Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

The Journal of Molecular Diagnostics, Vol. 23, No. 9, September 2021

the Journal of Molecular Diagnostics
jmdjournal.org

Prediction of False-Positive Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Molecular Results in a High-Throughput Open-Platform System

Ryan J. Martinez,* Nathan Pankratz,† Matthew Schomaker,‡ Jerry Daniel,§ Kenneth Beckman,§ Amy B. Karger,† Bharat Thyagarajan,† Patricia Ferreri,† Sophia L. Yohe,* and Andrew C. Nelson*

From the Division of Molecular Pathology and Genomics, Department of Laboratory Medicine and Pathology, the Department of Laboratory Medicine and Pathology,† School of Medicine, the M Health Fairview Molecular Diagnostics Laboratory,‡ and the University of Minnesota Genomics Center,§ University of Minnesota, Minneapolis, Minnesota*

Accepted for publication May 25, 2021.

Address correspondence to Andrew C. Nelson, M.D., Ph.D., 420 Delaware St. S.E., Mayo Mail Code 609, Minneapolis, MN 55455; or Sophia L. Yohe, M.D., 420 Delaware St. S.E., Mayo Mail Code 609, Minneapolis, MN 55455. E-mail: nels2055@umn.edu or yohe0001@umn.edu.

Widespread high-throughput testing for identification of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection by RT-PCR has been a foundation in the response to the coronavirus disease 2019 (COVID-19) pandemic. Quality assurance metrics for these RT-PCR tests are still evolving as testing is widely implemented. As testing increases, it is important to understand performance characteristics and the errors associated with these tests. Herein, we investigate a high-throughput, laboratory-developed SARS-CoV-2 RT-PCR assay to determine whether modeling can generate quality control metrics that identify false-positive (FP) results due to contamination. This study reviewed repeated clinical samples focusing on positive samples that test negative on re-extraction and PCR, likely representing false positives. To identify and predict false-positive samples, we constructed machine learning–derived models based on the extraction method used. These models identified variables associated with false-positive results across all methods, with sensitivities for predicting FP results ranging between 67% and 100%. Application of the models to all results predicted a total FP rate of 0.08% across all samples, or 2.3% of positive results, similar to reports for other RT-PCR tests for RNA viruses. These models can predict quality control parameters, enabling laboratories to generate decision trees that reduce interpretation errors, allow for automated reflex testing of samples with a high FP probability, improve workflow efficiency, and increase diagnostic accuracy for patient care. *(J Mol Diagn 2021, 23: 1085–1096; https://doi.org/10.1016/j.jmoldx.2021.05.015)*

The coronavirus disease 2019 (COVID-19) pandemic generated the need to rapidly implement high-throughput, widespread testing in the United States. The primary method for detecting severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the RNA virus responsible for COVID-19, is RT-PCR. RT-PCR is currently widely used and is the gold standard for the diagnosis of many infectious diseases.[1] RT-PCR has become the predominate diagnostic modality for viral disease as results are rapidly returned, it demonstrates high specificity and sensitivity, and it is relatively inexpensive.[2] However, as COVID-19 cases spread, it was quickly apparent that the need for testing outpaced health departments' and clinical laboratories' ability to provide testing. Thus, a multitude of RT-PCR and transcription-mediated amplification assays testing for SARS-CoV-2 have received emergency use authorization (EUA) from the US Food and Drug Administration (FDA) that includes both closed-platform and high-throughput, open-platform reactions (FDA, *https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/vitro-diagnostics-euas*,

S.L.Y. and A.C.N. contributed equally to this work as senior authors.
Disclosures: None declared.

https://doi.org/10.1016/j.jmoldx.2021.05.015

last accessed March 4, 2021). Indeed, approximately 336 million SARS-CoV-2 tests have been performed in the United States, with an increased need in current infectivity hot spots (CDC, *https://covid.cdc.gov/covid-data-tracker/#cases_totalcases*, last accessed March 4, 2021). Because of the rapid development and implementation of these assays, robust universal rules for interpretation and quality assurance of results have not been well defined in the clinical setting.

Guidelines for SARS-CoV-2 testing have been released by working groups and mainly focus on clinical scenarios of when to employ testing.[3] Many SARS-CoV-2 clinical tests receiving EUA only have interpretation guidelines for individuals under investigation, or those patients with an increased pretest probability of COVID-19 disease. Most of these same tests have not been validated as a screening technique for mildly symptomatic or asymptomatic patients, even though they are widely used in this manner. Testing of asymptomatic patients is a cornerstone for combating the spread of SARS-CoV-2 because transmission can be facilitated by patients with minimally or presymptomatic infections.[4] As these tests have not been fully studied in the setting of minimally symptomatic individuals, there is a potential for resultant errors. Attributing a positive SARS-CoV-2 result to an asymptomatic patient has an impact on the mental health, physical health, and socioeconomic well-being of that patient. A positive result also has wide-ranging impacts on infection control measures in the health care system and the community, such as isolation procedures in the hospital, closure of schools and daycares, and the halting of nursing home visits. In a low-prevalence population with low pretest probability, there is increased concern that a positive result is an error and represents a false-positive result. In addition, a false-positive SARS-CoV-2 result (FP; defined as a nonreproducible result on repeated extraction and RT-PCR) in a patient with significant symptoms due to other causes, such as a congestive heart failure patient with acutely worsening shortness of breath and cough, could lead to improper medical management. Thus, there is a need to understand error rates of SARS-CoV-2 assays to reduce the risk of false results. External quality assessments are traditionally performed on molecular assays by providing clinical laboratories with positive and negative samples and determining FP and false-negative rates from these blinded tests. False-negative results due to poor nasopharyngeal sampling or changes in anatomic viral replication have been identified and discussed elsewhere.[5] However, a recent study analyzing external quality assessments for RNA virus detection found FP rates ranged from 0% up to 16.7%, with a median of 2.3%.[6] Therefore, it is highly likely that a small portion of SARS-CoV-2 RT-PCR positive results are FPs and should be investigated.

As the need for SARS-CoV-2 testing was increasing, shortages of testing supplies developed. Thus, our clinical laboratory established a high-throughput, open-platform SARS-CoV-2 assay using the CDC developed primers and probes that could be easily adapted for multiple extraction methods.[7] To study FP results across multiple extraction systems, clinical samples with concern for contamination or those with low viral load, as defined by late C_T, underwent repeated extraction and testing from the primary sample. In general, FP results from PCR may be due to testing the wrong sample (due to mislabeling, sample mix-up, or reporting errors), cross-reactivity of the PCR assay, or contamination. Contamination is of particular concern in an open-platform assay when aerosolization from human or machine handling can cause the transfer of target genomic material between wells, especially when there are true-positive (TP; defined as reproducible results on repeated extraction and RT-PCR) samples that can exhibit as much as 10^6 range in virus quantity. Results of initial and repeated testing were compared to determine whether there were discordant results on repeat. Initially positive samples that were discordant were considered FP, whereas concordant samples were considered TP. Next, machine learning derived models were generated, taking into account the relative viral load of the index sample, the extraction method, and relative viral load of surrounding positive wells. These models were used as a clinical decision support tool to identify wells with a high probability of being an FP and to identify technical contributors to FP samples. Modeling identified FP samples that could be predicted across different extraction methods, with each model identifying different variables used to make FP predictions. From these models, we can identify potential technical improvements as well as identify samples with a high probability of FP results to act as an adjunct for laboratory clinical decision making and initiate automated sample re-extraction and repeat.

Materials and Methods

SARS-CoV-2 PCR Testing

Detection of SARS-CoV-2 RNA by RT-PCR has been detailed previously.[7] Individual RT-PCRs were performed in a 384-well plate using the US CDC-designed primers and probes specific for the N1 and N2 regions of the SARS-CoV-2 virus and human RNase P (RP). Nucleic acid extractions were performed using: i) manual, column-based methods [either Qiagen QIAamp Viral RNA Mini Kit (Qiagen, Germantown, MD) or Macherey Nagel Nucelospin RNA Virus kit (Macherey Nagel, Bethlehem, PA)] run in batches of 16 samples; ii) the semi-automated 16-sample throughput Promega Maxwell RSC Viral Total Nucleic Acid kit (Promega, Madison, WI) on the Maxwell RSC instrument; iii) the automated 96-sample throughput MagMAX Viral and Pathogen Nucleic Acid Isolation kit on the Thermo KingFisher Flex instrument (Thermo Fisher Scientific, Waltham, MA); or iv) the automated 96-sample throughput Zymo Quick-DNA/RNA Viral Mag Bead kit (Zymo Research, Irvine, CA) on the Tecan Fluent

instrument (Tecan Group, Männedorf, Switzerland). Note that the Maxwell RSC viral protocol has multiple manual pipetting steps of individual samples and was considered a manual protocol for modeling purposes. Extractions were performed per manufacturer's protocols, except as noted with the following modifications. For the Qiagen extraction, modifications included use of 100 μL of primary patient specimen and an elution volume of 50 μL. For the Macherey Nagel extraction, modifications included use of 100 μL of primary patient specimen and use of 650 μL of lysed sample loaded onto extraction columns. For extraction using Qiagen, Macherey Nagel, or Promega kits, batches of 16 primary nasopharyngeal samples were inactivated in a biosafety hood and extracted in parallel within a biosafety cabinet or on the Maxwell instrument and transferred to individually capped, two-dimensional–barcoded tubes arranged in two rows of eight samples for transport to the physically separate PCR setup facility; there, samples were opened with automated instrumentation and transferred into 96-well plates (combining six extraction batches of 16 samples) using automation-assisted pipetting instruments. Automation-assisted pipetting instruments were then used to aliquot a 384-well PCR assay plate from the 96-well sample master plates. For the automated extraction processes using KingFisher or Zymo kits, racks of 93 primary nasopharyngeal samples plus three controls were manually transferred in a biosafety cabinet into 96-well plates for viral inactivation, and then placed on the automated RNA extraction instrument to complete the procedure. Following completion of automated extraction, plates were sealed and transferred to the physically separate PCR setup facility; there, 96-well plates were unsealed and used for 384-well PCR assay plate setup as above. The assay limit of detection (LOD) is between 500 and 1600 viral copies per milliliter, with minor variation between extraction methods. A total of 500 viral copies correlated to C_T values of between 37 and 37.5 across 20 replicates performed using the EUA protocols (data not shown).

For studies on follow-up testing after a predicted or repeated FP result, PCR results were performed by one of four tests: the CDC N1/N2 PCR test described above, Xpert Xpress SARS-CoV-2 PCR assay (Cepheid, Sunnyvale, CA), Aptima SARS-CoV-2 transcription-mediated amplification assay (Hologic, Marlborough, MA), or Simplexa SARS-CoV-2 PCR assay (DiaSorin, Cypress, CA). All tests were performed as per manufacturer's instructions.

Interpretation of Results

Results were initially interpreted by an automated algorithm following CDC and FDA recommendations as follows:

N1 and N2 $C_T \leq 40$	Positive for SARS-CoV-2
Only N1 or N2 $C_T \leq 40$	Inconclusive for SARS-CoV-2
N1 and N2 C_T Undetected and RP $C_T \leq 38$	Negative for SARS-CoV-2
Undetected and RP $C_T > 40$	Invalid

Pathologists reviewed all autogenerated positive, inconclusive, and invalid results and multicomponent plots to confirm data quality and result interpretation; this review determined the need for repeated analysis based on data quality or plate layouts, indicating proximity of other positive samples. Inconclusive results with multicomponent plots demonstrating exponential amplification above a positive threshold (normalized reporter value > 0.5) with no concern for contamination were not repeated and resulted as positive, as per FDA recommendations. Positive or inconclusive samples with concern for contamination were re-extracted and repeated. Open-platform testing can be more prone to contamination of viral products due to aerosolization, plate seal removal, or pipettor drag during extraction or PCR setup.[8] To combat the risk of FP results in this open-platform laboratory-developed test (LDT), samples underwent repeated extraction and PCR if there was concern for contamination from surrounding wells, as indicated by a low relative viral load (RVL; see below for calculation) in the index well and high RVL in surrounding wells. Low RVL was defined at approximately <0.015 (-1.80 after $\log_{10}$ transformation), which correlates with N1/N2 C_T values of 34 and an RP C_T value of 28. RP C_T of 28 was selected as it was approximately the average value, and the N1/N2 C_T 34 value was empirically determined as it was approximately 50-fold less abundant than RP. Extraction method was also taken into account when selecting wells for repeat as manual and automated extractions empirically demonstrated different patterns of potential contamination in early quality control (QC) analyses (Supplemental Figure S1). Inconclusive samples with poor amplification, identified by multicomponent plot review, were re-extracted and repeated. Negative samples with RP C_T between 38 and 40 are re-extracted once to confirm the negative result. All invalid results were repeated once.

Data Preparation

Multiple data sets were generated from the analysis of the raw PCR data. The raw data set was first parsed for technically valid runs with no errors. Next, RVL was calculated for all autoscored positive and inconclusive samples (as defined by the interpretation algorithm described above) using the following formula: $[2^{(RP\ C_T - N1\ C_T)} + 2^{(RP\ C_T - N2\ C_T)}]/2$. This formula provides a normalized relative value of the amount of viral RNA in comparison to total human nucleic acid within each sample and was adapted from previous reports showing normalization was needed for reliable estimation of viral load across multiple samplings.[9] The formula used was adapted to incorporate both the N1 and N2

measurements as well as transform the C_T data to a linear scale. Next, geographic parameters were calculated that included east-west RVL, north-south RVL, surrounding RVL, and diagonal RVL (Supplemental Figure S1). For manually extracted samples, the extraction batch RVL of 16 samples was calculated by summing the individual RVLs for all positive samples in that batch, as these samples were handled simultaneously in a biosafety cabinet throughout the entire extraction procedure. The Tecan platform had specific plastic consumables with 16 compartments containing six pipette tips in a compartment (three horizontal, two vertical); the protocol included tip re-usage to minimize utilization of scarce plastic consumables, and these compartments were noted to contribute to contamination in experimental optimization and validation experiments. For Zymo/Tecan extracted samples, horizontal row RVL, vertical row RVL, and Tecan six-pack compartment RVL were calculated to account for movement of the automated pipetting heads across reagent and sample plates onboard the instrument. An example of geographic plate calculations is illustrated in Supplemental Figure S1. Some variables were derived from multiple wells (eg, east-west) with the RVLs of the wells summed to generate the final variable. For mathematical purposes, undetected C_T values were assigned a value of 46 as this was one cycle higher than the total number of 45 PCR cycles performed. Finally, a fold-change variable was generated by calculating the sum of the surrounding wells' RVL and dividing it by the index well RVL. This parsed and calculated data was defined as the cleaned data set. From the cleaned data set, four new parallel data sets were generated that included the following: i) primary data set, ii) final resulting data set, iii) repeated data set, and iv) modeling data set.

For the primary data set, samples were grouped on the basis of their unique sample identification number and parsed for only the first PCR result autocalled using the interpretation algorithm described above. Thus, the primary data set represents the preliminary test result generated by automated result classification without technologist or pathologist review for data quality and contamination potential before any human-determined repeated analysis was performed. For the final resulting data set, samples were grouped on the basis of their unique sample identification number and parsed for the final pathologist-reviewed result, which included the final result for samples subjected to repeated extraction and analysis. To generate the repeated data set, the cleaned data set was parsed and filtered specifically for samples that were repeated following re-extraction and linked by the unique sample identification number. Samples that underwent repeated PCR without repeated extraction were excluded from the analysis, as these were mostly due to PCR anomalies unrelated to concern for contamination. To generate the modeling data set from the repeated data set, any result initially autocalled as inconclusive in the primary data set was reclassified to positive, because the modeling approach needed a binary dependent variable for the result (positive or negative). No other result interpretations were modified. Repeated sample pairs in the modeling data set were then parsed on the basis of the results' interpretations for TP status (positive primary, positive repeat), which was assigned a value of 1; and for FP status (positive primary, negative repeat), which was assigned a value of 0.

Modeling

All RVL calculated variables in the modeling data set underwent further $\log_{10}$ transformation for modeling and graphical representation. Variables with a value of 0 were assigned the smallest value for that variable from the repeated data set before $\log_{10}$ transformation. Samples were separated into either manual or automatic extraction, and modeling was performed on centered and scaled data. FPs were predicted using a gradient boosting machine model using the *caret* version 6.0-86[10] package with a 10-fold repeated cross-validation. Partial dependence plots were generated using the *pdp* package version 0.7.0.[11] To generate confusion matrix tables, cutoff values for the probabilities were set to 0.5. For analysis of the final resulting data set, FP events were defined as a probability of $\leq$0.15 for manual, 0.1 for automated-KingFisher extraction, and 0.25 for automated-Tecan extraction. Data were analyzed in R version 4.0.2; code is available (*https://github.com/nelsonac-umn-lab/covid*, last accessed June 2, 2021). Statistical testing and graphical representation were performed using GraphPad Prism version 8.4.3 for MacOS (GraphPad Software, San Diego, CA) or using FlowJo Software for MacOS version 10.7.1 (Becton, Dickinson and Company, Ashland, OR).

Regulatory Statement and Data Availability

Utilization of clinical test results for the purposes of test validation and quality improvement was reviewed and approved by the Institutional Review Board (STUDY00009560 and STUDY00010603). The repeated data set used for modeling is included (Supplemental Table S1).

Results

Testing with our laboratory-developed RT-PCR assay for detection of SARS-CoV-2 viral RNA began on March 22, 2020. By August 12th, 2020, 206,445 samples were resulted with an overall 3.5% positive rate (Table 1) (see *Materials and Methods* for description of data generation). Results are composed of samples from both symptomatic and asymptomatic patients. The primary results generated before repeat of any samples demonstrated 1.4% invalid and 1.0% inconclusive results; however, repeated testing and pathologist review resolved most of these cases, and a much smaller number of samples remained inconclusive (0.1%) and invalid (0.4%) (Table 1). Early in this time period,

Table 1 Summarized Primary and Final SARS-CoV-2 Results

Result	Primary data set*	Final resulting data set*
Not detected	194,706 (94.3)	198,153 (96.0)
Positive SARS-CoV-2	6727 (3.3)	7273 (3.5)
Invalid	2894 (1.4)	777 (0.4)
Inconclusive	2103 (1.0)	202 (0.1)
QNS	14 (0.006)	35 (0.02)
No reaction	1 (0.0005)	5 (0.002)
Total	206,445	206,445

Data are given as *n* (%).
*See *Materials and Methods* for description of data set generation.
QNS, quantity not sufficient.

sample RNA was extracted using a manual column-based method or a low-throughput magnetic bead-based method for sets of 16 samples. Later, automated 96-well plate–based methods for RNA extraction were validated and used regularly. Manual extraction, based on extraction sets of 16 samples, was still performed daily on a smaller number of samples left over after plate-based RNA extraction. In total, three distinct extraction workflows were used for these clinical samples: manual extraction, automated KingFisher-based extractions, and automated Tecan-based extractions. Daily test volume increased as the amount of automation was increased (Figure 1A). By August, up to 6000 tests were performed daily. As testing volume increased, the prevalence of positive SARS-CoV-2 tests decreased from approximately 10% to 20% positivity to approximately 1.0% to 1.5% by early July, but then began to increase to approximately 5% in late July (Figure 1B). Validation, performance, and comparison studies for the SARS-CoV-2 RT-PCR test are described elsewhere.[7] By using this high-throughput, open-platform LDT RT-PCR assay with different extraction methods, testing was able to be rapidly scaled to account for increasing community needs with fewer supply bottlenecks associated with some proprietary, closed-platform approaches of detecting SARS-CoV-2.

Of the total samples assayed (36,692 manual, 169,564 automated, and 189 with no extraction method), 3311 were re-extracted, underwent subsequent PCR, and resulted (repeated data set) (Table 2). Samples that underwent automated extraction using either the Tecan or KingFisher methods are combined herein to compare manual versus automated methods. On retrospective data review, extraction methods for 54 repeated samples (4 positive samples, 1 sample with no reaction, and 46 not detected) were not identified. These samples were removed from modeling analysis but are present in the total data set. The most common reason samples were repeated following automated extraction was invalid primary results, whereas the most common reason samples were repeated following manual extraction was inconclusive primary results (Table 2). Analysis comparing TP and FP results demonstrated significantly more FP events when initial manual extraction methods were used (Fisher exact test, $P < 0.0001$) (Figure 2, A and B, and Table 3). Thus, FP events are identified in the SARS-CoV-2 PCR assay and are generated at different rates across extraction methods.

We hypothesized that samples with initial positive/inconclusive results returning negative results on repeat are likely FP events due to contamination; however, some samples could represent an initial TP and repeated false negative because of poor sample quality or inconsistent detection of low-concentration SARS-CoV-2 RNA. The data were used to evaluate for these two possibilities. First, if the FP events were associated with poor sample collection, the FPs would demonstrate reduced sample quality and reduced ability to detect viral RNA. The C_T values of the human internal positive control (RP) were compared across TP and FP samples within the same extraction type, finding no difference (one-way analysis of variance, manual comparison $P > 0.99$; automated comparison $P = 0.93$) (Figure 2C). This suggested the extent of sampling and/or sample integrity was not different between TP and FP samples. To explore the possibility of the inability to identify inadequate sample collection due to minimal difference between RP C_T values, all measured samples were

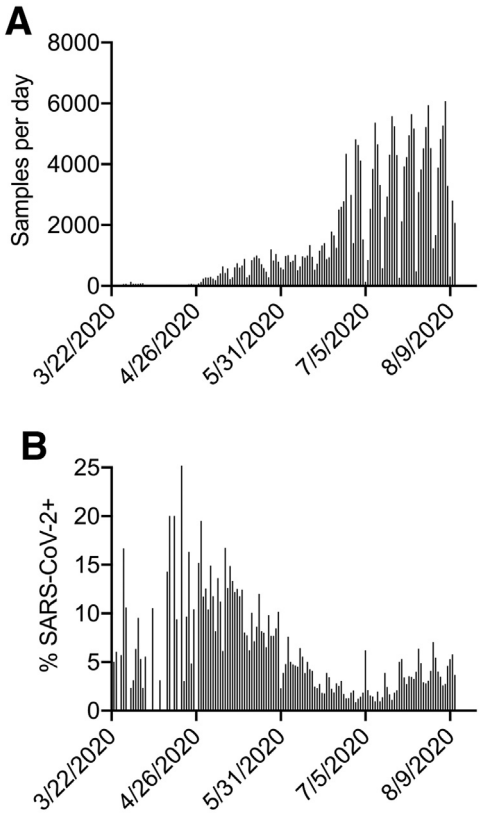

Figure 1 Overview of testing during time of study. **A:** Number of samples measured for SARS-CoV-2 as a function of time. **B:** Frequency of positive SARS-CoV-2 samples as a function of time. Bars represent a single day of tests.

Table 2 Summarized Results of SARS-CoV-2 Repeated Data Set

Results	Manual	Automated
Inconclusive → not detected	93 (40.0)	321 (10.4)
Invalid → not detected	39 (16.7)	1785 (58.0)
Not detected → not detected	37 (15.9)	187 (6.1)
Inconclusive → positive SARS-CoV-2	21 (9.0)	244 (7.9)
Invalid → invalid	14 (6.0)	287 (9.3)
Positive SARS-CoV-2 → not detected	13 (5.6)	35 (1.1)
Positive SARS-CoV-2 → positive SARS-CoV-2	11 (4.7)	84 (2.7)
Invalid → positive SARS-CoV-2	2 (0.9)	41 (1.3)
Not detected → inconclusive	1 (0.4)	3 (0.1)
Positive SARS-CoV-2 → inconclusive	1 (0.4)	2 (0.06)
Inconclusive → inconclusive	1 (0.4)	33 (1.1)
Invalid → QNS	—	21 (0.7)
Not detected → positive SARS-CoV-2	—	4 (0.1)
Not detected → invalid	—	5 (0.2)
Invalid → inconclusive	—	21 (0.7)
Inconclusive → invaild	—	4 (0.1)
Positive SARS-CoV-2 → invalid	—	1 (0.03)
Total	233	3078
All samples run (206,445)*	36,692	169,564

Data are given as *n* (%).
*A total of 189 samples were not coded for extraction method.
—, Not identified in this extraction method; QNS, quantity not sufficient.

Table 3 Summarized Results of SARS-CoV-2 Modeling Data Set

Results	Manual	Automated
Positive SARS-CoV-2 → not detected	98 (42.0)	330 (10.7)
Not detected → not detected	66 (28.3)	246 (8.0)
Invalid → not detected	25 (10.7)	1718 (55.8)
Positive SARS-CoV-2 → positive SARS-CoV-2	24 (10.3)	343 (11.1)
Invalid → invalid	11 (4.7)	317 (10.3)
Not detected → positive SARS-CoV-2	7 (3.0)	32 (1.0)
Invalid → positive SARS-CoV-2	1 (0.4)	79 (2.6)
Not detected → invalid	1 (0.4)	5 (0.2)
Positive SARS-CoV-2 → invalid	—	8 (0.3)
Total	233	3078

Data are given as *n* (%).
—, Not identified in this extraction method.

compared. RP C_T values were found to be normally distributed with a mean of 28.1 and an SD of 3.0 (data not shown), demonstrating approximately 95% of samples were between 34 and 22 C_T. Thus, there is a 2^{12} difference in RP transcript quantity and is likely enough difference between samples to identify poor quality specimens based on RP C_T value. Next, we hypothesized that the identification of FP results was due to samples with a viral concentration below this assay's LOD. The LOD for this assay is defined as the viral RNA quantity at which 95% (19 of 20) of replicates for positive samples are identified, which was associated with average C_T values of approximately 37.5. As the viral concentration falls below the LOD, the frequency of replicates producing a positive result will also decrease. To test this hypothesis, the RVL of all TP and FP samples in the repeated data set was compared, finding no significant difference between the manual extraction (one-way analysis of variance, manual comparison $P = 0.71$), but a trend toward lower RVL for FP in the automated extractions (one-way analysis of variance, $P = 0.11$) (Figure 2D). This does not rule out that samples with low viral load are contributing to the initially positive SARS-CoV-2 results that repeat negative, but it does suggest that low viral concentration is not clearly the main contributor to nonreproducible results and justifies detailed study of potential mechanisms driving contamination.

Sample contamination occurs in either a reproducible or a probabilistic manner. Technical method may introduce the risk for reproducible contamination, as most technicians/technologists, and all automated machines, perform tasks in an ordered manner (ie, pipetting of samples left to right or starting at well A1 and going to well A12). Stochastic contamination can occur due to aerosolization of RNA or

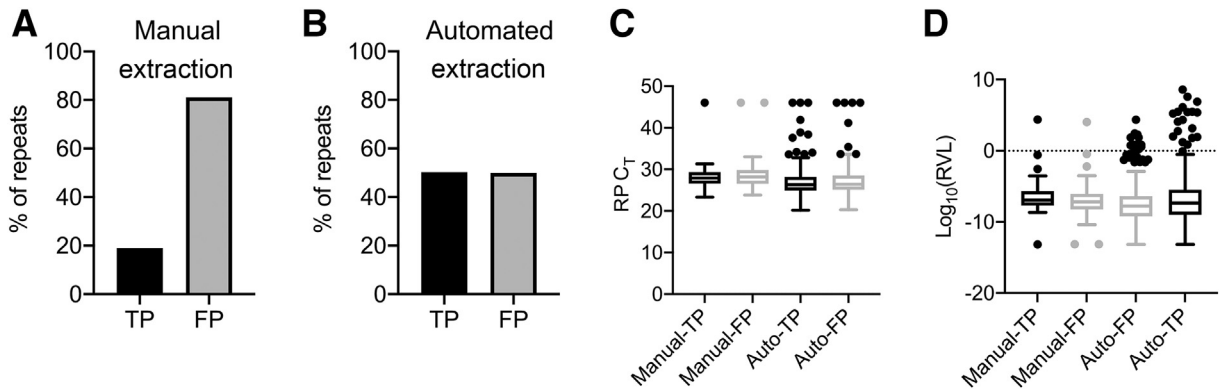

Figure 2 False-positive (FP) results are not explained by sample quality or relative viral load. **A:** Frequency of repeated manually extracted samples that were true positive (TP) and FP. **B:** Frequency of repeated automatically (Auto) extracted samples that were TP and FP. **C:** RNase P (RP) C_T values for repeated TP and FP samples that were initially manually or automatically extracted (Tukey box-and-whisker one-way analysis of variance with multiple comparisons). **D:** Relative viral load (RVL) of repeated TP and FP samples that were initially manually or automatically extracted (Tukey box-and-whisker one-way analysis of variance with multiple comparisons).

amplicon to surrounding wells and has been shown to occur in plate-based bacterial 16s RNA sequencing.[12] Even with stochastic contamination, there were still reproducible patterns of contamination in the 16s sequencing as abutting wells were more likely to exchange material than distant wells.[12] To investigate reproducible errors and potential sources of contamination in this open-platform assay, we set out to predict samples that had a high likelihood of being FP to identify the samples for repeat. Our basic assumptions were that wells directly surrounding a sample were most likely to contribute to contamination and that the relative difference in viral concentration between samples would impact the probability of an FP result. Because of technical differences across the three extraction methods, we hypothesized that certain directional relationships between neighboring samples could be more prominent contributors to the manner of contamination. On the basis of these assumptions, 12 different variables (Supplemental Table S2 and Supplemental Figure S1) were defined for analysis and data were modeled from each of the three extraction methods separately. The models' predictions showed a positive predictive rate of 87%, 93%, and 65% for detecting FPs using the manual, automated-KingFisher, and automated-Tecan extraction methods, respectively (Figure 3, A–C). Sensitivity of the models for the three extraction methods were 96%, 100%, and 67%, whereas the specificity was only 36%, 90%, and 66%, respectively (Figure 3, A–C). Identification of the top variables contributing to the models' predictions demonstrated that N2 C_T value contributed substantially in all three models, whereas unique extraction-specific values, such as the extraction batch RVL (manual) and the horizontal and Tecan shared pipette RVLs (automated-Tecan), were important in their respective methods (Figure 3, D–F). These results demonstrate that the extraction methods have different variables associated with prediction of FPs, indicating that contamination is produced in a method-dependent manner and that models need to be generated individually when applying this QC method to other assays.

The data used to generate the manual extraction model was further studied, as the manual method may be most broadly applicable to other laboratories and potentially the most likely to demonstrate reproducible error. The top three variables contributing to the manual model were studied first: N2 C_T, extraction batch RVL, and the calculated fold change of the measured well to the surrounding well RVL. Comparison of these three variables showed that the FPs exhibited higher values across all variables when compared with the TPs (Figure 4, A–C). These results support our hypothesis about potential sources of intersample contamination. First, the presence of multiple positive samples, particularly those with higher viral loads within a single manual extraction batch ($n = 16$ samples), is associated with FPs. Second, samples with higher or undetected viral N2 C_T values are more likely to be FPs. Last, large differences in the collective viral load of all adjacent samples surrounding an index well, compared with the index well itself, are more likely to identify an FP sample. However, none of these three variables was able to fully separate TP and FP samples, indicating the need for a more nuanced model.

Next, it was investigated whether more specific physical relationships between positive samples and the surrounding wells impacted the final result. First, the association of more specific geographic variables relating the RVLs of neighboring samples on east-west, north-south, and diagonal axes

A

Predicted data	Repeated data: Manual	False positives	True positives
	False positives	103	16
	True positives	4	9

B

Predicted data	Repeated data: Automated-KingFisher	False positives	True positives
	False positives	41	3
	True positives	0	27

C

Predicted data	Repeated data: Automated-Tecan	False positives	True positives
	False positives	214	112
	True positives	102	217

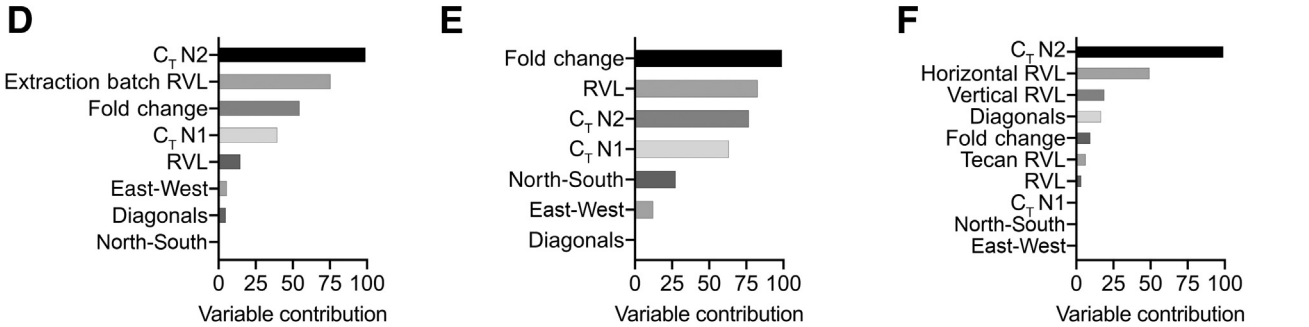

Figure 3 Machine learning–derived models identify false-positive (FP) samples. Confusion matrix from machine learning–derived models of predicting FP events of manually (**A**), automated-KingFisher (**B**), or automated-Tecan (**C**) extracted samples. Bar graph of the variable contribution to the models generated for the manual (**D**), automated-KingFisher (**E**), and automated-Tecan (**F**) extraction methods. For each model, individual variables used in the modeling were normalized to the maximal contributor variable. RVL, relative viral load.

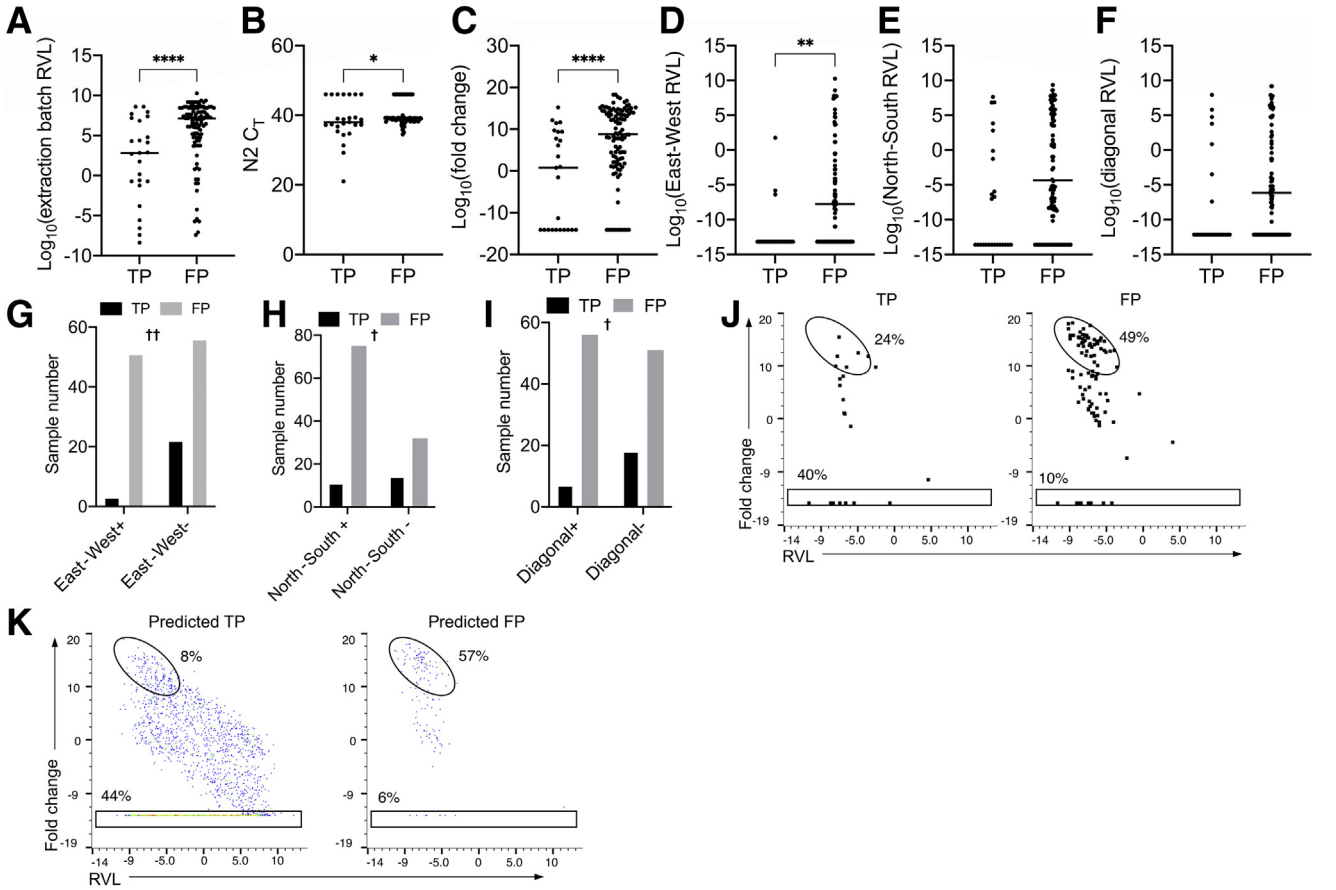

Figure 4 Identification of variables contributing to false positives (FPs) in manual extraction. **A:** Scatterplot of $\log_{10}$ transformed extraction batch relative viral load (RVL) of true-positive (TP) and FP samples (samples represented by a single point, two-tailed *t*-test, bar at median). **B:** Scatterplot of N2 C_T values of TP and FP samples (samples represented by a single point, two-tailed *t*-test, bar at median). **C:** Scatterplot of $\log_{10}$ transformed fold change of TP and FP samples (samples represented by a single point, two-tailed *t*-test, bar at median). **D**–**F:** Scatterplot of $\log_{10}$ transformed east-west (**D**), north-south (**E**), and diagonal (**F**) RVL of TP and FP samples (samples represented by a single point, two-tailed *t*-test, bar at median). **G**–**I:** Enumeration of TP and FP samples with (+) and without (−) a positive sample present in the east-west (**G**), north-south (**H**) or diagonal (**I**) position (for each bar graph, an individual Fisher exact test was performed for the unique geographic data and the *P* value is noted within the graph). **J:** Scatterplots of RVL by fold change for manually extracted samples that were repeated because of concern for FP results (points represent single sample), with gates showing the frequency of samples with positive surrounding wells (high fold change and low RVL; oval) and negative surrounding wells (low fold change and range of RVL; rectangle). **K:** Scatterplots of RVL by fold change for manually extracted samples that were predicted to be FP (**left panel**) or TP (**right panel**) (points represent single sample) with gates showing the frequency of samples with positive surrounding wells (high fold change and low RVL; oval) and negative surrounding wells (low fold change and range of RVL; rectangle). $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{****}P < 0.0001$; $^{\dagger}P < 0.05$, $^{\dagger\dagger}P < 0.01$ with a positive sample versus without a positive sample.

to the incidence of FP and TP results (Figure 4, D–F) were analyzed. The east-west RVL was higher in FP events, whereas the north-south and diagonal RVLs were not significantly different between TP and FP (Figure 4, D–F). Reasons for these directional differences associated with FP results are not certain but may suggest that the process of aliquoting into and out of two-dimensional–barcoded tubes for sample transfer to PCR setup may be a source for technical improvement based on our quality review of this specific workflow. It was also tabulated whether the wells surrounding TP and FP events did or did not contain a positive sample, irrespective of the relative viral load (Figure 4, G–I). Indeed, the presence of positive samples in the east-west, north-south, or diagonal wells increased the likelihood a sample was FP, suggesting the mere presence of a positive well of any RVL may be able to contaminate an adjacent well. Interestingly, though, a proportion of FPs demonstrated no positive samples in the immediately neighboring wells, suggesting contamination may also occur from outside the directly surrounding samples during processing. To determine how many FP and TP samples were adjacent to a positive well, the RVL and fold change measurements were correlated for manually extracted repeated samples (Figure 4J). Samples without surrounding positive wells were represented at the lowest RVL values (within rectangle gate), whereas those with positive adjacent wells were above this line of samples (within oval gate). Analysis revealed 40% of TP results did not have a surrounding positive sample, whereas 10% of FP results did not have a surrounding positive well (Figure 4J). Analysis of the RVL and fold change values for the full manually extracted data set with FP predictions showed 6% of samples had empty

surrounding wells, whereas 44% of TP samples had directly adjacent empty wells (lower box) (Figure 4K). Thus, the model predicts samples of low RVL with adjacent samples of high RVL are more likely to be predicted FP, whereas samples of high RVL without adjacent positive samples are more likely to be TP.

Next, the manual extraction model was interrogated to understand how the model variables predicted FP samples. First, the FP probability of manually extracted repeated samples was evaluated as a function of N1 and N2 C_T values (Figure 5, A and B). Samples with N1 and N2 C_T values of less than approximately 38 typically repeated as positive when extracted and measured again, whereas values greater than approximately 38 were more likely to not be detected on repeat. To compare the experimental and model data, one-dimensional partial dependency plots were generated for variables in the model (Figure 5, C–G). Partial dependency plots show the model's probability of predicting an FP sample as a function of the input variable. The N2 C_T (Figure 5D) and extraction batch RVL (Figure 5E) variables both show approximately 20% to 30% increases in the model FP probability at higher values. N1 C_T (Figure 5B), fold change (Figure 5F), and RVL (Figure 5G) also show increases in the probability of an FP sample at higher values, but the model prediction only changes minimally over the variable range. Comparison of the experimental data and model reveals the model assigns increased likelihood of FP results to a C_T value of >38 for both N1 and N2 and that even though the experimental N1 and N2 data look similar, N2 carries more weight in predicting FP than N1 in the model. Next, two-dimensional partial dependency plots were generated to investigate how two variables interacted within the prediction models (Figure 5H). As expected, low N1 and N2 C_T values were more likely to be TP, whereas samples with high N1/N2 C_T values are predicted FP. High extraction batch RVL values led the model to predict the sample as FP, even at low N2 C_T values, suggesting that more caution is necessary in QC review as more positive samples are batch handled in a biosafety cabinet simultaneously. Last, samples with low fold change were less likely to be FP, especially at N2 C_T values of less than approximately 38. This analysis better defines the interaction of variables that can be used to discriminate manually extracted TP and FP samples. As well, samples with a summed extraction batch RVL of $>10^5$ or a change of 10-fold compared with the positive surrounding wells were more likely to be FP.

Last, the three models were used to identify potential FP results in the final resulted data set that were not identified for repeated extraction by pathologist review. The probability threshold for identifying an FP in the models was increased such that the accuracy of FP predictions were maximized. Using this new probability threshold, the models predicted an additional 170 samples concerning for an FP result (0.08% of total results, 2.3% of positive/inconclusive results) that had not been identified for repeat. FP samples identified by retesting or modeling may represent low-viral titer samples in the presymptomatic or early

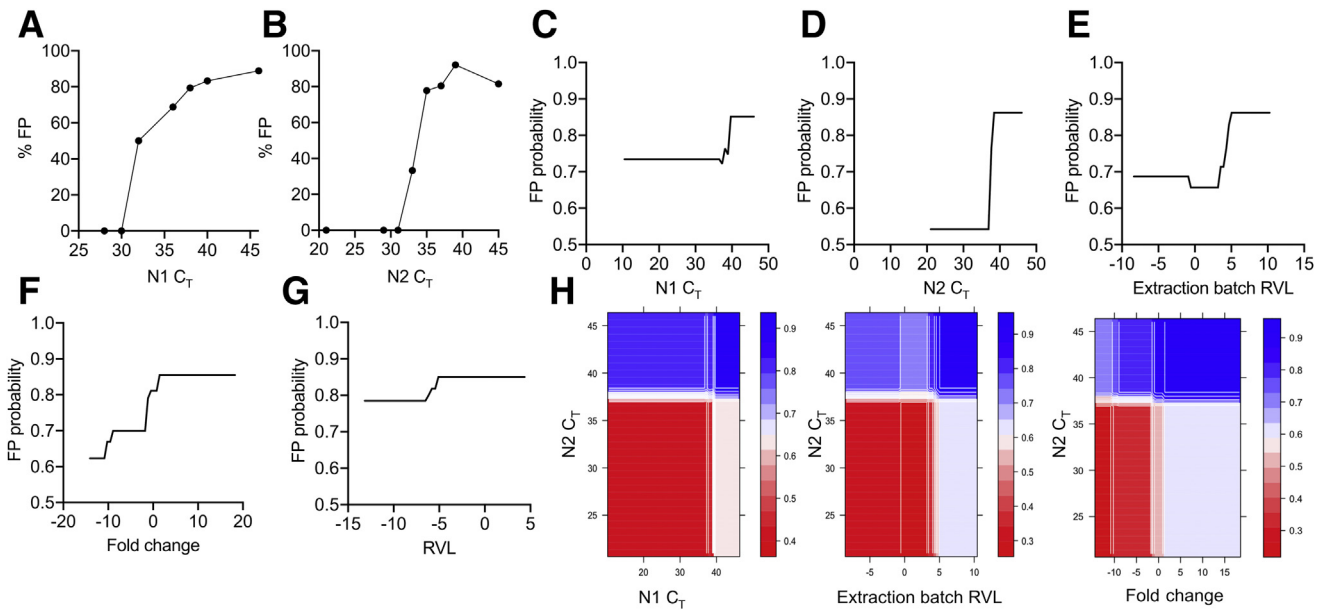

Figure 5 Identification of variables used for predicting false positives (FPs) in the manual extraction model. **A** and **B:** FP probability as a function of N1 C_T (**A**) and N2 C_T (**B**). The points represent the FP probability between that point and the next lowest C_T value point normalized to the total frequency of FP events. **C**–**G:** Partial dependency plots showing the probability of predicting an FP event in the manual extraction model for N1 C_T (**C**), N2 C_T (**D**), extraction batch relative viral load (RVL; **E**), fold change (**F**), and RVL as a function of variable values (**G**). **H:** Partial dependency heat map plots show the probability of predicting an FP event (red, low probability; blue, high probability) in the model as a function of N1 C_T versus N2 C_T (**left panel**), extraction batch RVL versus N2 C_T (**middle panel**), and fold change versus N2 C_T (**right panel**).

asymptomatic phase of infection that will progress and become TP results. To address this possibility, manually extracted FP samples (retested and predicted) were investigated to determine whether the patient received a follow-up test within the following 14 days. The value of 14 days was selected to balance the following considerations: allowance for progression of the viral illness, reduction in the likelihood of new infection, and maximization of the sample number for assessment. Sixteen patients were identified who had a second test within 14 days, finding 11 of these patients (69%) were negative on retesting. Thus, these models can be used as a decision-support tool when identifying samples needing repeat because of concerns for contamination. Given that the model misidentified predictions in the pathologist-selected repeated data set used for training, complementary pathologist or technologist review of the PCR data still serves an essential role in identifying FP events for the SARS-CoV-2 RT-PCR assay.

Discussion

Using data generated from repeated samples on an LDT high-throughput test, models were developed to predict false-positive SARS-CoV-2 PCR results and generate QC metrics for individual extraction methods. Employment of these models and identified variables will be used to improve the assay, as well as support clinical decision making when performing data quality review of results. Specifically, the manual extraction model identified high relative viral loads within each 16-sample extraction batch as a potential source of contamination. As well, the presence of a positive result in the surrounding wells, particularly in the east-west well, helped to identify FP results. From these findings, the protocol and technique can be altered in an attempt to reduce contamination during specific steps of processing. Last, we showed that an additional 2.3% (n = 170) of positive and inconclusive samples had a high likelihood of an FP result but were signed out as positive without re-extraction and repeated PCR. This highlights the importance of providing computational decision-support methods to pathologists or technologists performing QC review of data for molecular viral testing.

Analysis and identification of the variables likely contributing to FP results include the extraction batch RVL in the manual extraction model, as well as individual sample N2 and calculated fold change values in all models. Specifically, the proposed QC metrics to identify FP manually extracted samples are as follows: i) N2 C_T value (>38), ii) extraction batch RVL (>10^5), and iii) fold change (>10). These three variables demonstrated the largest contribution to predicting FPs and demonstrated cutoff values that could be used quickly for manual evaluation of specimen result quality. One important caveat of this work is that these models only inform the FP events generated using the modeled assay workflow and are not directly applicable without modification to other SARS-CoV-2 assays. However, use of this modeling framework to identify technical components contributing to FP results can be applicable to all open-platform RT-PCR assays. Monitoring of the variables identified above was incorporated into our laboratory's standard workflows for this assay and was applied to the analysis of a newly developed extraction-free SARS-CoV-2 RT-PCR assay using the CDC N1/N2/RP targets. A C_T cutoff and fold change variable were used to identify samples to automatically repeat using an orthogonal testing method, increasing turnaround time and accuracy of the assay.

Identification of extraction batch RVL as a potential predictor of FP results points to contamination that is hypothesized to be caused by aerosolization of samples during multiple manual extraction handling steps. This finding allows for specific protocol modifications and standardization of technique among operators to improve results and reduce contamination rates. In all models, N2 C_T values were identified as a predictor of FP results, whereas high N1 C_T values did not show the same ability to differentiate TP and FP results. This may point to differences in the efficiency of detecting N1 and N2 targets and may be a finding applicable to assays using this primer and probe set. For automated-Tecan samples, the model was less accurate at discriminating TP and FP results. Several hypotheses could be made for this finding. First, it is possible any of the models are overfit and identify insignificant differences within the repeated data that can distinguish the two groups but have no real-world applicability. Recursive feature elimination of variables was performed in an attempt to avoid overfitting, but no variables were identified for removal (data not shown). Second, a complex set of interactions could be driving FP events that are not easily identifiable using single- or two-variable analysis, as shown using partial dependency plots. Last, the results of the automated extraction model may point to only stochastic FP generation, with little to no reproducible sources of error, and may be due to mechanical differences in the specific performance of the different automated systems utilized in our laboratory. Nevertheless, this analysis shows the importance of generating a model to potentially identify and mitigate sources of error, allowing for improvement of technique and reduction of erroneous results across multiple methods.

Many of the results flagged by the models as potential FPs are composed of samples with higher C_T values. This is partially due to the conversion of inconclusive to positive results for modeling purposes. This finding is also due to the assumptions made about TP events and the selection of repeats generated for the training data set: it is unlikely a sample with high viral load is FP, whereas a sample with low viral RNA concentration is potentially explainable by contamination. Because most of the training data and therefore predictions are approaching the LOD, it is difficult to confidently distinguish between low viral titer and contamination. Follow-up testing of patients with predicted

and re-extracted FP samples shows minimal conversion to a TP result over the next 14 days, suggesting that the algorithm identifies contamination and not just early infection. Distinguishing between these two events may have different utility in certain clinical situations. For example, missing a TP with low viral load when performing contact tracing or during prehospital admission to a nonisolation ward could have severe consequences as patients in these clinical situations may be in a presymptomatic phase with low viral titers and could be at high risk to spread SARS-CoV-2 as their viral titer increases. Conversely, overcalling FP results when monitoring health care workers could lead to unnecessary staffing shortages in critical service areas. Furthermore, FP results for prehospital or presurgical admissions could confound patient placement and immediate management of emergent non−COVID-19 health issues, leading to suboptimal patient care and inefficient utilization of health care resources. These complexities stress the importance of clinical correlation in the interpretation of SARS-CoV-2 test results and highlight the major need for sufficient sample collection supplies and analytic reagents to offer subsequent sampling to patients with inconclusive, low viral load results.

Currently, this LDT defines inconclusive results as amplification and detection of only one of the two targets. It is unknown if inconclusive results are at increased risk of being an FP. Tests receiving EUA approach inconclusive results differently: some classify them as negative, some classify them as positive, and some leave it to the discretion of the laboratory (FDA, *https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/vitro-diagnostics-euas*, last accessed March 4, 2021). These EUA tests also have different C_T cutoff values for the respective positives, inconclusive results, and negatives. Some SARS-CoV-2 tests allow positives to be assigned with only one target amplifying at a C_T of 45, whereas others are negative at C_T values of >37. The sensitivity and specificity of the tests change with the alteration of C_T cutoff values and PCR efficiency, whereas the clinical purpose of the test should reflect how the C_T value thresholds are set.[13] Currently, SARS-CoV-2 testing focuses on increasing the sensitivity of the test and identifying any amount of viral RNA present in a sample, even if there is a concurrent increase in FPs. If instead the goal was to identify patients at risk for spreading infection, reducing the positive C_T cutoff would be more practical. Recent works have found that samples with C_T values of >35 or viral load $<10^5$ copies were unable to generate recoverable, culturable virus.[14,15] However, these results need to be interpreted cautiously as the studies were performed *in vitro* and C_T values have been shown to be assay dependent.[14,15] To address the questions of C_T values, transmissibility and test sensitivity thresholds, proficiency testing, and standardized samples may be useful. A recent SARS-CoV-2 proficiency test has shown 97% consensus for samples with approximately 5000 copies/mL and negative samples,[16] but correlating virus quantity with culture recoverability may be a more useful approach for clinical laboratories. Nonetheless, until these studies are performed, inconclusive results are difficult to interpret. In this study, inconclusive results that underwent repeat testing were coded as positive in training data sets for the purpose of modeling. When initially inconclusive results were reviewed in the repeated data set, 59% (409 of 690) were FP, whereas 41% (281 of 690) were TP. In contrast, review of initially positive results in the repeated data set showed 35% (55 of 158) were FP, whereas 65% (103 of 158) were TP (Fisher exact test, $P < 0.0001$, data not shown), demonstrating that inconclusive samples were found more frequently to be FP. Nonetheless, given the overall frequency of TP results in the initially inconclusive samples for this LDT assay, it would be inappropriate to simply call inconclusive results negative because it would result in the misdiagnosis of reproducible, positive samples. On the basis of these results, we propose two options for follow-up testing of a positive sample with high probability of an FP result. First, report the result as inconclusive and request that the patient is sampled and measured again in 48 to 72 hours to identify patients with low viral load early in infection. Second, if the patient cannot be tested again, perform repeated analysis of the primary sample either on the same platform or optimally on an orthogonal method. Clinical scenario will dictate the follow-up testing options, but this is a reasonable laboratory approach until more robust data are available to calibrate assay thresholds with clinical infectivity.

Finally, generation of these models will lead to improvement in multiple aspects of the test. First, it allows the laboratory to identify aspects of the technique that are reproducibly generating errors. Technical modifications will allow for remediation of these issues, and comparison of premodification and post-modification results will reveal the effect on error rates. Second, model implementation can generate automatic FP flags to improve workflow efficiency in triaging technical repeats, likely decreasing result turnaround times. Third, modeling can support clinical decision making and provide increased ability to identify potential errors. Anatomic pathology and radiology have begun to employ artificial intelligence for image analysis as an adjunct diagnostic tool, but artificial intelligence has not been widely used for molecular diagnostics clinical decision support.[17,18] For SARS-CoV-2 testing, we propose to use this method as a clinical decision support artificial intelligence to flag results that have a high FP probability. Of note, some of the models had poor positive predictive values (ie, automated-Tecan), and would be unlikely used as an artificial intelligence decision support. However, this same model could still be used to identify potential reproducible contamination issues. This metric could be incorporated into the clinical decision process and inform the reviewer if the sample should be repeated. This model may enforce confirmation bias while training the reviewer to miss other errors, but future studies will be needed to determine how

the adjunct tool alters clinical decision making. Nonetheless, this tool will be helpful for technique troubleshooting and clinical decision-making support and can be adapted to a wide variety of molecular diagnostic applications.

Acknowledgments

We thank Robyn Kincaid and Drs. Sophie Arbefeville, Aaron Barnes, and Ryan Langlois for helpful discussion in the preparation of this article; the University of Minnesota Genomics Center (including Benjamin Auch, Ray Watson, Lindsey Gengelbach, Darrell Johnson, Dr. Patrick Grady, Dr. Daryl Gohl, and Shea Anderson) and the M Health Molecular Diagnostic Laboratory and Infectious Disease Diagnostic Laboratory (including Kylene Karnuth, Michaela Leary, Shannon Gascoigne, and Jessica Gunderson) for staffing, support, and testing performance of the SARS-CoV-2 testing.

Author Contributions

R.J.M. analyzed the data and wrote the manuscript; N.P., P.F., A.B.K., and B.T. analyzed the data and edited the manuscript; M.S. and J.D. performed experiments, analyzed the data, and wrote the manuscript; K.B.B. designed the study, performed experiments, and edited the manuscript; S.Y. designed the study and edited the manuscript; A.C.N. designed the study, interpreted data, and wrote the manuscript. A.C.N. is the guarantor of this work and, as such, had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Supplemental Data

Supplemental material for this article can be found at *http://doi.org/10.1016/j.jmoldx.2021.05.015*.

References

1. Sethuraman N, Jeremiah SS, Ryo A: Interpreting diagnostic tests for SARS-CoV-2. JAMA 2020, 323:2249–2251
2. Mackay IM, Arden KE, Nitsche A: Real-time PCR in virology. Nucleic Acids Res 2002, 30:1292–1305
3. Hanson KE, Caliendo AM, Arias CA, Englund JA, Lee MJ, Loeb M, Patel R, El Alayli A, Kalot MA, Falck-Ytter Y, Lavergne V, Morgan RL, Murad MH, Sultan S, Bhimraj A, Mustafa RA: Infectious Diseases Society of America guidelines on the diagnosis of COVID-19. Clin Infect Dis 2020, [Epub ahead of print] doi:10.1093/cid/ciaa760
4. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J: Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science 2020, 368:489–493
5. Woloshin S, Patel N, Kesselheim AS: False negative tests for SARS-CoV-2 infection — challenges and implications. N Engl J Med 2020, 383:e38
6. Cohen AN, Kessel B, Milgroom MG: Diagnosing SARS-CoV-2 infection: the danger of over-reliance on positive test results. MedRxiv 2020. [Preprint] doi:10.1101/2020.04.26.20080911
7. Nelson A, Auch B, Schomaker M, Gohl D, Grady P, Johnson D, Kincaid R, Karnuth K, Daniel J, Fiege J, Fay E, Bold T, Langlois R, Beckman K, Yohe S: Analytical validation of a COVID-19 qRT-PCR detection assay using a 384-well format and three extraction methods. bioRxiv 2020, [Preprint] doi:10.1101/2020.04.02.022186
8. Espy MJ, Uhl JR, Sloan LM, Buckwalter SP, Jones MF, Vetter EA, Yao JDC, Wengenack NL, Rosenblatt JE, Cockerill FR, Smith TF: Real-time PCR in clinical microbiology: applications for routine laboratory testing. Clin Microbiol Rev 2006, 19:165–256
9. Duchamp MB, Casalegno JS, Gillet Y, Frobert E, Bernard E, Escuret V, Billaud G, Valette M, Javouhey E, Lina B, Floret D, Morfin F: Pandemic A(H1N1)2009 influenza virus detection by real time RT-PCR: is viral quantification useful? Clin Microbiol Infect 2010, 16:317–321
10. Kuhn M: Building predictive models in R using the caret package. J Stat Softw 2008, 28:i05
11. Greenwell BM: pdp: An R package for constructing partial dependence plots. R J 2017, 9:421–436
12. Minich JJ, Sanders JG, Amir A, Humphrey G, Gilbert JA, Knight R: Quantifying and understanding well-to-well contamination in microbiome research. MSystems 2019, 4:1–13
13. Caraguel CGB, Stryhn H, Gagné N, Dohoo IR, Hammell KL: Selection of a cutoff value for real-time polymerase chain reaction results to fit a diagnostic purpose: analytical and epidemiologic approaches. J Vet Diagn Investig 2011, 23:2–15
14. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, Niemeyer D, Jones TC, Vollmar P, Rothe C, Hoelscher M, Bleicker T, Brünink S, Schneider J, Ehmann R, Zwirglmaier K, Drosten C, Wendtner C: Virological assessment of hospitalized patients with COVID-2019. Nature 2020, 581:465–469
15. La Scola B, Le Bideau M, Andreani J, Hoang VT, Grimaldier C, Colson P, Gautret P, Raoult D: Viral RNA load as determined by cell culture as a management tool for discharge of SARS-CoV-2 patients from infectious disease wards. Eur J Clin Microbiol Infect Dis 2020, 39:1059–1061
16. Edson DC, Casey DL, Harmer SE, Downes FP: Identification of SARS-CoV-2 in a proficiency testing program. Am J Clin Pathol 2020, 154:475–478
17. Niazi MKK, Parwani AV, Gurcan MN: Digital pathology and artificial intelligence. Lancet Oncol 2019, 20:e253–e261
18. Lu MT, Ivanov A, Mayrhofer T, Hosny A, Aerts HJWL, Hoffmann U: Deep learning to assess long-term mortality from chest radiographs. JAMA Netw Open 2019, 2:1–14