False negative rate of COVID-19 PCR testing: a discordant testing analysis

Jamil N. Kanji1,2*, Nathan Zelyas1,3, Clayton MacDonald4, Kanti Pabbaraju5, Muhammad Naeem Khan6, Abhaya Prasad6, Jia Hu7,8, Mathew Diggle1,3, Byron M. Berenger5,9 and Graham Tipples1,10,11

Abstract

Background: COVID-19 is diagnosed via detection of SARS-CoV-2 RNA using real time reverse-transcriptase polymerase chain reaction (rtRT-PCR). Performance of many SARS-CoV-2 rtRT-PCR assays is not entirely known due to the lack of a gold standard. We sought to evaluate the false negative rate (FNR) and sensitivity of our laboratory-developed SARS-CoV-2 rtRT-PCR targeting the envelope (E) and RNA-dependent RNA-polymerase (RdRp) genes.

Methods: SARS-CoV-2 rtRT-PCR results at the Public Health Laboratory (Alberta, Canada) from January 21 to April 18, 2020 were reviewed to identify patients with an initial negative rtRT-PCR followed by a positive result on repeat testing within 14 days (defined as discordant results). Negative samples from these discordant specimens were re-tested using three alternate rtRT-PCR assays (targeting the E gene and N1/N2 regions of the nucleocapsid genes) to assess for false negative (FN) results.

Results: During the time period specified, 95,919 patients (100,001 samples) were tested for SARS-CoV-2. Of these, 49 patients were found to have discordant results including 49 positive and 52 negative swabs. Repeat testing of 52 negative swabs found five FNs (from five separate patients). Assuming 100% specificity of the diagnostic assay, the FNR and sensitivity in this group of patients with discordant testing was 9.3% (95% CI 1.5–17.0%) and 90.7% (95% CI 82.6–98.9%) respectively.

Conclusions: Studies to understand the FNR of routinely used assays are important to confirm adequate clinical performance. In this study, most FN results were due to low amounts of SARS-CoV-2 virus concentrations in patients with multiple specimens collected during different stages of infection. Post-test clinical evaluation of each patient is advised to ensure that rtRT-PCR results are not the only factor in excluding COVID-19.

Keywords: SARS-CoV-2, COVID-19, Discordant testing, False negative rate

Background

Accurate case detection with rapid isolation and contact tracing form critical elements of the public health response to COVID-19. With most emerging infections, initially available nucleic acid tests (NATs) may lack data on the frequency of false negative results which can unnecessarily lead to repeated testing.

Studies of false-negative (FN) results from respiratory samples for SARS-CoV-2 are variable demonstrating FN rates (FNRs) ranging from 1 to 30% [1, 2]. FN results can occur for numerous reasons including suboptimal specimen collection, testing too early in the disease process, low analytic sensitivity, inappropriate specimen type, low viral load, or variability in viral shedding [3–9]. Implications of FN results can be significant, potentially leading to positive case clusters and negative outcomes [10]. Current guidance from the World Health Organization (WHO) and others calls for repeat testing (including sampling of the lower respiratory tract) in individuals...
who continue to display symptoms of COVID-19 with continued infection prevention measures [9, 11, 12]. The optimal interval of repeat testing is not clear with different studies suggesting a range from 1 to 6 days following the first negative test [13, 14].

The current study was designed to assess the FNR and sensitivity for the laboratory-developed test rtRT-PCR (LDT) used for frontline SARS-CoV-2 testing in Alberta, Canada, by determining the number of FN results in patients with repeat specimens submitted.

Methods
Setting, patients, and clinical samples
In the province of Alberta, Canada (population 4.4 million people), SARS-CoV-2 testing was conducted exclusively at the provincial Public Health Laboratory for symptomatic patients during the first four months of the pandemic [15–17]. The first case was confirmed on March 5, 2020 [18]. Test results and patient demographics were extracted from the laboratory information system to identify patients between January 21 and April 18, 2020, with an initial negative SARS-CoV-2 result followed by a positive result on repeat testing within 14 days (one incubation period) hereon defined as discordant test results [11].

Acceptable specimens for SARS-CoV-2 testing included nasopharyngeal (NP), oropharyngeal (OP), deep nasal turbinate swabs, endotracheal aspirates, and bronchoalveolar lavages (see Additional file 1: Table S1). All collection kits were internally validated prior to use.

SARS-CoV-2 RNA detection
Nucleic acid extraction was performed on one of several platforms (see Additional file 1: Table S1). A LDT rtRT-PCR targeting the envelope (E) and RNA-dependent RNA-polymerase (RdRp) genes was used to detect SARS-CoV-2 RNA [19]. Samples with cycle threshold (Ct) values > 35 cycles were repeated in duplicate and considered positive if ≥ 2 of three results had an amplification curve. Invalid was used to refer to samples with PCR run errors such as instrument or internal control failure. The assay parameters and comparison to other assays used across Canada has been published [19, 20].

The negative samples from sets of discordant specimens were re-tested by rtRT-PCR for SARS-CoV-2 to evaluate for FNs. This was carried out by extracting nucleic acid from the original sample followed by testing using assays targeting three different genes: the E gene (using only the E gene target from the LDT in a singleplex format) and the N1/N2 portions of the nucleocapsid gene (see Additional file 1: Table S1) [21]. Evaluation of the CDC N1/N2 assay compared to the LDT demonstrated 94% positive agreement (95% CI 87.7–100%) and 100% negative agreement (see Additional file 2: Table S2).

The discordant samples were retrieved from storage at −70 °C and underwent one freeze–thaw cycle. Samples that had tested positive were assumed to be true positives (based on the validation study of the LDT assay demonstrating analytic specificity of 100%) [19]. A negative sample was considered to be a FN if repeat testing yielded a positive result for ≥ 2 of three gene targets (E gene, N1, and/or N2).

Evaluation of discordant swab quality
All swab sets identified as discordant were tested for the presence of human ribonuclease P (RNAse P) using an RT-PCR assay (see Additional file 1: Table S1) [21].

Statistical analysis
Statistical comparison of parametric variables was done using independent t-tests and non-parametric variables using the Wilcoxon matched-pairs signed rank test. Data analyses were conducted in Stata 14.2 software (Statacorp LP, 2015, College Station, USA).

Results
Between January 21 and April 18, 2020, 100,001 COVID-19 tests (95,919 patients) were completed with 1954 (2%) individual cases confirmed (see Additional file 3: Figure S1). Including repeat tests, the overall positivity rate was 2.2%.

Forty-nine (0.05%) were found to have discordant results (total 101 swabs including 46 patients with two swabs and 3 patients with three swabs). The median age of these patients was 72 years (range 25–97) with 69.4% being female and 26.5% requiring hospitalization (Table 1).

All 101 discordant swabs were available for further evaluation (herein identified as swab 1, swab 2, and swab 3) (Table 2). Original testing results of these 49 patients showed: swab 1 for all 49 patients was negative; swab 2 for 46/49 patients was positive, and swab 3 was positive for 3/3 patients. Repeat testing of swab 1 for each of the 49 patients using a combination of three alternate assays revealed five FN results (Table 2). Of these, 3/5 were NP swabs in UTM and 2/5 were Aptima® swabs used for deep nasal sampling. Ct values for repeat testing of swab 1 specimens among the three different assays ranged from 32.7 to 38.8 cycles (median 35.5). Five swab 1 specimens re-tested positive on the E gene assay and the CDC N2 assay; two swab 1 specimens re-tested positive by all three alternate assays. The mean times of collection (in days) between swab 1 and swab 2 for the FN and non-FN discrepant specimens were 6.1 (p = 0.06) and 3.3 (p = 0.20), respectively.
No significant differences in the Ct values for human RNase P were noted between swabs 1, 2, and 3 (see Additional file 4: Figure S2; all \( p \)-values > 0.05).

From the five FN specimens, 4/5 had swab 1 collected on or the day after date of symptom onset (DSO) (Table 3). The maximum duration between DSO and swab 1 was 9 days and swab 2 was eleven days. Swab 2 for all five patients was collected post-DSO (4–11 days). All patients with FN results had community-acquired SARS-CoV-2 infection; three were healthcare workers and three had exposure to a confirmed COVID-19 case. Based on the additional testing conducted, 5/101 negative swabs were considered FNs with 49/101 presumed to be true positives (TPs). Therefore, FNR (FN/[FN + TP]) in this subset of patients with discordant swabs is 9.3% (95% CI 1.5–17.0%). By extension, the sensitivity (1-FNR) of testing in this subset of discordant swabs is 90.7% (95% CI 82.6–98.9%).

**Discussion**

The major strength of this study lies in the large sample size (100,001 SARS-CoV-2 rtRT-PCR tests from 95,919 patients) from which discordant results were identified. Discordant results were found for 0.05% of all patients tested. Based on re-testing of 49 patients with discordant results, the FNR and sensitivity of our LDT in this subgroup of patients was approximately 9.3% and 90.7%, respectively.

The FNR calculated from our data analysis is comparable to other reports. Data from earlier in the pandemic reported FNRs of up to 30% [6] with a systematic review on the topic reporting ranges from 2 to 29% [2]. A large study from New York evaluating the clinical performance of SARS-CoV-2 molecular testing found that on average up to 17% of positives were missed by the first test [22], while another American study reported a FNR of 3.5% in patients with discordant swab results within a 7-day period [1]. Two other studies have estimated sensitivities ranging from 89 to 94.6% [22, 23].

In our study, specimen quality was not considered a contributing factor given human DNA content did not differ significantly across all the swabs. A similar approach using RNase P as a surrogate for quality of

| Variable | % |
|----------|---|
| Age (years) | 72 |
| Range | 25–97 |
| Sex | 15 30.6 |
| Female | 34 69.4 |
| Exposure to a known case | 38 77.6 |
| Yes | 18.4 |
| Acquisition | 9 38 |
| Health care | 18.4 |
| Community | 77.6 |
| Unknown | 4.0 |
| Healthcare worker | 14.3 |
| Hospitalization | 26.5 |
| Travel history | 8.2 |

| Table 1 Demographic characteristics of 49 patients with discordant swab results for COVID-19 |

| Swab 1 (n = 49 swabs) | Swab 2 (n = 49 swabs) | Swab 3 (n = 3 swabs) |
|-----------------------|-----------------------|----------------------|
| NP swab in UTM (%) | 29 (59.2) | 32 (65.3) | 2 (66.7) |
| Deep nasal turbinate swab (%) | 20 (40.8) | 9 (18.4) | 1 (33.3) |
| Oropharyngeal swab (%) | 0 | 8 (16.3) | 0 |
| Original swab test result (E gene Ct median; range) | 49 negative | 46 positive 3 negative (19.5; 13.0–35.5) (22.4; 16.2–37.8) | 3 positive 0 negative (13.9; 12.5–24.7) (16.5; 14.8–28.0) |
| (RdRp gene Ct median; range) | | | |
| Positive result on E gene assay (Ct range) | 5/49 (34.1–37.9) | 0/3 | ND |
| Positive result on CDC N1 assay (Ct range) | 2/49 (32.7–35.9) | 0/3 | ND |
| Positive result on CDC N2 assay (Ct range) | 5/49 (33.2–38.8) | 0/3 | ND |

Ct cycle threshold (cycles), E envelope, LDT laboratory developed test, ND not done, NP nasopharyngeal, RdRp RNA dependent RNA polymerase, UTM universal transport media

\( a \) Testing done using E gene/RdRp gene LDT SARS-CoV-2 rtRT-PCR

\( b \) Repeat testing conducted only on negative swabs to evaluate for false-negative results

\( c \) Only 3 of 49 patients had a third swab done
Table 3 Timeline (in days) of swab collection and epidemiologic information of five false negative discordant specimens

| Patient | −5 | −4 | −3 | −2 | −1 | 0  | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | Exposed to confirmed case? | Healthcare worker? | Community-or healthcare-acquired? |
|---------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-------------------------------|------------------|---------------------------------|
| 1       | DSO | S1 | S2 |    |    |    |    |    |    |    |    |    |    |    |    |    | Yes                          | Yes              | Community-acquired            |
| 2       | DSO | S1 | S2 |    |    |    |    |    |    |    |    |    |    |    |    |    | No                           | Yes              | Community-acquired            |
| 3       | DSO | S1 | S2 |    |    |    |    |    |    |    |    |    |    |    |    |    | Yes                          | No               | Community-acquired            |
| 4       | DSO | S1 | S2 |    |    |    |    |    |    |    |    |    |    |    |    |    | Unknown                      | No               | Community-acquired            |
| 5       | S1  |    |    | S2 |    |    |    |    |    |    |    |    |    |    |    |    | Yes                          | Yes              | Community-acquired            |

Negative numbers indicate days before DSO; day 0 indicates DSO; positive numbers indicate days post-DSO

DSO date of symptom onset, S1 swab 1, S2 swab 2
swab collection has been used in several other studies [3, 23, 24].

The five FNs were likely caused by changes in viral load and shedding over time. Based on Ct values, all FNs were found to have low levels of viral RNA. Four of five FN samples had early collections related to the DSO (from 5 days prior to symptom onset to 2 days post-symptom onset). The other FN sample was collected 9 days post-symptom onset with the swab found to be positive for this patient with routine testing having been collected 2 days later, which could be related to variable shedding after the acute phase of infection [8]. Variable shedding dynamics have also been noted by authors of a pooled analysis of 1330 samples with FNR estimated as 20% at three days post DSO, 38% on the DSO, and 67% on the day prior to DSO [4].

Three of five FN swabs were collected using an NP flocked swab in UTM and the other two were collected using the Aptima® swab and transport medium. While this may indicate that these swab types and media did not influence the FNR, more data is needed to support this. However, one study indicated that Aptima® products are as good or better than routine flocked NP/UTM swabs for detecting SARS-CoV-2, attributed in part to the preservatives in the Aptima® transport solution preventing RNA degradation [25].

The principal limitations of this study are its retrospective nature and that FN samples were biased towards patients undergoing repeat swab collection, likely due to high suspicion of COVID-19. Ideally, a cohort of negative patients would be tested using multiple NAT tests and re-tested prospectively, but this poses logistical challenges and would require a large number of patients to be screened. Another limitation is the assumption that all positives by the local LDT were true positives. However, the analytical specificity of the LDT is reported as 100% [19] and it demonstrated a high negative percent agreement with the CDC N1/N2 assay. Most other SARS-CoV-2 rtRT-PCR assays have shown high clinical specificities, making this a reasonable assumption [26, 27].

Conclusions
This work adds to the literature by demonstrating that the FNR of SARS-CoV-2 molecular assays is low [1, 27, 28] and subject to viral load dynamics over time. However, the interpretation of COVID-19 test results should be conducted in the overall context of each patient’s clinical presentation [9, 29], with repeat testing advised should post-test probability upon follow-up clinical evaluation remain high.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12985-021-01489-0.

Additional file 1: Table S1. Commercial products utilised in the process of SARS-CoV-2 rtRT-PCR testing

Additional file 2: Table S2. Comparison of LDT to CDC N1/N2 assay using 100 specimens selected randomly from those tested

Additional file 3: Figure S1. Results of COVID-19 testing from the first 100,001 tests completed

Additional file 4: Figure S2. RNase P detection (based on Ct value) between swab sets (a) on 46 patients who had two swabs that were discordant; and (b) on 3 patients with three swabs.

Abbreviations
Ct: Cycle threshold; DSO: Date of symptom onset; FN: False-negative; FNR: False-negative rate; LDT: Laboratory developed test; NATs: Nucleic acid tests; NP: Nasopharyngeal; OP: Oropharyngeal; PCR: Polymerase chain reaction; RdRp: RNA-dependent RNA-polymerase; RNase P: Human ribonuclease P; rtRT-PCR: Real-time reverse-transcriptase polymerase chain reaction; UTM: Universal transport media; WHO: World Health Organization.

Acknowledgements
There are no acknowledgements.

Authors’ contributions
JNK participated in the formal data analysis, writing of the original draft, as well as review and editing of the manuscript. NZ conceptualised the idea for the project and contributed to data curation, validation, project administration, methodology, investigation, as well as review and editing of the manuscript. CM contributed to the conceptualisation, methodology, formal analysis, data curation, writing of the original draft of the manuscript, as well as review and editing of manuscript. KP contributed to the methodology, investigation, provided laboratory resources, spent time with data curation, as well as review and editing of the manuscript. AP contributed to data curation as well as writing of the original draft, as well as review and editing of the manuscript. MD provided support with regards to manuscript review and editing along with initial critique of the project writeup. BMB provided support with regards to review and editing of the manuscript. GP provided support with regards to supervision, project administration, and review and editing of the manuscript. All authors read and approved the final manuscript.

Funding
This work did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate
The Human Research Ethics Board at the University of Alberta approved this study protocol (Reference Number Pro00100001).

Consent to publication
Not applicable.

Competing interests
All authors declare that there are no conflicts of interests or competing interests with regards to the material in this manuscript.

Author details
1 Public Health Laboratory, Alberta Precision Laboratories, University of Alberta Hospital, 8440 – 112 Street, Edmonton, AB T6G 2B7, Canada. 2 Division of Infectious Diseases, Department of Medicine, University of Alberta, Room
References

1. Long DR, Gombar S, Hogan CA, Greninger AL, O'Reilly-Shah V, Blyson-Cahn C, et al. Occurrence and timing of subsequent SARS-CoV-2 RT-PCR positivity among initially negative patients. Clin Infect Dis. 2020. https://doi.org/10.1093/cid/ciaa722.

2. Arevalo-Rodriguez J, Butrago-Garcia D, Simancas-Racines D, Zambrano-Achig P, del Campo R, Ciapponi A, et al. False-negative results of initial RT-PCR assays for COVID-19: a systematic review. medRxiv. 2020. https://doi.org/10.1101/2020.04.16.20066787.

3. Kinloch NN, Ritchie G, Brumme CJ, Dong W, Dong W, Lawson T, et al. Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention. Research use only 2019-novel coronavirus (2019-nCoV) Real-time RT-PCR primers and probes. https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel -primer-probes.html. Accessed 6 July 6 2020.

4. Prinz A. False negatives and reinfections: the challenges of SARS-CoV-2 RT-PCR testing. https://asm.org/Articles/2020/April/False-Negatives-and-Reinfections-the-Challenges-of. Accessed 5 July 2020.

5. Pan Y, Long L, Zhang D, Yuan T, Cui S, Yang P, et al. False-negative nucleic acid testing results for severe acute respiratory syndrome coronavirus 2 from thermal inactivation of samples with low viral loads. Clin Chem. 2020;66(6):794–801. https://doi.org/10.1093/clinchem/hvaa091.

6. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med. 2020;26(5):672–S. https://doi.org/10.1038/s41591-020-0869-5.

7. Thompson MJ, Ferber JR, Ondoul R, David D, Shifflett P, Meece JK, et al. Results of a pilot study using self-collected mid-turbinate nasal swabs for detection of influenza virus infection among pregnant women. Influenza Other Respir Viruses. 2015;9(3):155–60. https://doi.org/10.1111/irv.12309.

8. Avannis-Aghajani E, Sarkissian A, Fernandez F, Avannis-Aghajani A. Validation of the hologic aptima unisex and multitest specimen collection kits used for endocervical and male urethral swab specimens (aptima swabs) for collection of samples from SARS-CoV-2-infected patients. J Clin Microbiol. 2020. https://doi.org/10.1128/JCM.00753-20.

9. Afzal A. Molecular diagnostic technologies for COVID-19: limitations and challenges. J Adv Res. 2020. https://doi.org/10.1016/j.jare.2020.08.002.

10. Thompson MJ, Ferber JR, Ondoul R, David D, Shifflett P, Meece JK, et al. Results of a pilot study using self-collected mid-turbinate nasal swabs for detection of influenza virus infection among pregnant women. Influenza Other Respir Viruses. 2015;9(3):155–60. https://doi.org/10.1111/irv.12309.

11. Doll ME, Pryor D, Mackey D, Doem CD, Bryson A, Bailey P, et al. Utility of retesting for diagnosis of SARS-CoV-2 COVID-19 in hospitalized patients: impact of the interval between tests. Infect Control Hosp Epidemiol. 2020. https://doi.org/10.1017/ice.2020.224.

12. Ait-Yaziz H, Zhu H, Zhan C, Chen C, Lv W, et al. Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology. 2020;296(2):E32–40. https://doi. org/10.1148/radiol.2020200642.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.