

1 Title: “Informative Use of Cycle-Threshold Values to Account for Sampling Variability in  
2 Pathogen Detection”

3 Authors: Aidan M. Nikiforuk<sup>1,2</sup>, Agatha N. Jassem<sup>1,3</sup>

4 Affiliations:

5 <sup>1</sup>The British Columbia Center for Disease Control, Vancouver, BC, V5Z 4R4, Canada

6 <sup>2</sup> School of Population and Public Health, The University of British Columbia, Vancouver, BC, V6T  
7 1Z4, Canada

8 <sup>3</sup> Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver,  
9 BC V6T 1Z4, Canada

10

11

12

13

14

15

16 Word count (text): 673

17 Tables: 0

18 Figures: 2

19

20 **Corresponding Author**: Aidan M. Nikiforuk

21 email: [Aidan.Nikiforuk@bccdc.ca](mailto:Aidan.Nikiforuk@bccdc.ca)

22 telephone: 1-204-430-5084

23 Address: The British Columbia Center for Disease Control,  
24 Vancouver, BC, V5Z 4R4, Canada

25 **Keywords:** SARS-CoV-2, COVID-19, RT-qPCR, Diagnostic Validity, Standardization,  
26 Normalization, Sampling Variability

27  
28 **Abbreviations:** real-time quantitative polymerase chain reaction, RT-qPCR; cycle threshold, Ct;  
29 Ribonuclease P, RNaseP; Envelope, E; Pearson's correlation coefficient, PCC; simple linear  
30 regression, SLR.

31  
32 **Abstract:**

33 Nucleic acid amplification tests, like real-time polymerase chain reaction, are widely  
34 used for pathogen detection; however, their interpretation rarely accounts for sampling  
35 variability. Instead, cycle threshold values are often categorized reducing precision. We  
36 describe how pathogen cycle threshold values can be normalized to endogenous host gene  
37 expression to correct for sampling variability and compare the validity of this approach to  
38 standardization with a standard curve. Normalization serves as a valid alternative to  
39 standardization, does not require making a standard curve, increases precision, accounts for  
40 sampling variability, and can be easily applied to large clinical or surveillance datasets for  
41 informative interpretation.

42 **Introduction:**

43 Nucleic acid amplification tests like reverse transcription quantitative real-time  
44 polymerase chain reaction (RT-qPCR) are routinely used in clinical laboratories for pathogen  
45 detection [1]. The prototypical RT-qPCR assay includes two primer and probe sets (targets) for a  
46 specific pathogen and one primer probe set (target) for an endogenous host gene. Results are  
47 interpreted from a range of cycle-threshold (Ct) values, samples with a Ct value below a certain

48 predetermined cut-off are categorized as positive [2]. Categorizing Ct values transforms a  
49 quantitative measurement into a categorical one limiting precision of the measurement which  
50 may lead to misinterpretation. Yet, Ct values are broadly interpreted categorically because, it  
51 provides a simpler alternative to using them as numerical value, which requires making a  
52 standard curve and additional analysis.

53 To encourage more informative use of Ct values as a numerical value, we investigate  
54 using relative gene expression (normalization) in lieu of a standard curve (standardization) for  
55 interpretation [3]. Normalization has the benefits of re-purposing the host gene target,  
56 accounting for sampling variability by measuring the ratio of the pathogen target(s) to host  
57 target, it can be performed retrospectively and automated for large clinical or surveillance  
58 datasets.

#### 59 **Methods:**

60 To validate relative gene expression as a quantitative way to interpret Ct values we  
61 used a dataset of n = 212 clinical test results from persons who tested positive for COVID-19  
62 from 24/3/2020 to 9/5/2020. Specimens from persons seeking a diagnostic test for SARS-CoV-2  
63 infection were collected by nasopharyngeal swab and nucleic acid extraction was performed  
64 using the Viral RNA isolation kit on the MagMAX-96™ platform (ThermoFisher). Host (*RNaseP*)  
65 and viral gene (*Envelope*, *Nucleocapsid*) targets were assayed by RT-qPCR, as previously  
66 described. A 9-replicate, 5-fold, 1:10 dilution of SARS-CoV-2 synthetic RNA was used to make a  
67 standard curve and Ct values were transformed to log<sub>10</sub> GE/mL using simple linear  
68 regression[4]. Relative gene expression was calculated between the viral Envelope (*E*) and  
69 *RNaseP* targets using a derivation of the method proposed by Livak and Schmittgen ( $2^{-\Delta Ct}$ ) [5].

70 Pearson's correlation and simple linear regression were used to estimate the relationship  
71 between crude, normalised and standardized RT-qPCR measurements. Data analysis was  
72 performed using R Statistical Software Version 4.1.0, p-values of less than  $\alpha = 0.05$  were  
73 considered statistically significant.

#### 74 **Results:**

75 To illustrate how normalizing the pathogen gene target(s) to host gene target accounts  
76 for sampling variability we made a scatterplot of normalized SARS-CoV-2 viral load to crude viral  
77 *E* gene Ct (Figure 1). At a single Ct value (e.g. Ct 15) the quantifiable amount of SARS-CoV-2  
78 genome differs by 5-fold, which corresponds to 10,000 viral genomes or a 4-log change. This  
79 difference was undetectable without correcting for sampling variability and serves as an  
80 example of how interpretation of untransformed Ct values may lead to information bias and  
81 incorrect inference.

82 We then compared normalized SARS-CoV-2 viral load to standardized viral load to test  
83 if relative gene expression serves as a valid alternative to making a standard curve (Figure 2).  
84 Normalized viral load has a strong positive linear relationship with standardized measurements  
85 and variability seems homogenous across the range of observed values, indicating  
86 comparability (SLR,  $R^2 = 0.942$ ,  $P < 0.001$ ).

#### 87 **Discussion:**

88 We demonstrate that using relative gene expression to normalize pathogen load  
89 quantifies specimen sampling variability and serves as a valid alternative to analysis using a  
90 standard curve. Normalization has the benefits of: repurposing host gene targets, accounting

91 for sampling variability (reducing information bias), more informative use of Ct values as a  
92 numeric measure than a categorical one and can be retrospectively performed on large clinical  
93 datasets. Relative gene expression makes the assumption that the target host gene selected as  
94 an endogenous control possess equivalent inter-individual expression, which may not always be  
95 the case in those infected with a pathogen.

96           Given the well document disadvantages of interpreting Ct values categorically or as a  
97 crude numeric measure for pathogen detection[6], we recommend that clinicians consider  
98 making more informed use and account for sampling variability by normalization as described.

99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115

116 **References:**

- 117 [1] Kralik P, Ricchi M. A Basic Guide to Real Time PCR in Microbial Diagnostics: Definitions,  
118 Parameters, and Everything. *Front Microbiol* 2017;8.  
119 <https://doi.org/10.3389/fmicb.2017.00108>.
- 120 [2] Tom MR, Mina MJ. To Interpret the SARS-CoV-2 Test, Consider the Cycle Threshold  
121 Value. *Clinical Infectious Diseases* 2020;71:2252–4. <https://doi.org/10.1093/cid/ciaa619>.
- 122 [3] Bustin S. Absolute quantification of mRNA using real-time reverse transcription  
123 polymerase chain reaction assays. *J Mol Endocrinol* 2000;25:169–93.  
124 <https://doi.org/10.1677/jme.0.0250169>.
- 125 [4] Nikiforuk AM, Kuchinski KS, Twa DDW, Lukac CD, Sbihi H, Basham CA, et al. The  
126 contrasting role of nasopharyngeal angiotensin converting enzyme 2 (ACE2) transcription  
127 in SARS-CoV-2 infection: A cross-sectional study of people tested for COVID-19 in British  
128 Columbia, Canada. *EBioMedicine* 2021;66:103316.  
129 <https://doi.org/10.1016/j.ebiom.2021.103316>.
- 130 [5] Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time  
131 Quantitative PCR and the 2- $\Delta\Delta$ CT Method. *Methods* 2001;25:402–8.  
132 <https://doi.org/10.1006/meth.2001.1262>.
- 133 [6] Dahdouh E, Lázaro-Perona F, Romero-Gómez MP, Mingorance J, García-Rodríguez J. Ct  
134 values from SARS-CoV-2 diagnostic PCR assays should not be used as direct estimates of  
135 viral load. *Journal of Infection* 2021;82:414–51.  
136 <https://doi.org/10.1016/j.jinf.2020.10.017>.

137

138 **Ethical Approval:**

139 Ethical approval for the study was obtained from the University of British Columbia  
140 human ethics board (H20-01110). Written informed consent was not required as per ethics  
141 board approval. Participant data was de-identified prior to analysis; the results of this non-  
142 interventional observational study were not linked back to any identifying patient records. The  
143 study was deemed to be of minimal risk.

144

145 **Author Contributions:**

146 A.M. Nikiforuk: Conceived, designed, and performed the analysis, wrote the manuscript,  
147 reviewed, and edited the manuscript. A.N. Jassem: Conceived the analysis, discussed the  
148 results, reviewed, and edited the manuscript.

149 **Declaration of Competing Interest:**

150 The authors have no competing interests to declare.

151 **Funding Source:**

152 This work was funded by the Canadian Institutes of Health Research (#434951) and Genome  
153 British Columbia (COV-55).

154

155

156

157

158

159

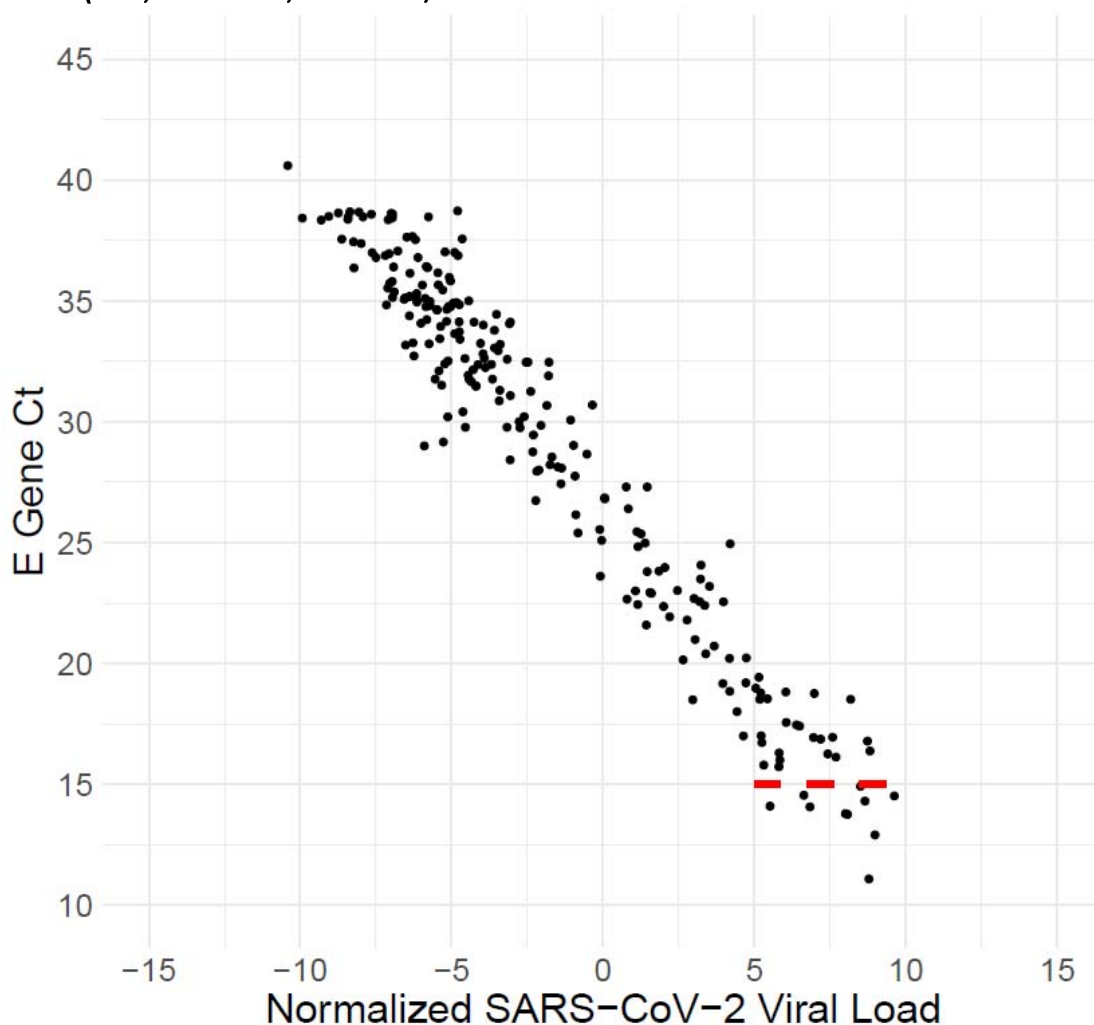
160

161

162

163

164 **Figure 1: Scatterplot Visualizing the Linear Relationship Between Normalized SARS-CoV-2**  
165 **Viral Load and SARS-CoV-2 E gene Cycle-Threshold Value, for n = 212 Nasopharyngeal**  
166 **Specimens (PCC,  $r = -0.970$ ,  $P < 0.001$ ).**



167  
168 Scatterplot of normalized SARS-CoV-2 viral load and *Envelope* gene (*E*) cycle threshold values (Ct) assayed by RT-  
169 qPCR in n = 212 nasopharyngeal specimens collected from people tested for COVID-19 in British Columbia. The  
170 transformed and untransformed variables show a strong negative relationship, quantified by Pearson's correlation  
171 coefficient ( $r = -0.970$ ,  $P < 0.001$ ). The red segmented line shows that at a constant E gene Ct value of 15, SARS-  
172 CoV-2 viral load may vary by a 5-fold difference which corresponds to 10,000 viral genomes. This dramatic  
173 difference would not have been detectable without adjusting for sampling variability in the normalization process;  
174 therefore, Ct values alone possess little quantitative information regarding viral load.

175

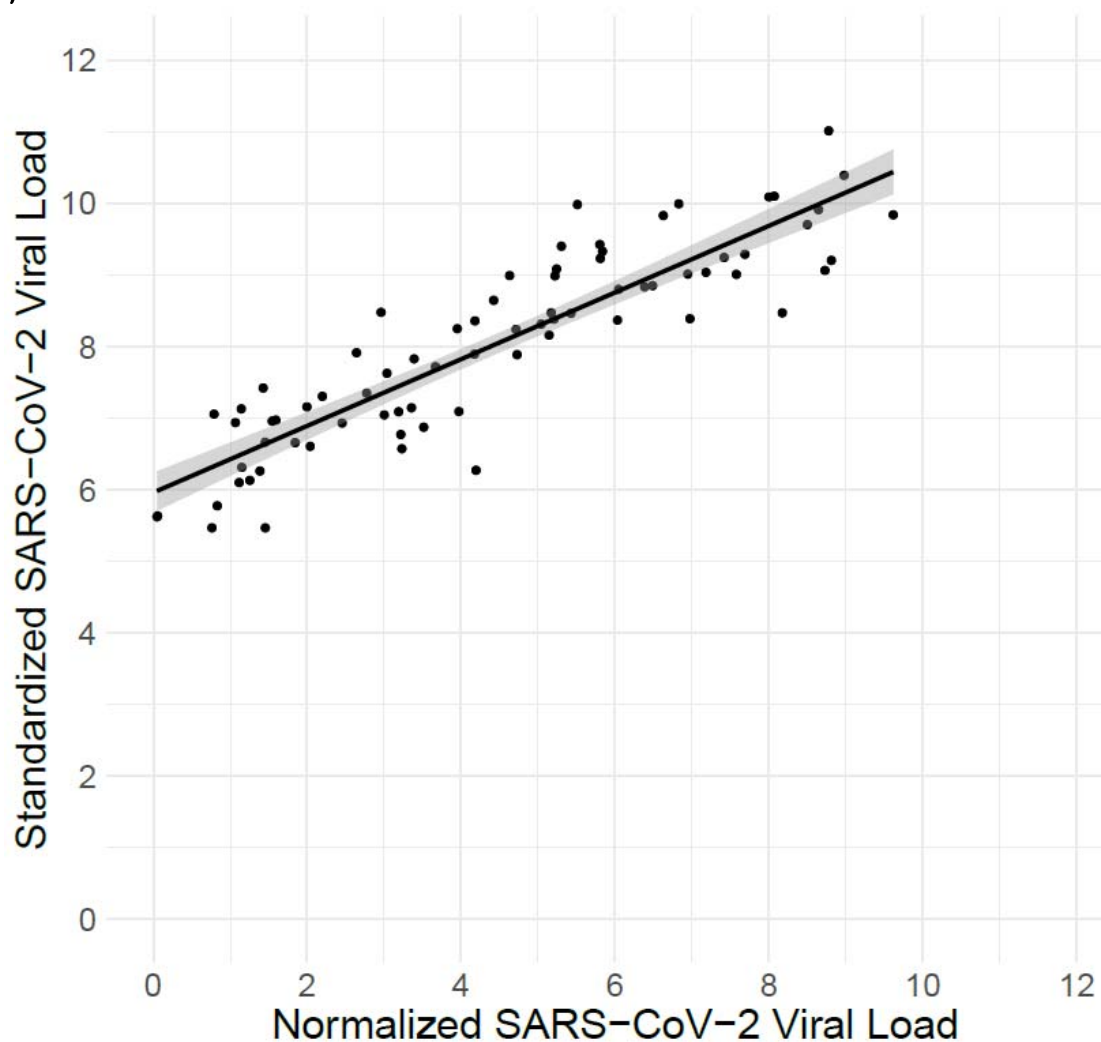
176

177

178



179 **Figure 2: Linear Relationship between normalized SARS-CoV-2 viral load and SARS-CoV-2 E**  
180 **gene Cycle-Threshold Value, for n = 212 Nasopharyngeal Specimens (SLR,  $R^2 = 0.942$ ,  $P <$**   
181 **0.001).**



182  
183 Scatterplot of normalized and standardized SARS-CoV-2 viral load assayed by RT-qPCR in n = 212 nasopharyngeal  
184 specimens collected from people tested for COVID-19 in British Columbia. The two variables show a strong positive  
185 relationship and shared variation, quantified by simple linear regression ( $R^2 = 0.942$ ,  $P < 0.001$ ). Black line and  
186 shading show the model fit and 95% confidence interval.

187