An Update on the False-Negative Rates of Reverse Transcriptase Polymerase Chain Reaction-Based SARS-CoV-2 Tests

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 near or 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 finding cast doubt on screening efforts intended to enable early SARS-CoV-2 detection. However, new 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

Several studies have reported longitudinal measurements of SARS-CoV-2 RT-qPCR detection in COVID-19 patients[4–10]. These data were collected and analyzed by Kucirka et al.[1] using a Bayesian analysis to fit a polynomial logistic regression of detection rate. Unfortunately, polynomial regression without regularization is sensitive to overfitting and extrapolation to periods with sparse data is often problematic. Thus as an alternative framework, we implemented a method replacing the polynomial assumptions with a moving average autoregressive state space model (ARMA-SSM), where the probability of detection and its rate of change is simply assumed to be correlated from day to day rather that forming any specific linear pattern. This model formulation avoids giving undue influence to sparse data and more adequately quantifies uncertainty in these regions. In addition, we focus on detection rate relative to the onset of symptoms rather than assuming a fixed 5 day delay from infection to symptom onset, an assumption which recent research has suggested is unlikely to reflect real world conditions[2].

The detection rates estimated by the ARMA-SSM and polynomial model for the day of and days following symptom onset 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 a single patient tested prior to the onset of symptoms was available in this data hindering any inferences in this critical presymptomatic period (Figure 1). In spite of this, the polynomial model produces precise estimates prior to the onset of symptoms with little inferred potential for error. In contrast, the ARMA-SSM model 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 ARMA-SSM model 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 4 additional studies[2,3,10,11] by following the PubMed search strategy described previously[1]. These new data contained 407 tests from 95 patients who tested positive for SARS-CoV-2, including 21 patients measured prior to symptom onset. Combined with the previous seven studies, the data contain the results of 1601 tests.

Fitting the ARMA-SSM model 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 ARMA-SSM model with combined data estimates RT-qPCR detection rates average 74% (95% CrI: 61–84%) on the day of symptom onset, increase to 81% (95% CrI: 72-87%) at 4 days after symptoms and then fall steadily as time passes from symptom onset. Similarly, detection rate drops as time moves earlier from symptom onset. For example, at 4 days prior to symptom onset, the detection rate is estimated at 51% (95% CrI:A 29-72%).

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 is predicted at 89% (95% CrI: 84–95%) in Xiao et al. [11].

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 levels (and potentially less transmissible levels [2,12]) 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

Where raw data was 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. [11] and tests detecting ≥ 1 PCR target were counted as positives.

The Bayesian autoregressive moving average state space model 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. 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 (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] Roman Wlfel, Victor M. Corman, Wolfgang Guggemos, Michael Seilmaier, Sabine Zange, Marcel A. Mller, Daniela Niemeyer, Terry C. Jones, Patrick Vollmar, Camilla Rothe, Michael Hoelscher, Tobias Bleicker, Sebastian Brnink, Julia Schneider, Rosina Ehmann, Katrin Zwirglmaier, Christian Drosten, and Clemens Wendtner. Virological assessment of hospitalized patients with covid-2019. *Nature*, 581: 465–469, May 2020. ISSN 1476-4687. doi: 10.1038/s41586-020-2196-x.

[5] Li Guo, Lili Ren, Siyuan Yang, Meng Xiao, De Chang, Fan Yang, Charles S. Dela Cruz, Yingying Wang, Chao Wu, Yan Xiao, Lulu Zhang, Lianlian Han, Shengyuan Dang, Yan Xu, Qiwen Yang, Shengyong Xu, Huadong Zhu, Yingchun Xu, Qi Jin, Lokesh Sharma, Linghang Wang, and Jianwei Wang. Profiling early humoral response to diagnose novel coronavirus disease (covid-19). *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, March 2020. ISSN 1537-6591. doi: 10.1093/cid/ciaa310.

[6] Kostas Danis, Olivier Epaulard, Thomas Bnet, Alexandre Gaymard, Sphora Campoy, Elisabeth Bothelo-Nevers, Maude Bouscambert-Duchamp, Guillaume Spaccaferri, Florence Ader, Alexandra Mailles, Zoubida Boudalaa, Violaine Tolsma, Julien Berra, Sophie Vaux, Emmanuel Forestier, Caroline Landelle, Erica Fougere, Alexandra Thabuis, Philippe Berthelot, Raphael Veil, Daniel Levy-Bruhl, Christian Chidiac, Bruno Lina, Bruno Coignard, Christine Saura, and Investigation Team. Cluster

of coronavirus disease 2019 (covid-19) in the french alps, 2020. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, April 2020. ISSN 1537-6591. doi: 10.1093/cid/ciaa424.

[7] Eu Suk Kim, Bum Sik Chin, Chang Kyung Kang, Nam Joong Kim, Yu Min Kang, Jae Phil Choi, Dong Hyun Oh, Jeong Han Kim, Boram Koh, Seong Eun Kim, Na Ra Yun, Jae Hoon Lee, Jin Yong Kim, Yeonjae Kim, Ji Hwan Bang, Kyoung Ho Song, Hong Bin Kim, Ki Hyun Chung, Myoung Don Oh, and Korea National Committee for Clinical Management of COVID-19. 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*, 35:e142, April 2020. ISSN 1598-6357. doi: 10.3346/jkms.2020.35.e142.

[8] Juanjuan Zhao, Quan Yuan, Haiyan Wang, Wei Liu, Xuejiao Liao, Yingying Su, Xin Wang, Jing Yuan, Tingdong Li, Jinxiu Li, Shen Qian, Congming Hong, Fuxiang Wang, Yingxia Liu, Zhaoqin Wang, Qing He, Zhiyong Li, Bin He, Tianying Zhang, Yang Fu, Shengxiang Ge, Lei Liu, Jun Zhang, Ningshao Xia, and Zheng Zhang. Antibody responses to sars-cov-2 in patients of novel coronavirus disease 2019. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, March 2020. ISSN 1537-6591. doi: 10.1093/cid/ciaa344.

[9] Lei Liu, Wanbing Liu, Yaqiong Zheng, Xiaojing Jiang, Guomei Kou, Jinya Ding, Qiongshu Wang, Qianchuan Huang, Yinjuan Ding, Wenxu Ni, Wanlei Wu, Shi Tang, Li Tan, Zhenhong Hu, Weitian Xu, Yong Zhang, Bo Zhang, Zhongzhi Tang, Xinhua Zhang, Honghua Li, Zhiguo Rao, Hui Jiang, Xingfeng Ren, Shengdian Wang, and Shangen Zheng. A preliminary study on serological assay for severe acute respiratory syndrome coronavirus 2 (sars-cov-2) in 238 admitted hospital patients. *Microbes and infection*, 22:206–211, 2020. ISSN 1769-714X. doi: 10.1016/j.micinf.2020.05.008.

[10] 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.

[11] 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.

[12] 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.

[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/`.

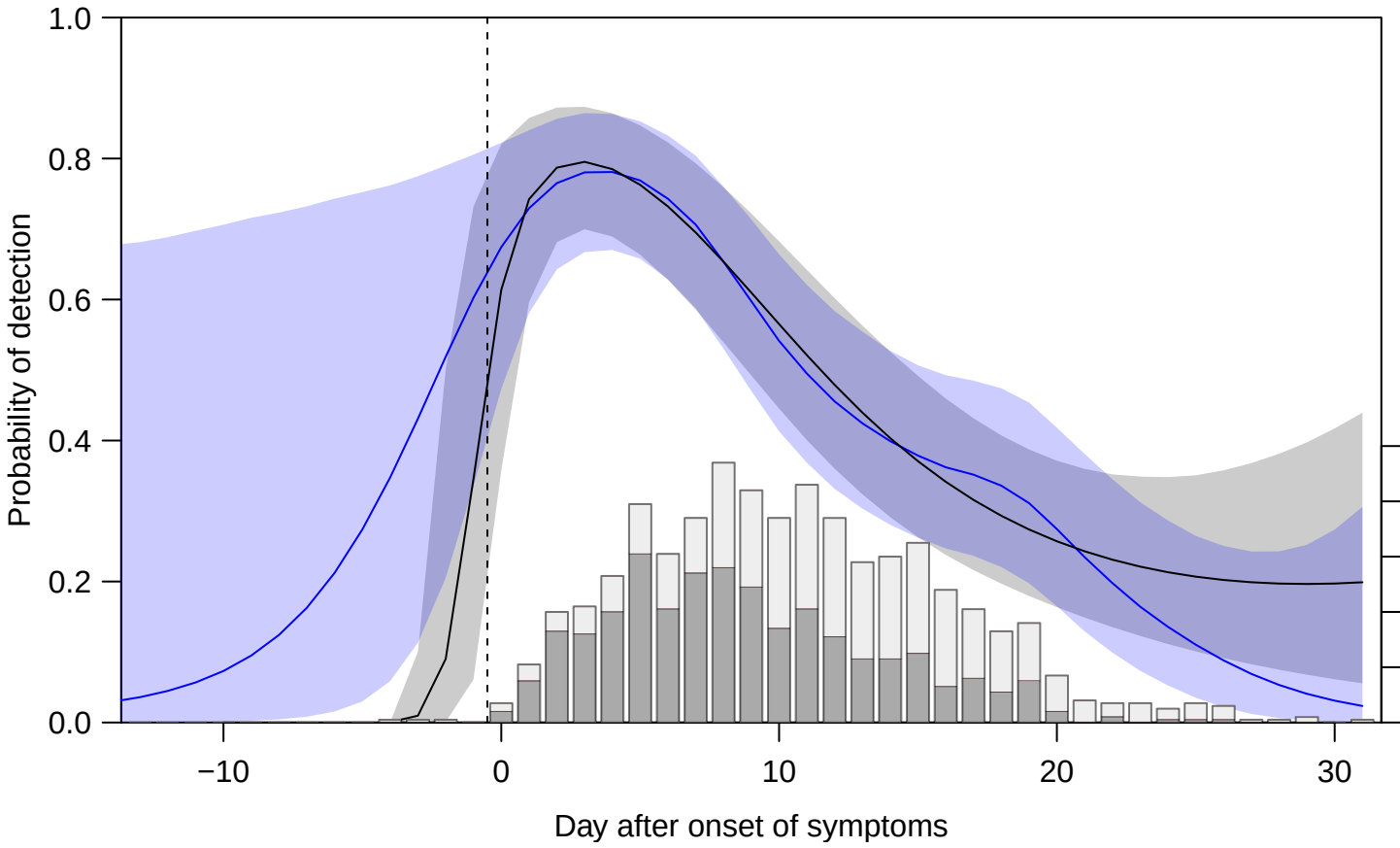

Figure 1: Comparison between polynomial and autoregressive moving average models. Bars show positive (dark) and negative (light) test results from previously published studies of RT-qPCR collected by Kucirka et al.[1]. Curves show estimates of SARS-CoV-2 detection probability estimated by polynomial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

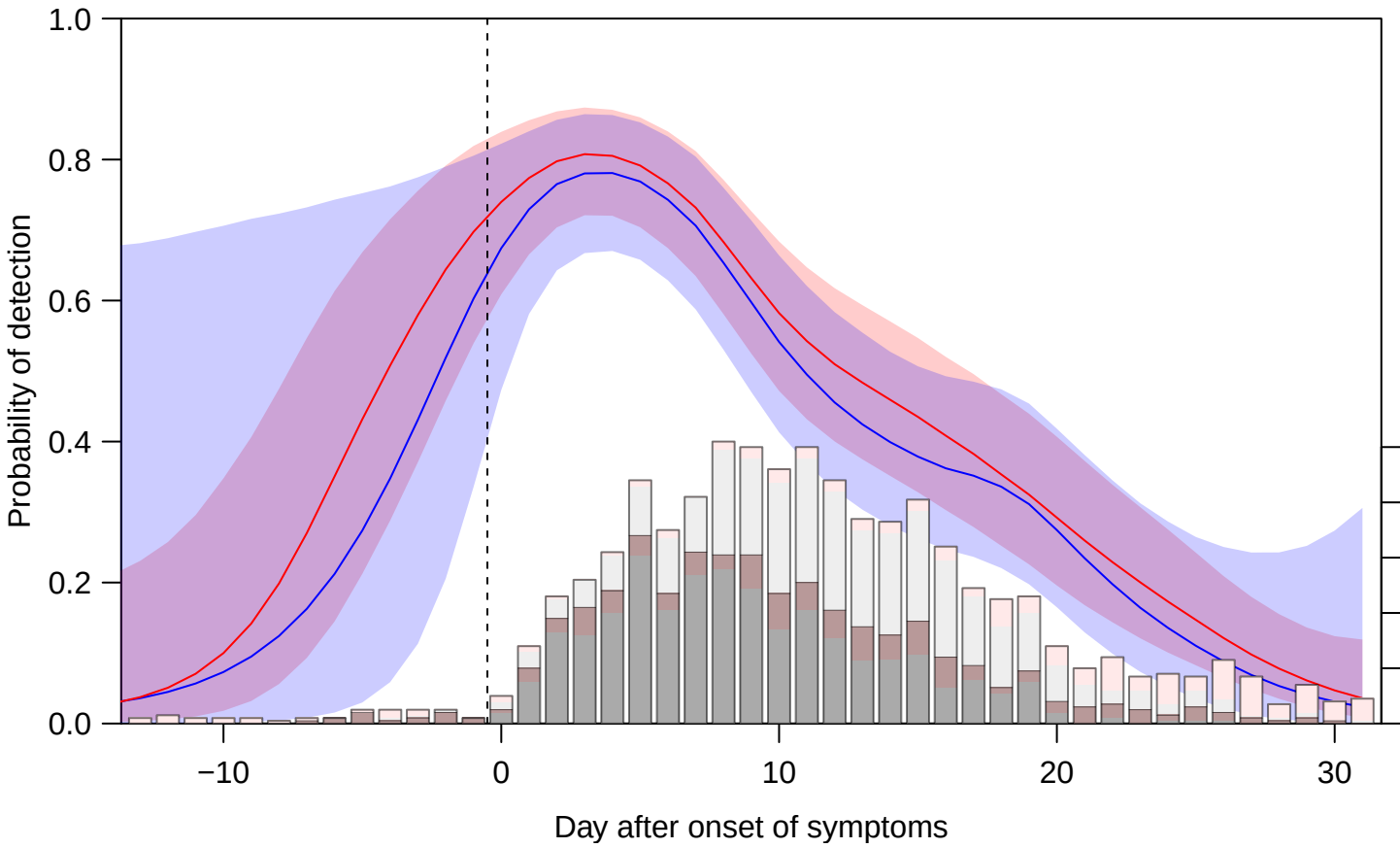

Figure 2: Comparison between the estimates of RT-qPCR detection probability before and after the addition of additional data. Curves show detection probability estimated by the autoregressive moving average state space model before (blue, as in Figure 1) and after (red) the addition of new data including 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).