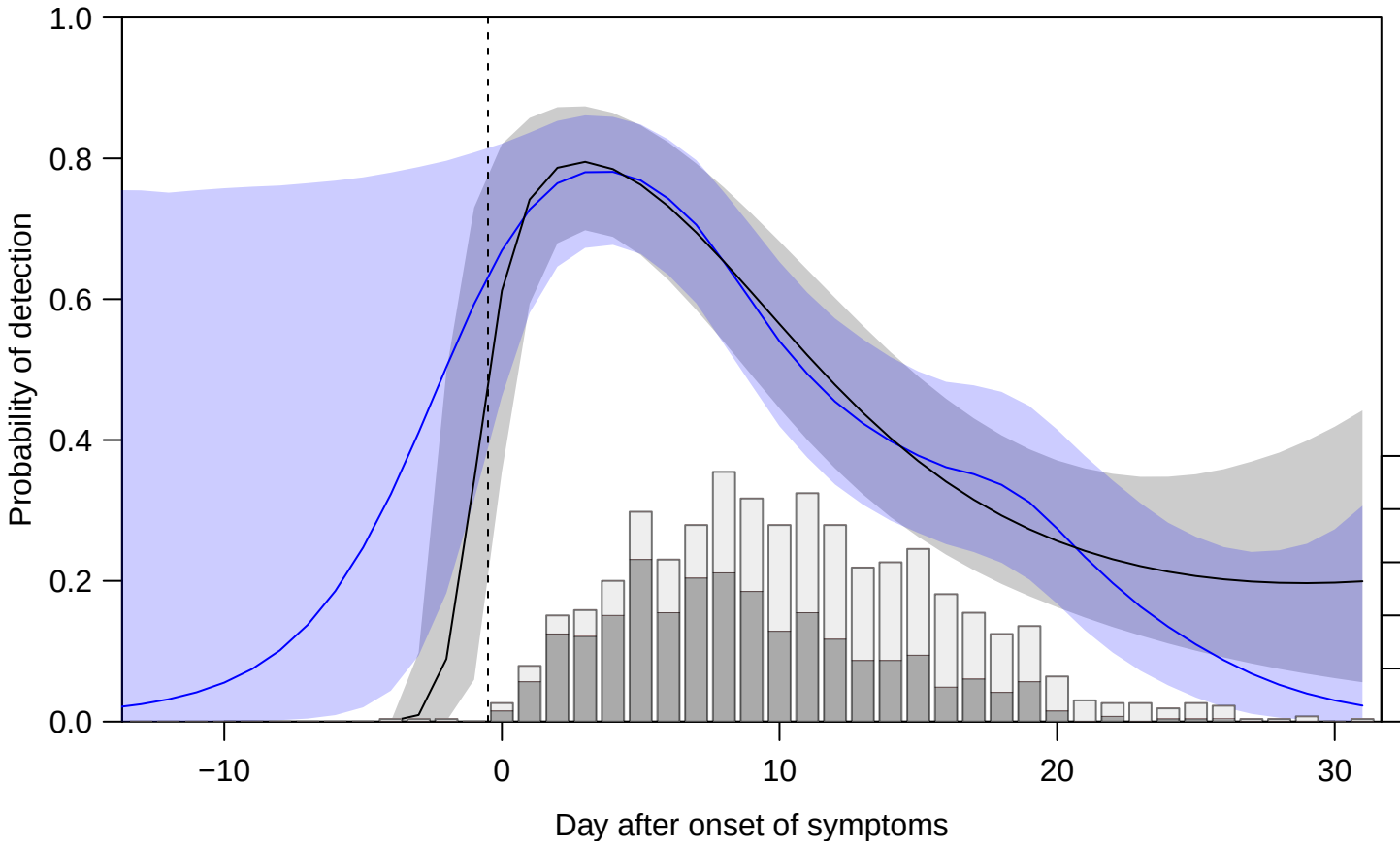


Figure 1: Comparison between polynomial and autoregressive moving average models. Bars show positive (dark) and negative (light) test results from previously published studies of RT-qPCR collected by Kucirka et al.[1]. Curves show estimates of SARS-CoV-2 detection probability estimated by polynomial regression (black) and an autoregressive moving average state space model (blue) and their 95% credible intervals (shading).

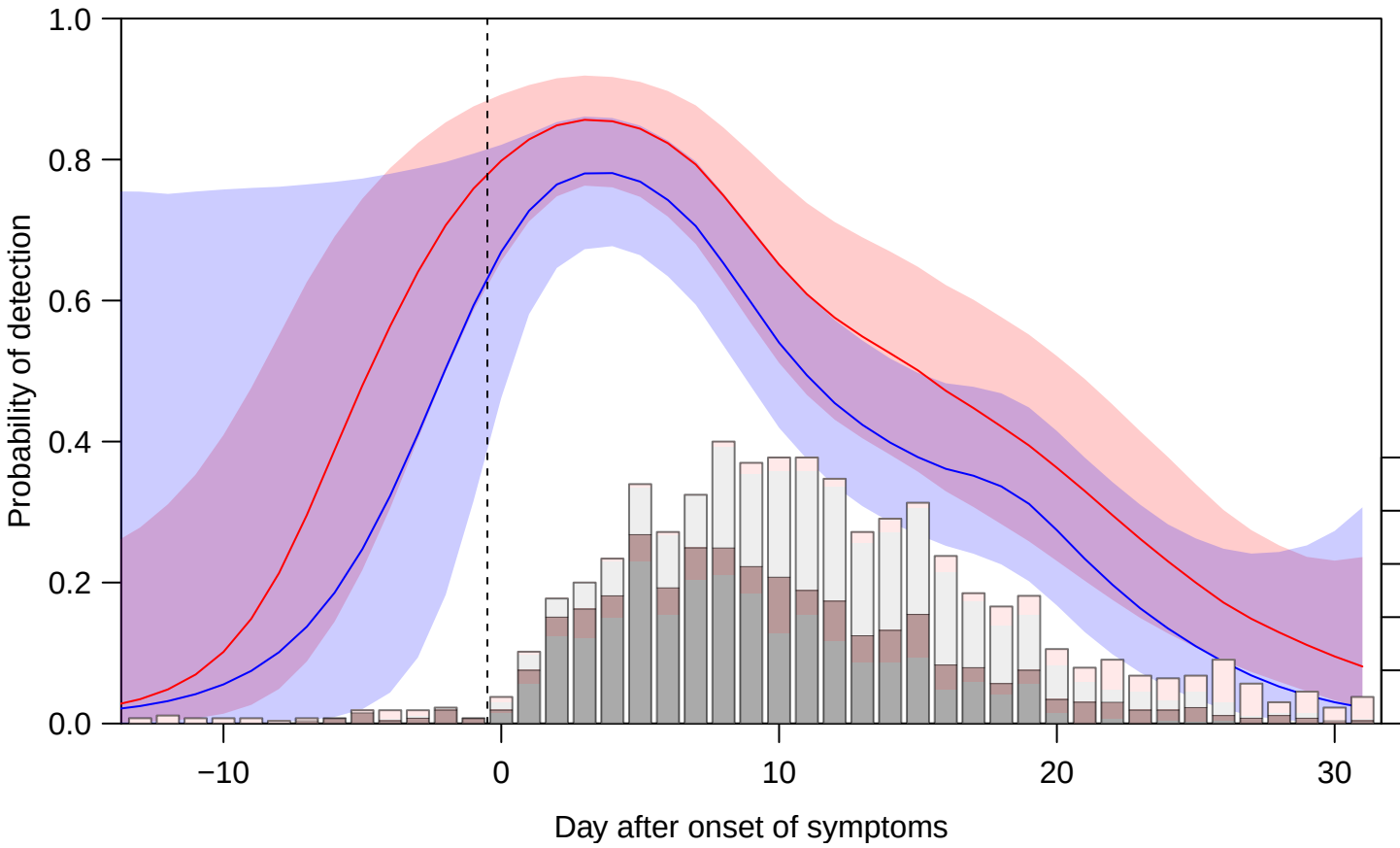


Figure 2: Comparison between the estimates of RT-qPCR detection probability before and after the addition of additional data. Curves show detection probability estimated by the autoregressive moving average state space model before (blue, as in Figure 1) and after (red) the addition of new data including tests from 21 patients prior to the onset of symptoms. Shading shows 95% credible intervals. Bars show positive (dark) and negative (light) test results from previously published studies of RT-qPCR collected by Kucirka et al. [1] (grey) and new data (pink).