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.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Harmonization of SARS-CoV-2 reverse transcription quantitative PCR tests to the first WHO international standard for SARS-CoV-2 RNA

Malaya K. Sahoo \textsuperscript{a}, ChunHong Huang \textsuperscript{a}, Mamdouh Sibai \textsuperscript{a}, Daniel Solis \textsuperscript{a}, Benjamin A. Pinsky \textsuperscript{a,b,*}

\textsuperscript{a} Department of Pathology, Stanford University School of Medicine, Stanford, CA, United States
\textsuperscript{b} Division of Infectious Diseases and Geographic Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA, United States

**ABSTRACT**

**Background:** Cycle threshold (Ct) values from SARS-CoV-2 reverse transcription quantitative PCR (RT-qPCR) tests are used to measure viral burden. Calibration to the First WHO International Standard for SARS-CoV-2 RNA may improve quantitative inter-assay agreement.

**Methods:** WHO standard was tested using four emergency use authorized RT-qPCRs to generate calibration curves and evaluate Ct value differences. Harmonization of two assays, Cepheid Xpert Xpress SARS-CoV-2 targeting E and nucleocapsid (N2) [Xpert (E) and Xpert (N2)] and a laboratory-developed test targeting E [LDT (E)], was assessed using 93 positive upper respiratory samples. Platform (target) pairs were compared via Bland-Altman analysis and Passing-Bablok regression.

**Results:** Ct values with the WHO standard were comparable across platforms and targets, except Xpert (N2) for which the mean difference was a median of 3.68 cycles (Interquartile Range, IQR = 3.23 to 3.76 cycles) greater than other platform (target) pairs. Using clinical samples, the mean difference of Xpert (N2) to LDT (E) was 3.64 cycles (95% Confidence Interval, CI = 1.51 to 5.76). After calibration, the mean difference of Xpert (N2) to LDT (E) was 0.08 log\textsubscript{10} IU/mL (95% CI = -0.56 to 0.71) and the regression was $y = 1.00x + 0.08$ (95% CI slope = 0.93 to 1.07, 95% CI intercept = 0.28 to 0.42).

**Conclusions:** Calibration to the WHO standard resulted in the harmonization of two RT-qPCR tests, whereas analysis by Ct value alone may have led to erroneous quantitation. Harmonization to the WHO standard has the potential to improve the generalizability of clinical associations with SARS-CoV-2 RNA levels.

1. Introduction

Nucleic acid amplification tests (NAATs) serve as the reference methods for detection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) RNA in respiratory tract samples. Though many NAATs utilize reverse transcription quantitative PCR (RT-qPCR) and have the capability of quantitation, to date, all but one has been emergency use authorized by the US Food and Drug Administration (FDA) for qualitative testing only. Nevertheless, quantitative testing, primarily using cycle threshold (Ct) values, has been investigated to predict the development of severe COVID-19 and to identify individuals with active infection [1–3]. These studies have been generally inconsistent, at least in part due to the use of multiple different RT-qPCR methods. SARS-CoV-2 proficiency testing has revealed that Ct values can vary substantially within and between RT-qPCR methods [4]. Calibration of quantitative assays using an international reference standard is an approach that has been widely used to harmonize tests that measure viral burden in clinical samples [5].

In this study, we utilized the First WHO International Standard for SARS-CoV-2 RNA (NIBSC code 20/146) to calibrate four RT-qPCR assays: the Stanford Health Care Clinical Virology Laboratory SARS-CoV-2 RT-PCR Emergency Use Authorized Laboratory Developed Test [6,7], Hologic Panther Fusion SARS-CoV-2 Assay, PerkinElmer SARS-CoV-2 Real-time RT-PCR Assay, and Cepheid Xpert Xpress SARS-CoV-2. We then assessed harmonization of the in-house RT-qPCR with the Xpert Xpress using 93 SARS-CoV-2 positive upper respiratory samples.

2. Materials and methods

**Ethics Statement.** This study was conducted with Stanford Institutional Review Board approval (protocol 57519), and individual consent was waived.

**RT-qPCR Methods.** The following four assays were included in the study: Stanford Health Care Clinical Virology Laboratory SARS-CoV-2 RT-PCR Emergency Use Authorized Laboratory Developed Test, which targets the envelope (E) gene [platform (target); LDT (E)]; Hologic Panther Fusion, which targets two regions in ORF1ab [Fusion (ORF1ab)]; PerkinElmer SARS-CoV-2 Real-time RT-PCR Assay, which targets the nucleocapsid (N) and ORF1ab genes [PE (N) and PE (ORF1ab)]; and Cepheid Xpert Xpress SARS-CoV-2, which targets the E and N genes [Xpert (E) and Xpert (N2)]. Testing was performed according to the Emergency Use Authorization package inserts.

* Corresponding author at: Department of Pathology, Stanford University School of Medicine, Stanford, CA, United States.
E-mail address: bpinsky@stanford.edu (B.A. Pinsky).

https://doi.org/10.1016/j.jcv.2022.105242
Received 10 June 2022; Received in revised form 7 July 2022; Accepted 14 July 2022
Available online 16 July 2022
1386-6532/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Reference Material. The First WHO International Standard for SARS-CoV-2 RNA (code 20/146) was obtained from the National Institute for Biological Standards and Controls (NIBSC, Hertfordshire, UK). The WHO material was reconstituted in nuclease free water according to the package insert, and was subsequently diluted to 6.0, 5.0, 4.0, and 3.0 IU/mL in viral transport media (VTM) from pooled SARS-CoV-2-negative upper respiratory swab specimens.

Analytical and Clinical Evaluation. The WHO dilution series was tested using the four assay methods in triplicate over 3 days, except Cepheid Xpert Xpress SARS-CoV-2 which was tested in duplicate on day 3. Log\(^{10}\) IU/mL concentrations were calculated from the cycle threshold (Ct) values using ordinary least squares regression in Microsoft Excel (Microsoft, Redmond, WA) (Supplemental Table 1). Precision was analyzed in R, using a custom script implementing the formula described in Chesher, 2008 [8] (Supplemental Tables 2 and 3).

Clinical Evaluation. Ninety-three residual upper respiratory swab specimens collected in VTM were tested in parallel using the Stanford Clinical Virology Laboratory SARS-CoV-2 RT-PCR LDT and Cepheid Xpert Xpress SARS-CoV-2, LDT (E), Xpert (E), and Xpert (N2) Ct values and log\(^{10}\) IU/mL concentrations were compared via Bland-Altman analysis which was performed using Prism version 6.0 g (GraphPad, La Jolla, CA). Log\(^{10}\) IU/mL concentrations for each platform (target) pair were compared using Passing-Bablok regression in R using the mcr package (https://cran.r-project.org/web/packages/mcr/index.html).

### Table 1

| Platform (Target) Nominal concentration log\(^{10}\) IU/mL | Overall mean Ct | Mean difference Platform (Target) – LDT (E) | Mean difference Xpert (N2) - Platform (Target) |
|----------------------------------------------------------|-----------------|---------------------------------------------|-------------------------------------------------|
| LDT (E) 3.0                                             | 33.02           | 0.00                                        | 4.24                                            |
| Fusion (ORF1ab)                                        | 3.0             | 33.44                                       | 0.42                                            |
| PE (N) 3.0                                             | 33.55           | 0.53                                        | 3.72                                            |
| PE (ORF1ab)                                           | 3.0             | 33.36                                       | 0.34                                            |
| Xpert (E) 3.0                                         | 3.0             | 33.46                                       | 0.44                                            |
| Xpert (N2) 3.0                                        | 3.0             | 37.26                                       | 4.24                                            |
| LDT (E) 4.0                                             | 29.86           | 0.00                                        | 3.69                                            |
| Fusion (ORF1ab)                                        | 4.0             | 30.41                                       | 0.55                                            |
| PE (N) 4.0                                             | 29.95           | 0.09                                        | 3.60                                            |
| PE (ORF1ab)                                           | 4.0             | 29.86                                       | 0.00                                            |
| Xpert (E) 4.0                                         | 4.0             | 30.29                                       | 0.43                                            |
| Xpert (N2) 4.0                                        | 4.0             | 33.55                                       | 3.69                                            |
| LDT (E) 5.0                                             | 26.54           | 0.00                                        | 3.75                                            |
| Fusion (ORF1ab)                                        | 5.0             | 27.14                                       | 0.60                                            |
| PE (N) 5.0                                             | 26.93           | 0.39                                        | 3.36                                            |
| PE (ORF1ab)                                           | 5.0             | 26.61                                       | 0.07                                            |
| Xpert (E) 5.0                                         | 5.0             | 27.30                                       | 0.76                                            |
| Xpert (N2) 5.0                                        | 5.0             | 30.29                                       | 3.75                                            |
| LDT (E) 6.0                                             | 23.19           | 0.00                                        | 3.86                                            |
| Fusion (ORF1ab)                                        | 6.0             | 23.91                                       | 0.72                                            |
| PE (N) 6.0                                             | 23.56           | 0.37                                        | 3.49                                            |
| PE (ORF1ab)                                           | 6.0             | 23.37                                       | 0.17                                            |
| Xpert (E) 6.0                                         | 6.0             | 24.08                                       | 0.88                                            |
| Xpert (N2) 6.0                                        | 6.0             | 27.05                                       | 3.86                                            |

LDT, Laboratory-Developed Test; Fusion, Hologic Panther Fusion; PE, PerkinElmer; Xpert, Cepheid GeneXpert; E, Envelope gene; ORF1ab, Open Reading Frame-1ab; N, Nucleocapsid gene.

3. Results

The Ct values observed using the WHO standard were generally comparable across platforms and targets, except for Xpert (N2) for which the mean difference was a median of 3.68 cycles (Interquartile Range, IQR = 3.23 to 3.76 cycles) greater than the other platforms and targets (Table 1). In the absence of calibration, such a Ct difference may erroneously be considered 12.8-fold (2\(^{3.68}\)) or 1.1 log\(^{10}\) lower in concentration. When excluding Xpert (N2), the median of the mean difference of other platforms and targets was 0.42 cycles (IQR = 0.30 to 0.56...
cycles) greater than LDT (E).

Given the Ct differences observed between Xpert (N2) and other platforms and targets with the WHO standard, the Xpert and LDT platforms were selected for further evaluation using 93 residual clinical upper respiratory specimens. Comparison of Xpert (E) to LDT (E) Ct values showed a mean difference of 0.99 cycles [95% Confidence Interval (CI) = -1.09 to 3.06] (Fig. 1A). The mean difference of Xpert (N2) to LDT (E) and Xpert (N2) to Xpert (E) were 3.64 cycles (95% CI =1.51 to 5.76) and 2.65 cycles (95% CI = 1.53 to 3.77), respectively (Fig. 1B and 1C).

To evaluate the impact of harmonization, the regression equations obtained from testing the WHO standard were used to calculate log_{10}...
IU/mL concentrations from the Ct values of the clinical specimens. When harmonized to log_{10} IU/mL, the quantitative agreement as measured by mean difference, approached zero for all comparisons: Xpert (E) to LDT (E) = -0.12 log_{10} IU/mL (95% CI = -0.76 to 0.53); Xpert (N2) to LDT (E) = 0.08 log_{10} IU/mL (95% CI = -0.56 to 0.71); Xpert (N2) to Xpert (E) = 0.19 log_{10} IU/mL (95% CI = -0.13 to 0.52) (Fig. 2A-C). Similarly, Passing-Bablok regression of the harmonized values showed the following equations: Xpert (E) to LDT (E), y = 1.00x + 0.08 (95% CI slope = 0.93 to 1.08, 95% CI intercept = 0.25 to 0.48); Xpert (N2) to LDT (E), y = 1.00x + 0.08 (95% CI slope = 0.93 to 1.07, 95% CI intercept = 0.28 to 0.42); and Xpert (N2) to Xpert (E), y = 1.00x + 0.16 (95% CI slope = 0.96 to 1.04, 95% CI intercept = 0.01 to 0.35) (Fig. 2D-F). Based on this analysis, the 95% confidence intervals of the slope for these WHO harmonized platform (target) pairs included one, demonstrating the absence of proportional bias. However, these regressions revealed a slight positive systematic bias; the 95% confidence intervals of the intercepts were greater than zero for all platform (target) comparisons.

Finally, the WHO standard showed commutability among these platform (target) pairs, with quantitation of the WHO dilutions falling within the 95% confidence intervals of both the Bland-Altman and Passing-Bablok analyses of the clinical samples (Fig. 2A-F).

4. Discussion

In this work we calibrate four SARS-CoV-2 RT-qPCR tests to the WHO International Standard for SARS-CoV-2 RNA (NIBSC code 20/146). Then using 93 clinical upper respiratory specimens we demonstrate that this harmonization improves quantitative agreement between platform (target) pairs compared to estimated viral burden using Ct values alone. Finally, we show that the WHO standard is commutable, or quantitated similarly to clinical specimens, using the LDT and Xpert tests. Limitations of this study include its single-institution design, the assessment of a small number of the available SARS-CoV-2 RT-qPCR assays, and the inclusion of a single sample matrix. Future work will require multisite studies to evaluate commutability and the impact of calibration to the WHO standard on additional RT-qPCR platforms and specimen types, including saliva and bronchoalveolar lavage fluid.

Throughout the pandemic, Ct values have been used as a surrogate for viral burden, and myriad Ct thresholds have been proposed to predict COVID-19 severity, assess transmissibility as approximated via viral culture, inform removal of isolation, and evaluate the performance of rapid antigen tests [2,3,9]. Though these thresholds may be internally valid, particularly if a single RT-qPCR is used, the generalizability of these results to different institutions performing different assays are notably limited. The situation is further complicated by the numerous platforms that individual laboratories have been required to implement to maintain testing during supply chain interruptions and to provide turnaround times appropriate to the clinical context [10]. Harmonization to the WHO standard, and in the future, secondary standards calibrated to the WHO material, provide opportunities to improve quantitative agreement between SARS-CoV-2 RT-qPCR platforms. Such harmonization may help accelerate the development of broadly applicable thresholds and expand the clinical utility of quantitative SARS-CoV-2 testing.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jcv.2022.105242.

References

[1] D.D. Rhoads, B.A. Pinsky, The truth about SARS-CoV-2 cycle threshold values is rarely pure and never simple, Clin. Chem. 68 (2021) 16–18.
[2] V.P. Shah, W.H. Farah, J.C. Hill, L.C. Hassett, M.J. Binninger, J.D. Yao, M. H. Murad, Association between SARS-CoV-2 cycle threshold values and clinical outcomes in patients with COVID-19: a systematic review and meta-analysis, Open Forum Infect. Dis. 8 (2021) eoh453.
[3] M.J. Binninger, Can testing predict SARS-CoV-2 infectivity? The potential for certain methods to be surrogates for replication-competent virus, J. Clin. Microbiol. 59 (2021), e0046921.
[4] D. Rhoads, D.R. Peaper, R.C. She, F.S. Nolte, C.M. Wojewoda, N.W. Anderson, B. S. Pitt, College of American pathologists (CAP) microbiology committee perspective: caution must be used in interpreting the cycle threshold (Ct) value, Clin. Infect. Dis. 72 (2021) e683–e686.
[5] R.T. Hayden, A.M. Caliendo, Persistent challenges of interassay variability in transplant viral load testing, J. Clin. Microbiol. 58 (2020).
[6] C.A. Hogan, M.K. Sahoo, C. Huang, N. Garamani, B. Stevens, J. Zehnder, B. A. Pinsky, Comparison of the panther fusion and a laboratory-developed test targeting the envelope gene for detection of SARS-CoV-2, J. Clin. Virol. 127 (2020), 104383.
[7] V.M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D.K. Chu, T. Bleicker, S. Brunsink, J. Schneider, M.L. Schmidt, D.G. Mulders, B.L. Haagmans, B. van der Veer, S. van den Brink, L. Wijman, G. Goderski, J.L. Romerte, J. Ellis, M. Zambon, M. Peiris, H. Goossens, C. Reusken, M.P. Koopmans, C. Drosten, Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR, Euro Surveill. (2020) 25.
[8] M.B. Miller, Understanding, verifying, and implementing emergency use authorization molecular diagnostics for the detection of SARS-CoV-2 infection, Cochrane Database Syst. Rev. 3 (2021), CD013705.
[9] D. Rhoads, E. Dinh, E. Vianin, C. Davenport, D. Emperador, Y. Takwoingi, J. Cunningham, S. Beese, J. Domen, J. Dretzke, D. The present study included a small number of the available SARS-CoV-2 RNA (NIBSC code 20/146). Then using 93 clinical upper respiratory specimens we demonstrate that this harmonization improves quantitative agreement between platform (target) pairs compared to estimated viral burden using Ct values alone. Finally, we show that the WHO standard is commutable, or quantitated similarly to clinical specimens, using the LDT and Xpert tests. Limitations of this study include its single-institution design, the assessment of a small number of the available SARS-CoV-2 RT-qPCR assays, and the inclusion of a single sample matrix. Future work will require multisite studies to evaluate commutability and the impact of calibration to the WHO standard on additional RT-qPCR platforms and specimen types, including saliva and bronchoalveolar lavage fluid.

Throughout the pandemic, Ct values have been used as a surrogate for viral burden, and myriad Ct thresholds have been proposed to predict COVID-19 severity, assess transmissibility as approximated via viral culture, inform removal of isolation, and evaluate the performance of rapid antigen tests [2,3,9]. Though these thresholds may be internally valid, particularly if a single RT-qPCR is used, the generalizability of these results to different institutions performing different assays are notably limited. The situation is further complicated by the numerous platforms that individual laboratories have been required to implement to maintain testing during supply chain interruptions and to provide