medRxiv preprint doi: https://doi.org/10.1101/2020.04.26.20080911; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

False positives in reverse transcription PCR testing for SARS-CoV-2

Andrew N. Cohen[1]*, Bruce Kessel[2]

[1] Center for Research on Aquatic Bioinvasions, Richmond CA, USA.
[2] John A. Burns School of Medicine, University of Hawai'i, Honolulu HI, USA.
* Correspondence: acohen@bioinvasions.com

Abstract

Large-scale testing for SARS-CoV-2 by the reverse transcription polymerase chain reaction is a key part of the response to the COVID-19 pandemic, but little attention has been paid to the potential frequency and impacts of false positive results. In the absence of data on the clinical specificity of SARS-CoV-2 assays, we estimate a conservative false positive rate from external quality assessments of similar viral assays, and show that this rate may have large impacts on the reliability of positive test results. This has clinical and case management implications, affects an array of epidemiological statistics, and should inform the scale of testing and the allocation of testing resources. Measures to raise awareness of false positives, reduce their frequency, and mitigate their effects should be considered.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.26.20080911; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Testing for SARS-CoV-2 employs the reverse transcription polymerase chain reaction (RT-PCR) to detect diagnostic sequences in the virus' RNA genome. The results are used to diagnose and determine the treatment of patients, to sequester individuals, to allocate the use of personal protective equipment, and to direct contact tracing. At the policy level test results are used to model rates of spread, calculate hospitalization and death rates, assess the number and significance of asymptomatic carriers, shape public policy on social distancing, and inform decisions about the spatial and temporal allocation of medical resources. Test accuracy is thus of paramount importance, yet little attention has been paid to the potential frequency and impacts of false positive results (Supplementary Text). Although SARS-CoV-2 RT-PCR assays are widely reported to have 100% specificity (Table S1), this refers to the absence of cross-reactivity with non-target genetic material (analytical specificity), not to the potential for incorrect results in the real-world implementation of testing (clinical specificity) where contamination or human error can generate false positives during sample collection, transport or analysis.

External quality assessments (EQAs) test the implementation of medical diagnostic assays by providing participating laboratories with blind panels of positive and negative samples. The laboratories assay these panels using their normal procedures and report the results to the EQA manager, who compiles and analyzes the results. Since external quality assessments have not yet been performed on SARS-CoV-2 assays, and in the absence of any other data on the clinical specificity of these assays, we conducted a meta-analysis to determine the range of false positive rates (FPRs) in EQAs of similar assays (see Methods). We compiled data and calculated FPRs for 43 EQAs of RT-PCR assays of RNA viruses, conducted in 2004-2019. Each EQA involved between three and 174 laboratories, which together provided results for 4,113 blind panels containing 10,538 negative samples (Table 1). Though the reported information on the laboratories' methods was incomplete, it appears that 99.8% of the panels were assayed with RT-PCR and 0.2% with other nucleic acid amplification-based methods (RT-loop-mediated isothermal amplification or RT-reverse polymerase amplification). In EQAs where the information was provided, 17% of the RT-PCR assays were conventional and 83% were real-time.

In all, 336 of the 10,538 negative samples (3.2%) were reported as positive. We considered two data sets comprising all 43 EQAs (full data set), and the 37 EQAs that analyzed at least 100 negative samples (subset). FPRs in each EQA ranged from 0-16.7% for the full data set, and 0-8.1% for the subset. There was no correlation between FPR and Year for the full data set (n=43, r=0.147, p=0.346), and a weak downward correlation for the subset (n=37, r=0.327, p=0.056) (Figure S1). The median and interquartile range were lower for the full data set (median=2.3%, interquartile range=0.8-4.0%) than for the subset (median=2.5%, interquartile range=1.2-4.0%) (Figure S2).

We conservatively used the lower of the 25th percentile FPRs from the two data sets in our models. This FPR value (0.8%) is further conservative in that it does not include false positives produced during sampling (*1*), or account for any increased error rate stemming from the rapid expansion of SARS-CoV-2 testing and the use of novel diagnostic assays (*2*). Published analyses of SARS-CoV-2 testing indicate a range of FNRs from 0% to 52.2% (Table S2). We used an FNR of 25% as input to the model; sensitivity analyses performed over the range of 0-50% showed little effect on the reliability of positive results (Figures S2-S6).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.26.20080911; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Even a low FPR reduces the reliability of positive results when prevalence is low. This is apparent in model results based on test data for 50 U.S. states and 77 countries (Figures 1A, 2A), which suggest that some testing is targeted too broadly to be useful in regions with low test positivities (toward the right side of the panels). However, some elements of a testing program may be of value even if the average reliability of positive results in a region is extremely low. Modeling disaggregated data (for example, separately modeling data for diagnostic, screening and surveillance testing) could guide the reallocation of test resources from program elements whose positive results are too unreliable to be useful into more focused and reliable testing.

Modeling daily test data reveals the trajectories of test program reliability (Figures 1B-D, 2B-C). Some regions continue to test broadly even as test positivity drops to low levels and the modeled reliability of positive results approaches or reaches zero. In Montana, for example, the model suggests that after around April 25th most positive results, especially for asymptomatic individuals, were probably false positives, but an average of nearly 3,000 tests continued to be conducted in the state each day. South Korea arrived at a similar point by April 7 but continued to conduct over 6,000 tests a day.

The reliability of positive results dropped to near zero in these cases when test positivity approached the estimated FPR. However, even at positivities up to around four times the FPR, over 20% of positive results are likely to be false positives. Unless other respiratory diseases are pervasive in a test population, most of these false positive individuals would likely be asymptomatic, which could at least partially explain reports of large numbers of asymptomatic carriers of SARS-CoV-2.

False positives also affect the interpretation of individual test results. Statements from health agencies and officials suggest that positive results from SARS-CoV-2 tests are more trustworthy than negative results (*3*, *4*). However, over a wide range of likely scenarios, the opposite is true (see Figures 1, 2, S3-S6, wherever positive predictive value is less than negative predictive value). The FPR acts on samples from the uninfected fraction of the population, returning positive results for some uninfected individuals, while the FNR acts on the infected fraction, returning negative results for some infected individuals. When prevalence is low, the uninfected fraction is much greater than the infected fraction, so that even a low FPR can have a larger effect than a high FNR.

The false positives reported in the EQAs were probably due to contamination either from a positive sample analyzed at the same time (cross-contamination), or more likely from genes amplified from prior positive samples or positive controls (carryover contamination). The amplification of nucleic acids makes PCR-based assays highly sensitive, but also highly vulnerable to minute levels of sample contamination which can produce false positives that are indistinguishable from true positives. False positives can also be produced by sample mix-ups (*5*) or data entry errors.

Some non-EQA studies have reported false positives in RT-PCR assays for other coronaviruses (*6*, *7*), and there is some evidence for false positives specifically in SARS-CoV-2 assays. Four studies conducting sensitivity or cross-reactivity assessments on SARS-CoV-2 RT-PCR assays

medRxiv preprint doi: https://doi.org/10.1101/2020.04.26.20080911; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

reported false positives when negative samples were tested, presumably due to contamination in those laboratories (Tables S3, S4). Several cases of false positives in regular SARS-CoV-2 testing have been reported in the media (Supplementary Text).

Though numerous studies have addressed FNRs in SARS-CoV-2 RT-PCR testing (Table S2), we found virtually no discussion of FPRs in these tests in the scientific or medical literature (Supplementary Text). False positive results, especially when unanticipated, have clinical and case management consequences, including waste of personal protective equipment, waste of human resources in contact tracing, and unnecessary sequestering of uninfected individuals, sometimes along with infected individuals (*8*). A false positive test result could impede a correct diagnosis, delaying or depriving patients of appropriate treatment. False-positive patients introduce noise into clinical observations, which may hinder the development of improved COVID-19 medical care based on clinical experience. If antibody or antiviral treatments become available for COVID-19 patients, or prophylactic treatments for asymptomatic or mildly symptomatic individuals with positive test results, false-positive individuals could be subjected to medically inappropriate treatments *(9*). Individuals that have falsely tested positive may be less likely to avoid future exposure to infected individuals, believing they have immunity, and for the same reason may not seek vaccination when it becomes available. Clinical trials could lose statistical power by unwittingly enrolling false-positive individuals, who would be exposed to potentially harmful side effects without any mitigating potential for benefit. False positives also affect an array of epidemiological statistics, including the asymptomatic ratio.

The impact of false positives in SARS-CoV-2 testing could be mitigated by increasing awareness of the probability of false positives; by improving estimates of false positive rates with appropriately-designed external quality assessments, or assessing results retrospectively with serological tests; and by reducing the frequency of false positives by requiring two independent positive tests to classify an individual as positive. In previous disease outbreaks the World Health Organization and the U.S. Centers for Disease Control and Prevention warned about the potential for false positives in RT-PCR testing, restricted testing to individuals most likely to have the disease to avoid generating excessive false positives, and required confirmation of positive results by a second test (*10, 11*). These warnings and requirements are absent from these organizations' guidance on SARS-CoV-2 (*12, 13*). Perhaps they should be reinstated.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.26.20080911; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References

1 Qian Y, Zeng T, Wang H, Xu M, Chen J, Hu N, Chen D, Liu Y. Safety management of nasopharyngeal specimen collection from suspected cases of coronavirus disease 2019. Int J Nurs Sci 2020. DOl: 10.1016/j.ijnss.2020.03.012

2 Lippi G, Simundic AM, Plebani M. Potential preanalytical and analytical vulnerabilities in the laboratory diagnosis of coronavirus disease 2019 (COVID-19). Clin Chem Lab Med 2020 March 16 (Epub ahead of print).

3 Fact Sheet for Patients: CDC - 2019-nCoV Real-Time RT-PCR Diagnostic Panel. U.S. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/downloads/Factsheet-for-Patients-2019-nCoV.pdf (2020).

4 Johnson, C.Y. A 'negative' coronavirus test result doesn't always mean that you aren't infected. *Washington Post* https://www.washingtonpost.com/science/2020/03/26/negative-coronavirus-test-result-doesnt-always-mean-you-arent-infected (2020).

5 Lau KA, Theis T, Gray J, Rawlinson WD. Ebola preparedness: diagnosis improvement using rapid approaches for proficiency testing. J Clin Microbiol 2017;55:783-90.

6 Patrick DM, Petric M, Skowronski DM, Guasparini R, Booth TF, Krajden M, McGeer P, Bastien N, Gustafson L, Dubord J, MacDonald D. An outbreak of human coronavirus OC43 infection and serological cross-reactivity with SARS coronavirus. Canadian Journal of Infectious Diseases and Medical Microbiology. 2006;17(6):330-6.

7 Koetz A, Nilsson P, Lindén MV, Van Der Hoek L, Ripa T. Detection of human coronavirus NL63, human metapneumovirus and respiratory syncytial virus in children with respiratory tract infections in south-west Sweden. Clinical microbiology and infection. 2006 Nov 1;12(11):1089-96.

8 Wang Y, Wang Y, Chen Y, Qin Q. Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID-19) implicate special control measures. *J Med Virol.* 2020;92(6):568-576. doi: 10.1002/jmv.25748

9 Molloy PJ, Persing DH, Berardi VP. False-positive results of PCR testing for Lyme disease. *Clin Infect Dis.* 2001;33(3):412-413. doi: 10.1086/321911

10 World Health Organization. Use of laboratory methods for SARS diagnosis. https://www.who.int/csr/sars/labmethods/en/#lab

11 U.S. Centers for Disease Control and Prevention. Public Health Guidance for Community-Level Preparedness and Response to Severe Acute Respiratory Syndrome (SARS). Version 2. Supplement F: Laboratory Guidance. May 21, 2004.

12 World Health Organization. Laboratory testing for coronavirus disease (COVID-19) in suspected human cases. Interim guidance 19 March 2020.

13 U.S. Centers for Disease Control and Prevention. Evaluating and testing persons for Coronavirus Disease 2019 (COVID-19). Revised May 3, 2020. https://www.cdc.gov/coronavirus/2019-nCoV/hcp/clinical-criteria.html

Acknowledgements

We thank Michael Milgroom, Elisa Liberti, Dominic Chow and Thomas Taylor for their assistance. **Author contributions:** AC designed and conducted the analyses. AC and BK collaborated on the review of clinical implications and writing the report. **Competing interests:** The authors have no competing interests. **Data and materials availability:** All data is available in the manuscript or supplementary materials.

Supplementary Materials:

Methods
Supplementary Text
Figures S1-S6
Tables S1-S5
References

medRxiv preprint doi: https://doi.org/10.1101/2020.04.26.20080911; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

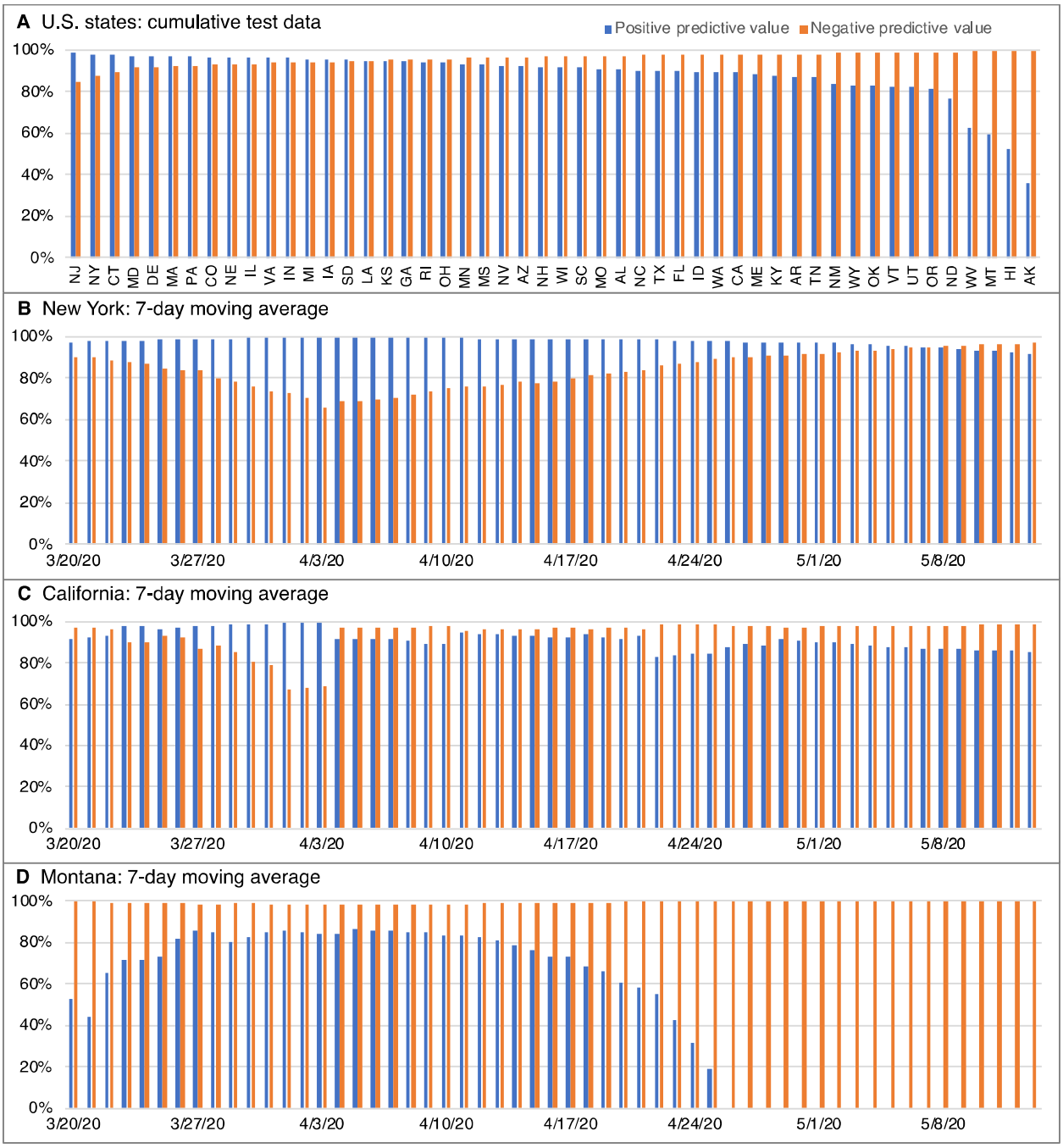

Figure 1. Reliability of SARS-CoV-2 test results in the United States. Positive predictive value (the probability that a positive result is true) and negative predictive value (the probability that a negative result is true) calculated with a false negative rate of 25% and a false positive rate of 0.8%. (A) Results for the 50 U.S. states based on cumulative test data through May 13, 2020. States arranged left to right in order of decreasing test positivity. (B-D) Reliability trajectories based on the previous-7-day moving average for states with positive test results with high reliability (New York), reliability starting to decline (California), and steeply declined reliability (Montana). Test data are from The COVID Tracking Project (https://covidtracking.com/about-data accessed May 14, 2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.26.20080911; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

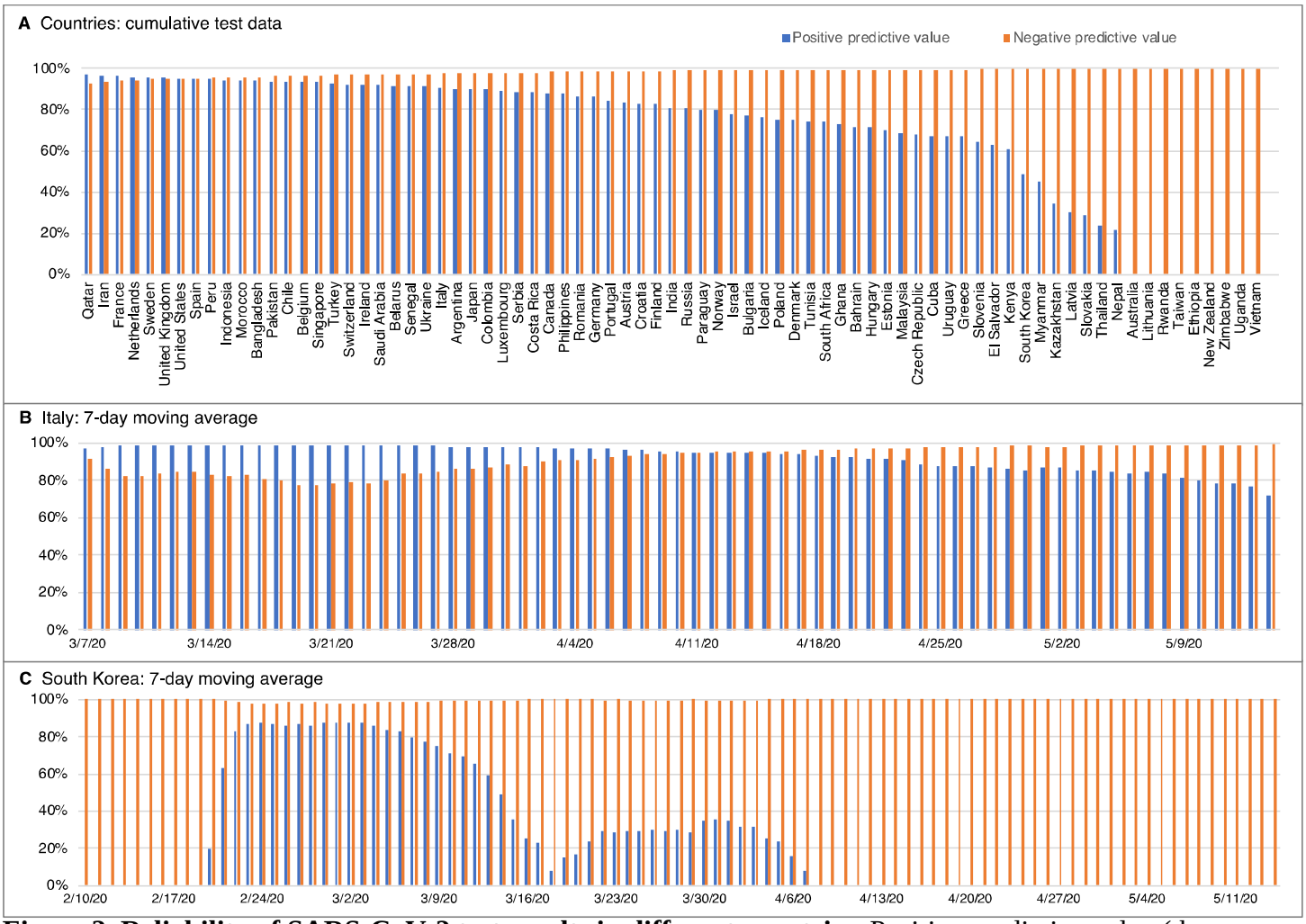

Figure 2. Reliability of SARS-CoV-2 test results in different countries. Positive predictive value (the probability that a positive result is true) and negative predictive value (the probability that a negative result is true) calculated with a false negative rate of 25% and a false positive rate of 0.8%. (A) Results for 77 countries based on cumulative test data through the most recent available date (between April 29 and May 14, 2020). Countries arranged left to right in order of decreasing test positivity. (B-C) Reliability trajectories based on the previous-7-day moving average for countries with positive test results with declining reliability (Italy) and sharply declined reliability (South Korea). Cumulative test data are from Our World in Data (https://github.com/owid/covid-19-data/tree/master/public/data/ accessed May 14, 2020). Daily test data are from the Italian Ministry of Health (http://www.salute.gov.it/portale/nuovocoronavirus/archivioNotizieNuovoCoronavirus.jsp?lingua=italiano&menu=notizie&p=dalministero&area=nuovocoronavirus&notizie.page=0 accessed May 14, 2020) and the South Korean Centers for Disease Control and Prevention (https://www.cdc.go.kr/board/board.es?mid=&bid=0030 accessed May 14, 2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.26.20080911; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 1. False positive rates in external quality assessments of RNA virus assays[a]

Virus	Number of EQAs	Dates of EQAs	Laboratories per EQA	Negative Samples per EQA	False Positive Rates[b]
SARS-CoV	1	2004?	58	174	2.3-6.9%
MERS-CoV	3	2014-17	49-99	49-1,134	<0.6%-1.0%
Influenza A viruses	17	2007-2019	64-174	114-332	<0.6-7.0%
Hepatitis C virus	8	2005-07	5-104	21-728	2.1-7.0%
Hepatitis Delta virus	1	2015?	28	112	5.4%
Chikungunya virus	2	2007, 2014	31-56	108-297	1.9-8.1%
Chikungunya, Dengue	1	2015	20	40	2.5%
Dengue virus	1	2013	16	16	6.3%
Zika virus	1	2016	50	504	2.8%
Rift Valley Fever virus	1	2012	30	117	3.4%
Measles virus	1	2014	41	123	0.8%
Ebola virus	5	2014-16	3-82	3-317	0.3-16.7%
4 arboviruses	1	2017	51	204	4.9%

[a] "<" indicates a false positive rate below the detection limit; treated as zero in the analyses.
[b] See Table S5 for full data and references.