Supplementary Materials

The Limit of Detection Matters; The Case for Benchmarking SARS-CoV-2 Testing

Ramy Arnaout,[a,c,f #] Rose A. Lee,[a,b,c] Ghee Rye Lee,[e] Cody Callahan,[d] Annie Cheng,[a] Christina F. Yen,[b,c] Kenneth P. Smith,[a,c] Rohit Arora,[a,c] James E. Kirby,[a,c,#]

[a]Department of Pathology, Beth Israel Deaconess Medical Center, Boston, MA, USA

[b]Division of Infectious Diseases, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA

[c]Harvard Medical School, Boston, MA, USA

[d]Department of Radiology, Beth Israel Deaconess Medical Center, Boston, MA, USA 02215

[e]Department of Surgery, Beth Israel Deaconess Medical Center, Boston, MA, USA 02215

[f]Division of Clinical Informatics, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA USA 02215

#Co-Corresponding Authors (order determined alphabetically)
Ramy Arnaout and James E. Kirby
Beth Israel Deaconess Medical Center
330 Brookline Avenue - YA309
Boston, MA 02215
jekirby@bidmc.harvard.edu
Phone: 617-667-3648
Fax: 617-667-4533

Supplementary Methods: Conversion from Ct to Viral Load

Derivation. RT-PCR depends on the exponential amplification of template to form product, which is measured using a fluorescence signal that is directly proportional to the concentration of product (e.g. an intercalating fluorophore). Fluorescence signal is monitored over time as repeat cycles of PCR are performed. In the early cycles, the signal is typically below the detection threshold and so appears as a flat baseline; this is the lag phase. As exponential amplification continues, the signal crosses the detection threshold and exponential growth becomes apparent; this is the log phase. As amplification continues further, formation of double-stranded product and primer consumption inhibit the amplification reaction [1, 2], causing the growth of the signal to stall; this is the plateau phase. Typically a signal threshold (τ) is chosen within the region where the log phase is apparent; the cycle at which the signal crosses τ is called the Ct value. Note that because monitoring is continuous, the Ct value can be fractional (e.g. 6.54 cycles, as opposed to cycle 6 or cycle 7).

To convert from Ct value to viral load requires an expression that relates the two. To derive this expression, we first consider the relationship between signal intensity (y) and cycle number (x), which as described above is exponential:

$$y(x) = y_0 e^{kx}$$

Here y_0 is the intensity of the input material (generally far too low to measure directly). Recall that intensity is directly proportional to viral load in the sample (v); we can therefore substitute $y = Cv$, where C is a proportionality constant in units of fluorescence intensity units per copies/mL:

$$y(x) = Cv_0 e^{kx}$$

We use e^k, where k denotes the rate of growth, for ease of illustrating the exponential form of this expression: in every cycle, the amount of product at the end of the cycle is e^k times the amount at the start of the cycle; thus e^k is a ratio. Because PCR results in a doubling of product in each cycle (at maximum efficiency and ignoring measurement error, inhibitors, and other extenuating factors), we know that e^k will have a value in the vicinity of 2. However we also know that inhibition of the polymerase by PCR product will make this ratio fall with each cycle, and we will need to measure this fall or else risk underestimating the amount of starting material. Therefore it is useful to make a substitution for e^k, to the ratio ρ, yielding

$$y(x) = Cv_0\rho^x \qquad \text{Eq. 1}$$

and because ρ is expected to vary (fall) with x, to write ρ as a function of x:

$$y(x) = Cv_0\rho(x)^x \qquad \text{Eq. 2}$$

We can measure ρ(x) for a given PCR reaction from the signal-vs-cycle plot. In the interval x_1 to x_2, the amount of signal will increase from y_1 to y_2 by some multiple that depends on ρ:

$$y_2 = y_1\rho^{x2-x1}$$

Because ρ is what we desire to measure, we solve for it, yielding

$$\rho = (y_2/y_1)\text{^}(1/(x_2-x_1)) \qquad \text{Eq. 3}$$

We acquired screenshots of the signal-vs-cycle curve for 50 randomly chosen positive samples and extracted the data using WebPlotDigitizer [3]. This yielded ~200 datapoints for each sample at a density of 7-30 datapoints per PCR cycle. We then used Eq. 3 to measure ρ at every

datapoint. To minimize sampling error, we actually measured ρ on 20 x_1-x_2 intervals centered around each datapoint (consecutive datapoints, next-nearest neighbors, and so on), binned to every 1/10th of a cycle, and took the median ρ for each bin, resulting in smooth curves of efficiency (=ρ-1) vs. cycle number (e.g. Fig S1).

As expected, following a short interval during which efficiency appeared to increase as a result of signal crossing the detection threshold and becoming quantifiable, efficiency peaked and then fell with cycle number. The peak efficiency exceeded 1 for several samples, as is possible from literature on RT-PCR, which mentions measured efficiencies up to 1.3 (ρ up to 2.3) [4]. The peak efficiency reliably occurred 1.41±0.93 (mean±stdev) cycles after the machine-reported Ct, a not surprising finding, since the Abbott Ct or FCN is based on modeling of peak efficiency, i.e., the cycle number at the so called maxRatio [5]. Also as expected, there was a negative association across samples between efficiency and Ct number, well fit (R^2=0.82) by a linear relationship with Theil-Sen slope m=-0.028/cycle and intercept b=1.34 (Fig. S2). This relationship provides a measure of ρ(x) from Eq. 2, or more precisely, it provides ρ(Ct), which is needed to solve for v_0 as a function of Ct.

Using ρ(Ct) to account for the decreasing efficiency with Ct, the signal threshold $\boldsymbol{\tau}$ that corresponds to Ct can be approximated as

$$\boldsymbol{\tau} = Cv_0 \times \prod(mx+b)$$

because every cycle results in a fold increase over the previous cycle, only with falling efficiency (m is negative). Note the product is taken from x=1 to x=Ct-1. (To see how this equation relates to Eq. 1 above, note too that if efficiency did not fall, m would be 0, and the mx term would

disappear; note further that if efficiency remained at 100% in this scenario, then ρ=b=2.) Taking the log of both sides:

$$\log\boldsymbol{\tau} = \log C + \log v_0 + \sum(\log(mx+b))$$

where now the sum is from x=1 to x=Ct-1. Because the data we have gives fluorescence at fractional cycles, it is useful to convert from discrete (here, per-cycle) growth to continuous growth:

$$\log\boldsymbol{\tau} = \log C + \log v_0 + \int \log(mx+b)dx$$

with the same limits as above. Integrating and evaluating at these limits and simplifying slightly:

$$\log\boldsymbol{\tau} = \log C + \log v_0 + (b/m+Ct-1)\times\log(m(Ct-1)+b) - (Ct-1) - ((b/m+1)\times\log(m+b) - 1)$$

$$\log\boldsymbol{\tau} = \log C + \log v_0 + (b/m+Ct-1)\times\log(m(Ct-1)+b) - (b/m+1) \times\log(m+b) - Ct + 2 \qquad \text{Eq. 4}$$

From the manufacturer and our validation, we know that at the LoD v_0=100 copies/mL (v_L) and Ct=26.06 (Ct_L). Therefore:

$$\log\boldsymbol{\tau} = \log C + \log v_L + (b/m+Ct_L-1) \times\log(m(Ct_L-1)+b) - (b/m+1) \times\log(m+b) - Ct_L + 2 \qquad \text{Eq. 5}$$

Subtracting Eq. 5 – Eq. 4 and moving $\log v_0$ to the left-hand side:

$$\log v_0 = \log v_L + (Ct_L-1+b/m)\times\log(m(Ct_L-1)+b) - (Ct-1+b/m)\times\log(m(Ct-1)+b) + Ct-Ct_L \qquad \text{Eq. 6}$$

For clarity, we can make the following substitutions:

$A = m(Ct_L-1)+b$: a constant reflecting the efficiency at Ct_L

$B = m(Ct-1)+b$: the efficiency at the observed Ct

$C = \log v_L + A/m \times \log A - Ct_L$: a constant encompassing the measured LoD and efficiency falloff

This yields the somewhat more readable

$$\log v_0 = -B/m \times \log B + Ct + C \qquad \text{Eq. 6}$$

Exponentiating both sides yields the desired expression for v_0 as a function of Ct, with v_L, Ct_L, m, and b as above.

Validation. The model was validated using calibration control material obtained from the United States Food and Drug Administration. The inactivated, tissue culture cell-grown SARS-CoV-2 reference material was provided with virus quantified in arbitrary units/mL rather than in viral genome copies/mL. Therefore, to aid in comparison of model predictions and the calibration standard, serial 10-fold dilutions of the standard were prepared in quadruplicate and finer dilutions prepared surrounding the assay limit of detection (LoD) in replicates of ten. The dilutions were then tested using the Abbott M2000 RealTime PCR assay to correlate LoD with Ct. The LoD was determined using simple logistic regression (Prism 8.0 for MacOS, Graphpad Software, LLC, San Diego, CA) based on the detection percentage at each dilution. The LoD of the Abbott assay was also determined by testing serial dilutions of the SeraCare standard (see Materials and Methods, main text), a standard quantified in viral genome copies/mL by droplet digital PCR, and was determined to be 50 genome copies/mL by simple logistic regression. Therefore, we assigned a value of 50 genome copy/mL to the serial dilution and cycle threshold of the FDA reference standard at the assay LoD. The Ct value for the next 10-fold more concentrated reference standard dilution was assigned a value of 500 copies/mL, etc.

The Ct values predicted by the model and obtained from the reference material at each serial log10 titer are plotted in Fig. S3. The concentration of the available reference material only allowed replicate tested at the five log10 dilutions spanning 50 to 5 x 10^5 genome copies/mL. However, within this range, the model and calibrator show almost an identical slope and nearly complete overlap. Therefore, we conclude that Eq. 6 is highly predictive of the relationship between Ct and viral load. We expect that significant deviations from model predictions are less likely to occur at very early cycle thresholds, i.e., greater than 10^6 genome copies/mL, as PCR reaction inhibition observed at high cycles numbers is limited and therefore amplification parameters are more predictable. Therefore, we predict that the extension of model predictions to 10^9 genome copies/mL is likely to be reasonably accurate.

References

1. Jansson L, Hedman J. Challenging the proposed causes of the PCR plateau phase. Biomolecular detection and quantification **2019**; 17: 100082.
2. Kainz P. The PCR plateau phase - towards an understanding of its limitations. Biochimica et biophysica acta **2000**; 1494(1-2): 23-7.
3. Rohatgi A. WebPlotDigitizer. 4.2 ed. San Francisco, CA, **2019**.
4. Svec D, Tichopad A, Novosadova V, Pfaffl MW, Kubista M. How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. Biomolecular detection and quantification **2015**; 3: 9-16.
5. Shain EB, Clemens JM. A new method for robust quantitative and qualitative analysis of real-time PCR. Nucleic acids research **2008**; 36(14): e91.

Figure S1

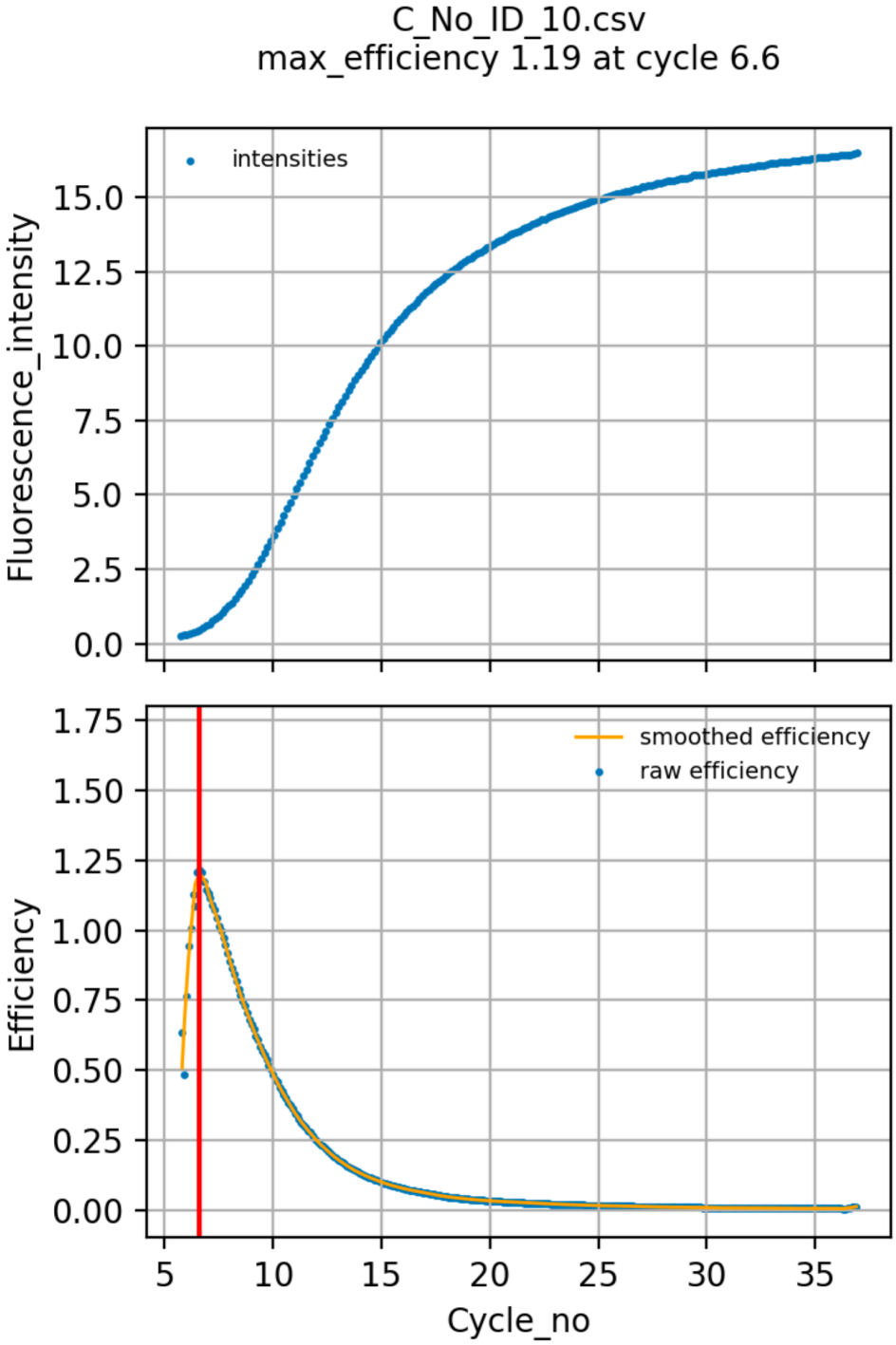

C_No_ID_10.csv
max_efficiency 1.19 at cycle 6.6
intensities
Fluorescence_intensity
15.0
12.5
10.0
7.5
5.0
2.5
0.0
smoothed efficiency
raw efficiency
Efficiency
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
5
10
15
20
25
30
35
Cycle_no

Figure S2

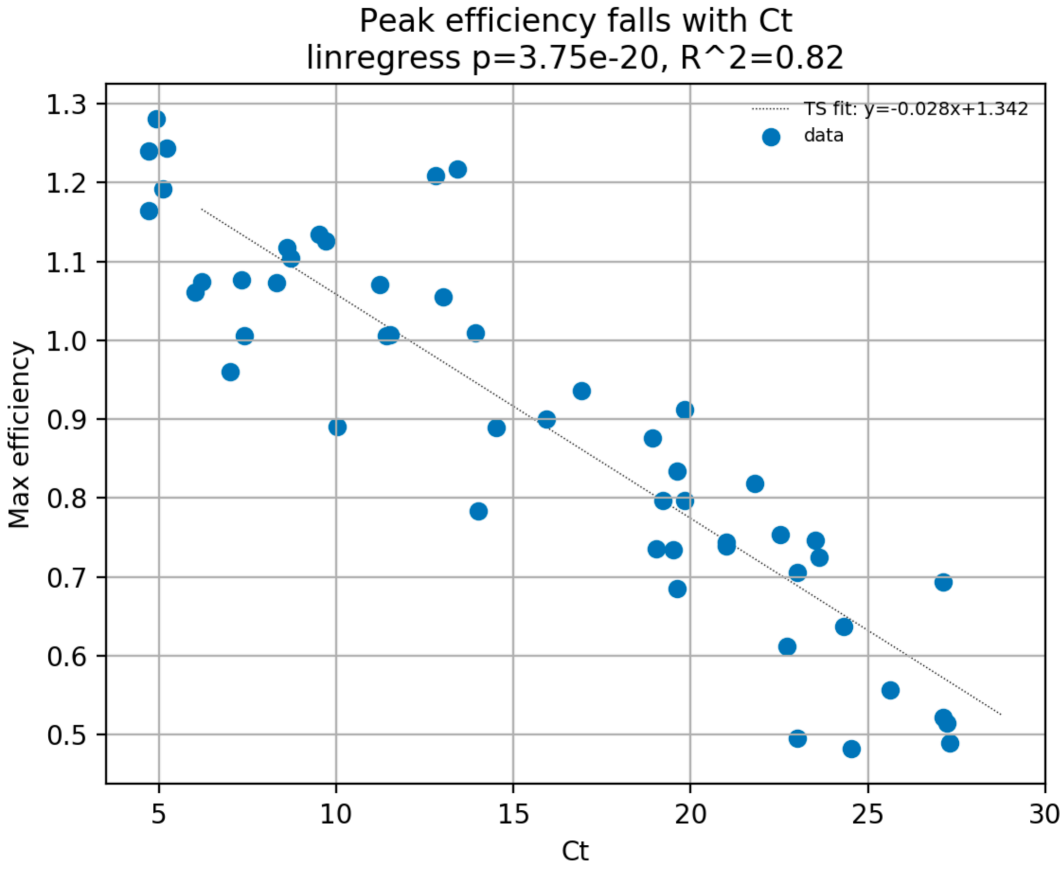
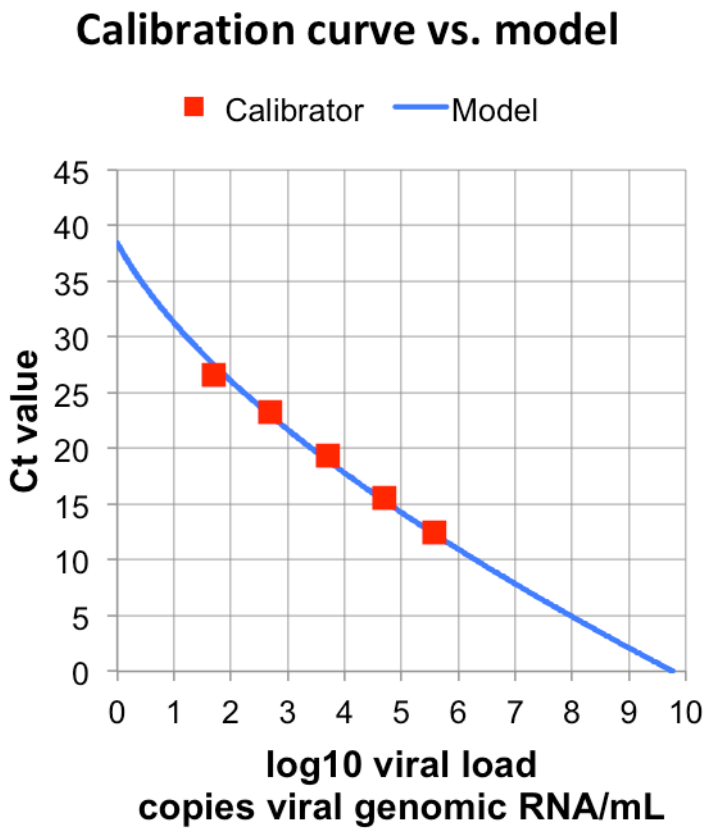

Peak efficiency falls with Ct
linregress p=3.75e-20, R^2=0.82
TS fit: y=-0.028x+1.342
data
Max efficiency
Ct
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
5
10
15
20
25
30

Figure S3


Calibration curve vs. model
Calibrator
Model
45
40
35
30
25
20
15
10
5
0
Ct value
0 1 2 3 4 5 6 7 8 9 10
log10 viral load
copies viral genomic RNA/mL