RESEARCH ARTICLE

ct2vl: Converting Ct Values to Viral Loads for SARS-CoV-2 RT-qPCR Test Results

Elliot D. Hill, MSc,[a†] Fazilet Yilmaz, MD,[b†] Cody Callahan,[a†] Annie Cheng,[a] Jasper Braun, PhD,[a] Ramy Arnaout, MD, DPhil[a, c*]

Beth Israel Deaconess Medical Center, Division of Clinical Pathology, Department of Pathology, Boston, MA USA 02215[a]; Department of Pathology, Rhode Island Hospital, Warren Alpert Medical School of Brown University, Providence, MA USA 02903[b]; Beth Israel Deaconess Medical Center, Division of Clinical Informatics, Department of Medicine, Boston, MA USA 02215[c]

Compiled June 16, 2022
This is a draft manuscript, pre-submission

Address correspondence to rarnaout@bidmc.harvard.edu.

ABSTRACT RT-qPCR is the de facto reference method for detecting the presence of SARS-CoV-2 genomic material in infected individuals (1). Although RT-qPCR is inherently quantitative and despite SARS-CoV-2 viral loads varying by 10 orders of magnitude and therefore being potentially highly clinically informative, in practice SARS-CoV-2 RT-qPCR results are usually reported qualitatively as simply positive or negative. This is both because of the mathematical complexity of converting from C_t values to viral loads and because the same C_t value can correspond to orders-of-magnitude differences in viral load depending on the testing platform (2, 3, 4). To address this problem, here we present *ct2vl*, a Python package designed to help individual clinical laboratories, investigators, and test developers convert from C_t values to viral loads on their own platforms, using only the data generated during validation of those platforms. It allows any user to convert C_t values to viral loads and is readily applicable to other RT-qPCR tests. *ct2vl* is open source, has 100% code coverage, and is freely available via the Python Package Index (PyPI).

IMPORTANCE Up to now, COVID-19 test results have been reported as positive vs. negative, even though "positive" can mean anywhere from 1 copy of SARS-CoV-2 virus per milliliter of transport media to over 1 *billion* copies/mL, with attendant clinical consequences. Democratizing access to this quantitative data is the first step toward its eventual incorporation into test development, the research literature, and clinical care.

KEYWORDS: SARS-CoV-2, Python, RT-qPCR, C_t value, viral load, limit of detection

INTRODUCTION

The real-time reverse-transcription polymerase chain reaction, commonly known as quantitative RT-PCR or RT-qPCR, is a standard method for testing human samples for the presence of viruses such as HIV-1 (human immunodeficiency virus type 1), HCV (hepatitis C virus), and, since 2020, SARS-CoV-2 (5, 6). In RT-qPCR, the tiny amount of genetic material originally present in a positive patient sample is copied by a polymerase enzyme over repeat cycles, resulting in exponential amplification that eventually leads to detectable amounts of genetic material (7). The cycle number at which the detection threshold is reached is called the C_t value. Because the reaction is monitored continuously, the threshold may be crossed between cycles, leading to the alternative term fractional cycle number [FCN] (8). The more starting material, the fewer cycles are needed for signal to cross the threshold. Thus, the smaller the C_t value, the greater the amount of starting material.

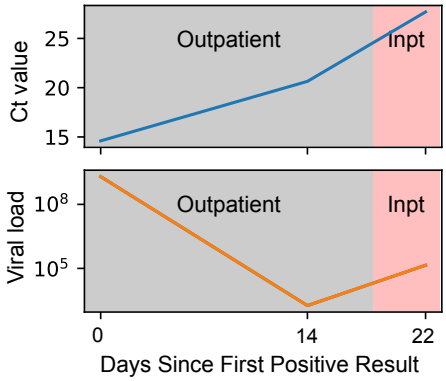

FIG 1 C_t **value vs. viral load (in copies/mL). (a) In this patient,** C_t **values trended consistently upward. Clinically, it would be reasonable to interpret this trend as reflecting a continuous fall in viral load and thereby expect clinical improvement, all other things being equal. Yet the patient worsened between Day 14 and 22, necessitating hospitalization. It happened that the Day 14 result was from a different platform from the Day 0 and Day 22 results; the** C_t **value for Day 14 cannot be interpreted on the same scale as the other two datapoints; the plot is misleading. (b) Conversion to viral load shows the true picture: a rebound in viral load, consistent with the worsening clinical picture.**

RT-qPCR results can be reported either qualitatively (positive/negative) or quantitatively. For most infections, quantitative results are usually reported not as a C_t value but as a viral load: the number of copies of viral genomic material present per unit volume of sample (i.e., a concentration). The most common unit is copies/mL. An important advantage of viral loads over C_t values is that viral loads are consistent across platforms. C_t values are not, due to platform-specific differences in polymerase, amplification conditions, the signal-detection method, whether "dark cycles" are run, and other factors. For example, for SARS-CoV-2, a C_t value of 26 corresponds to a viral load of 100 copies/mL of viral transport media on one FDA-approved platform and nearly 500,000 copies/mL on another (9). This platform-to-platform variability can make C_t values difficult to interpret and can lead to mistaken conclusions about a patient's clinical course (Fig. 1). An additional advantage is the direct correlation between disease burden and viral load, as opposed to the "golf score" inverse correlation with C_t.

In the scramble to create, approve, and validate tests for COVID-19, most SARS-CoV-2 RT-qPCR tests were not validated to output viral load (10). Yet viral loads can be valuable indicators of where a patient is in the course of infection, as well as the likelihood of being infectious. For these reasons, it has been our experience that clinical staff often ask laboratorians informally to report C_t values for their patients, outside of the health record. The difficulty in interpreting C_t values has persisted into the third year of the pandemic. As time goes on, the likelihood that a patient or caregiver will encounter RT-qPCR results from platforms with disparate C_t-value scales cannot but increase, increasing the chance of diagnostic error.

Fortunately, the correspondence between C_t value and viral load in RT-qPCR is well understood mathematically, and the validation studies that laboratories must perform in order to bring a test online can provide the data necessary to convert from C_t to viral load on clinical samples (11). In past work, we wrote computer code to convert from C_t values to viral loads to help reveal and quantify the range of viral burden in the patient population (12) and to compare the sensitivity and utility of testing from different anatomical sources quantitatively and generalizably (e.g. saliva vs. nasal secretions vs. nasopharyngeal secretions) (13, 9). We later expanded on this code to

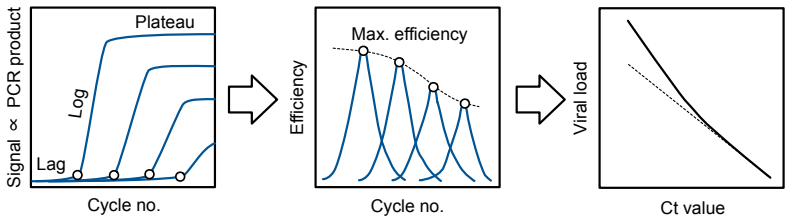

FIG 2 Overview of *ct2vl*'s approach. Left: Signal vs. cycle number traces are analyzed to find maximum replication efficiency (open circles). Middle: The fall in maximum efficiency vs. cycle number is fit by a curve (dotted line). Right: Together with an anchor point (Ct at the LoD), this curve is used to convert Ct values to viral loads (solid line). Without this curve, viral load would be underestimated (dotted line).

provide viral loads for new platforms brought online at our institution. However, this code applied only to our own platforms, despite SARS-CoV-2 viral loads being a global need. To address this need, here we present a much-expanded new Python package called *ct2vl* intended to make it straightforward to convert from C_t values to viral loads on any platform.

METHODS

Mathematical derivation. Traditionally, conversion from C_t values to viral loads has required first creating a standard curve spanning a range of viral loads at least as large as what is observed in clinical practice. However, standard curves can be time consuming and expensive, especially when viral loads range over as many orders of magnitude as they do with SARS-CoV-2 ($\geq$ 1 billion-fold between the lowest and highest viral loads encountered in clinical practice) (12). Fortunately, reliable C_t-to-viral load conversion can also be performed mathematically based on the well understood biochemical principles of PCR (14, 11, 8). This mathematical approach requires only (1) time series of signal vs. cycle number for positive samples and (2) an anchor point—the C_t value for a given viral load—such as labs routinely measure, in replicate, when validating the limit of detection (LoD) before bringing a platform online.

Both the equation we used and its experimental validation have been described in detail in previous work (12), so we review the approach only briefly here. PCR generally exhibits three phases: a lag phase set by the stochasticity of polymerase first encountering template molecules (and in practice the platform's detection threshold), an exponential (or "log") phase during which the amount of product roughly doubles each cycle, and finally a plateau phase due to inhibition of the enzyme by the (now-copious) product it has produced. A detection threshold is crossed during the exponential phase; in fact at least one large diagnostics company determines the threshold, and thereby the C_t or FCN, from the cycle at which maximum exponential growth is observed (8). Absent additional considerations, converting from C_t to viral load would involve simply fitting this relationship using an S-shaped function (e.g. Gompertz' growth curve).

In practice, fitting S-shaped curves is not straightforward. A good fit requires careful weighting of datapoints in different parts of the curve, and in practice no S-shaped function—Gompertz, sigmoid, logistic, or Chapman—precisely captures the part of the curve that is the most important for viral load determination, the exponential phase (15). Moreover, there is no guarantee that the details of the PCR formulation (e.g., multiple targets; internal controls) will not affect the details of the fit. Therefore, an alternative approach is to fit only the key determinant of the region of interest: the

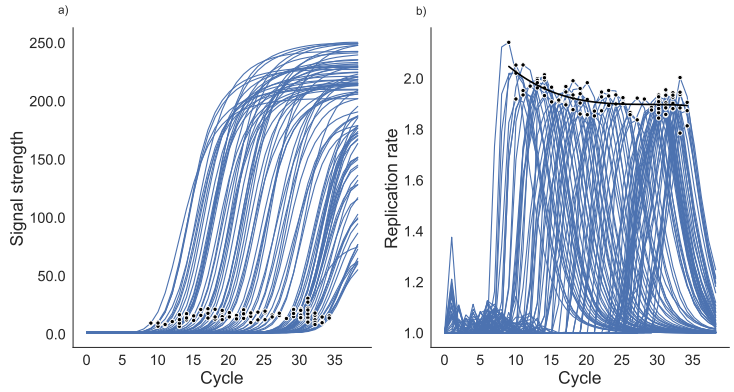

FIG 3 a) PCR traces and the points at the maximum replication rate (black dots). b) the replication rate for each trace with the points where maximum replication rate is reached. Notice that the max replication rate falls as cycle increases.

decrease in maximum polymerase efficiency that is frequently observed with increasing cycle number, especially in situations where template is admixed with an internal-control template whose amplification can compete with the template for polymerase, and inhibit polymerase, in more complex ways than a model of a single template might capture.

Exponential growth with decreasing replication rate yields the following equation for viral load, v_0, as a function of C_t value (see the supplementary information of Arnaout et al., 2021 for the derivation):

$$\rho(\beta) = X\beta \tag{1}$$

$$\log v_0 = \log v_L + \int_0^{C_{t_L}} \log \rho(\beta) dx - \int_0^{C_t} \log \rho(\beta) dx \tag{2}$$

Here, v_L and C_{t_L} give the anchor point: the simplest anchor point is to let v_L be the limit of detection (LoD) (12) and C_{t_L} be the C_t value at the LoD. ρ is a polynomial fit of maximum replication rate vs. cycle at maximum replication rate (a slight change from (12)). These constitute the parameters of this model (see Implementation, below). Maximum replication rate and the cycle at maximum replication rate are derived from time series data of the form amount of material vs. cycle number (Fig. 2a). The amount of material is most often measured as a fluorescence intensity (e.g. of an intercalating fluorophore present in the RT-qPCR reaction mix).

Implementation. To parameterize Equation (1), we must find the coefficients, β, of the polynomial regression fit between the maximum replication rate and the cycle at maximum replication rate. To calculate max replication rate and cycle at max replication rate, a set of PCR traces for positive samples are obtained from the platform and processed as follows. The initial 3 cycles are removed because these initial values of PCR traces are often noisy and may interfere with the estimation of maximum replication rate. Negative signal-intensity values are considered noise and therefore set to 0. The data is smoothed, ensuring monotonic increase (PCR product cannot decrease; slight/transient decreases sometimes observed during the lag phase are attributed to signal-detection noise). These steps result in denoised traces. Examples of processed traces are plotted in Fig. 3a. Even after denoising, some noisy measurements of replication rate can be observed in the early cycles (Fig. 3b).

Maximum replication rate is then calculated as the largest ratio of the signal at a

given cycle to the signal at the previous cycle. A polynomial regression is fit to the relationship between maximum replication rate and cycle at maximum replication rate, yielding β. The degree of the polynomial is chosen via a cross-validation grid search over degrees $1, 2, 3$ (linear, quadratic, and cubic), providing a more robust update to the description Arnaout 2021. With β estimated and v_L and C_{t_L} provided by the user, the integrals in equation (1) are calculated numerically by *ct2vl*. Equation (1) is now calibrated and is ready to convert C_t values to viral loads. The user is spared interaction with the mathematics (see Usage, below).

Calibration and validation datasets. The FDA-approved Abbott Alinity m Real-Time PCR assay SARS-CoV-2 RT-qPCR testing platform was used for this analysis. Results for the Abbott m2000 have been previously described (12). To validate *ct2vl*'s accuracy, we compared the viral load predictions from *ct2vl* to a validation dataset composed of 40 C_t values and their corresponding viral loads from two independent calibration series of viral loads on the same Abbott Alinity m SARS-CoV-2 RT-qPCR machines. First, we calibrated *ct2vl* on 96 positive PCR traces from one of the Alinity machines, using equation (1) and the known (experimentally confirmed) LoD = 100 copies/mL and mean C_{t_L} = 37.83 for this machine (see Usage, below). We then used *ct2vl* to convert the 48 C_t values from the validation dataset to viral loads and compared these predicted viral loads to the ground-truth viral loads in the validation dataset to estimate the prediction error. The calibration and validation datasets are provided as Supplementary Information.

For the validation dataset, the genome copy number was based on the reference standard produced by SeraCare (AccuPlex SARS-CoV-2 Reference Material Kit, catalog number 0505-0126). This control material consists of replication-incompetent, enveloped, positive-sense, single- stranded RNA Sindbid virus into which SARS-CoV-2 PCR targets detected by Abbott SARS-CoV-2 RT-qPCR assays have been cloned. This control material was quantified by the manufacturer using digital droplet PCR, and diluted into viral transport medium for analysis.

Sensitivity analysis. To estimate *ct2vl*'s sensitivity to the C_{t_L} parameter, we replaced the mean value of 37.83 with each of 23 different C_{t_L} measurements in separate calibration tests (these values were averaged together to get the mean), while holding all other parameters fixed, and measured the prediction error (| predicted viral load - known viral load |) on the validation dataset (for which viral load was known). To estimate *ct2vl*'s sensitivity to β, we bootstrap-resampled our calibration data 1000 times (randomly sampling the same number of traces, with replacement), refit the polynomial regression (in Equation (1)) on each bootstrapped sample, then calculated the *ct2vl* prediction error on the validation dataset for each bootstrap sample. To estimate total confidence intervals, i.e., the cumulative effect of variation in CtL and β, we bootstrap-refit β for each of the 23 C_{t_L} values.

Code coverage. In computer science, code coverage (or test coverage) is a measure of how much of the code is covered by test suites. Complete coverage means every line of code has been tested for proper function. For *ct2vl*, code test coverage was determined using the *pytest-cov* package.

RESULTS

Python package overview. The Python package *ct2vl* takes RT-qPCR-reaction time series as input to parameterize an equation describing the relationship between C_t value and viral load (Equation (1)); after this calibration step, it converts new C_t values to viral load for a given C_{t_L} and v_L. The package can be used as a command line tool or imported into Python programs. The package has 100% test code coverage and has

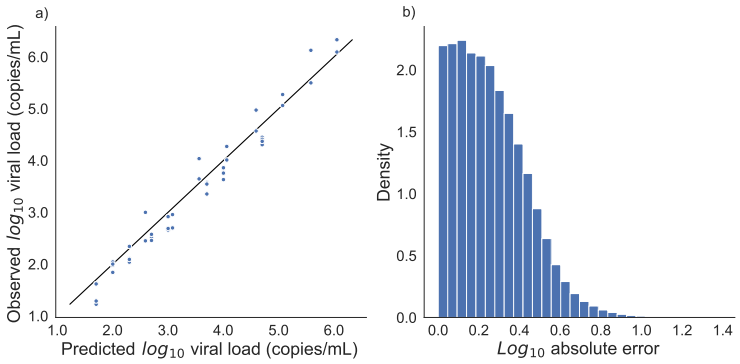

FIG 4 Predicted vs. observed viral load. (a) Viral loads predicted by *ct2vl* compared observed viral loads in a validation dataset of known viral loads (Seracare). (b) Histogram of absolute prediction error. Notice that the majority of the error is below half a log10 unit. Since viral loads can range over 10 orders of magnitudes, this error is within a clinically useful range.

been tested on macOS, Ubuntu, and Windows.

Calibration and validation results. For our calibration dataset, the coefficients of the polynomial fit between max replication rate and cycle at max replication rate were 2.27, -2.48e-02, 4.06e-04 (Fig. 3). For the validation dataset, predicted values demonstrated excellent agreement with observed values (Pearson's $r = 0.99$, $p < 0.001$). The mean absolute error between predicted and observed viral load was 0.21 ± 0.28 log10 units ($mean \pm 2std$), meaning that predicted viral loads were accurate to within 2.5 ± 1.5 fold, highly accurate considering that viral loads range over 10 orders of magnitude in SARS-CoV-2 infection (12). Consistent with this finding, R^2 was 0.97 for a linear fit between predicted and observed viral loads, with slope 1:1, demonstrating the accuracy as well as precision of *ct2vl* over the full range of the six orders of magnitude of available validation data (Fig. 4).

Sensitivity analysis results. Regarding sensitivity to the C_{t_L} parameter, we found an absolute prediction error of 0.25 ± 0.33 log10 units ($mean \pm 2std$). Bootstrapping β parameters gave a mean absolute error of 0.24 ± 0.30 log10 units. Lastly, to measure overall sensitivity, varying C_{t_L} and bootstrapping β parameters simultaneously gave a mean absolute error of $.25 \pm 0.35$ log10 units (Fig. 5).

Installation. *ct2vl* requires Python 3.7 or higher to be installed. Assuming *pip* is installed, to install *ct2vl*, at the command line, run

```
$ pip install ct2vl
```

Command-line usage. To calibrate *ct2vl* run

```
$ python3 -m ct2vl calibrate <traces> <LoD> <Ct_at_LoD>
```

Here <infile> is a csv file containing the positive traces where each row is a PCR reaction trace and each column is a time step in that trace. See the example file positive_traces.csv in the Supplementary Information, for which the command would be:

```
$ python3 -m ct2vl calibrate traces.csv 100.0 37.83
```

Once *ct2vl* has been calibrated using the above command, C_t values can be converted to viral loads by typing

```
$ python3 -m ct2vl convert <Ct>
```

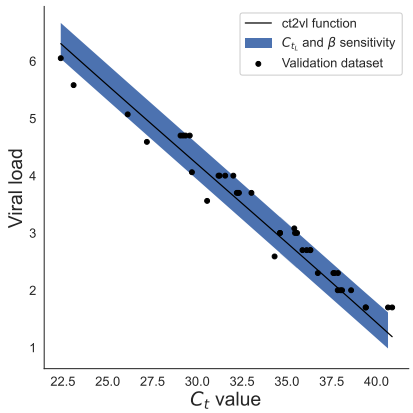

FIG 5 C_t **value vs. viral load. The validation dataset (black dots) with the *ct2vl* prediction function (mean** $\pm$ **2std) when** C_{t_L} **was varied and the calibration parameters (**β**) were bootstrapped.**

One or multiple C_t values can be passed. For example, to convert a C_t value of 23.1 to a viral load, with a LoD of 100 copies/mL and a corresponding C_t value of 38.73:

```
$ python3 -m ct2vl convert 23.1
```

The output will be printed to the screen in a text table with integer row numbers, the LoD and Ct-at-LoD (C_{t_L}) used, the C_t value that was input, the viral load in units of copies/mL, and the viral load in log10 units:

	LoD	Ct_at_LoD	Ct	viral_load	log10_viral_load
1	100.00	37.83	25.32	299427.73	5.48

To convert several C_t values to viral loads:

```
$ python3 -m ct2vl convert 25.32 30.11 35.95
```

Output:

	LoD	Ct_at_LoD	Ct	viral_load	log10_viral_load
1	100.00	37.83	25.32	299427.73	5.48
2	100.00	37.83	30.11	14030.16	4.15
3	100.00	37.83	35.95	336.52	2.53

Output can be saved to a file by providing a file path to the optional flag '–output', like so:

```
$ python3 -m ct2vl convert 100.0 37.83 23.1 --output viral_loads.tsv
```

Here, the tabular output will have been saved to a tab-delimited text file called viral_loads.tsv.

Python-package usage. For users who are familiar with the Python programming language and environments, *ct2vl* can also be used programmatically as follows:

```
>>> from ct2vl.ct2vl import CT2VL
>>> converter = CT2VL('traces.csv', LoD=100.0, Ct_at_LoD=37.83)
>>> viral_loads = converter.ct_to_viral_load(Ct=[25.3, 30.1, 35.9])
>>> viral_loads
[299427.732571, 14030.156190, 336.522756]
```

Availability. *ct2vl* is freely available on the Python Package Index (PyPI) or via GitHub.

DISCUSSION

The COVID-19 pandemic necessitated development and approval of RT-qPCR tests at a rate that outstripped the ability to convert results reliably from C_t values to viral loads. As COVID-19 becomes endemic, the probability that patients will receive results on multiple platforms can only rise, as will the need to manage cases where viral load will need to be monitored over time. *ct2vl* facilitates calculation of viral loads for any platform, based on a laboratory's own validation data. Because the mathematics of RT-qPCR are more complicated for real-world clinical tests, which often contain internal controls and multiple targets, than in stripped-down experimental systems, and because accurate assessment of maximum efficiency is more important for viral load estimation than fitting the entirely of the curve (lag, log, and stationary phase), *ct2vl* concentrates on the most important cause of deviation from pure exponential growth—the fall of replication efficiency with cycle—and fits its empirically/phenomenologically, as opposed to shoe-horning a particular S-shaped curve from the many such curves that exist.

Comparison against a calibration curve of well described SARS-CoV-2 standard (Seracare) demonstrated excellent performance of *ct2vl*, including robustness to sensitivity analysis. Total error was less than half a log10 unit, acceptable performance relative to the 10 log10 units over which viral loads vary in SARS-CoV-2 infection and comparable to the error in HIV viral load testing. *ct2vl* is made available free of charge with a completely open-source codebase and 100% code test coverage, to facilitate customization and incorporation into laboratory workflows. We note that *ct2vl* is applicable to any situation in which calibration is available and polymerase replication rate does not rise with cycle number (i.e., all conventional PCR). It is hoped that it will prove useful beyond SARS-CoV-2, even as the world continues to have to manage successive waves of COVID-19 (16).

SUPPLEMENTARY INFORMATION

The traces used to calibrate ct2vl for this study and validation data can be found here.

REFERENCES

1. **Böger B, Fachi MM, Vilhena RO, Cobre AF, Tonin FS, Pontarolo R**. Jan 2021. Systematic Review with Meta-Analysis of the Accuracy of Diagnostic Tests for COVID-19. Am J Infect Control 49 (1):21–29. doi: 10.1016/j.ajic.2020.07.011.
2. **Raju S, Anderson NW, Robinson E, Squires C, Wallace MA, Zhang R, Burnham CAD, Parikh BA**. Nov 2021. Comparison of 6 SARS-CoV-2 Molecular Methods and Correlation With the Cycle Threshold Distribution in Clinical Specimens. J Appl Lab Med 6 (6):1452–1462. doi: 10.1093/jalm/jfab086.
3. **Rhoads D, Peaper DR, She RC, Nolte FS, Wojewoda CM, Anderson NW, Pritt BS**. May 2021. College of American Pathologists (CAP) Microbiology Committee Perspective: Caution Must Be Used in Interpreting the Cycle Threshold (Ct) Value. Clin Infect Dis 72 (10):e685–e686. doi:10.1093/cid/ciaa1199.
4. **Wirden M, Feghoul L, Bertine M, Nere ML, Le Hingrat Q, Abdi B, Boutolleau D, Ferre VM, Jary A, Delaugerre C, Marcelin AG, Descamps D, Legoff J, Visseaux B, Chaix ML**. Sep 2020. Multicenter Comparison of the Cobas 6800 System with the RealStar RT-PCR Kit for the Detection of SARS-CoV-2. J Clin Virol 130:104573. doi: 10.1016/j.jcv.2020.104573.
5. **Gibellini D, Vitone F, Schiavone P, Ponti C, La Placa M, Re MC**. Apr 2004. Quantitative Detection of Human Immunodeficiency Virus Type 1 (HIV-1) Proviral DNA in Peripheral Blood Mononuclear Cells by SYBR Green Real-Time PCR Technique. J Clin Virol 29 (4):282–289. doi: 10.1016/S1386-6532(03)00169-0.
6. **Le Guillou-Guillemette H, Lunel-Fabiani F**. 2009. Detection and Quantification of Serum or Plasma HCV RNA: Mini Review of Commercially Available Assays. Methods Mol Biol 510:3–14. doi:10.1007/978-1-59745-394-3_1.
7. **Jung R, Soondrum K, Neumaier M**. Sep 2000. Quantitative PCR. Clin Chem Lab Med 38 (9):833–836. doi:10.1515/CCLM.2000.120.
8. **Shain EB, Clemens JM**. Aug 2008. A New Method for Robust Quantitative and Qualitative Analysis of Real-Time PCR. Nucleic Acids Res 36 (14):e91. doi:10.1093/nar/gkn408.
9. **Callahan C, Ditelberg S, Dutta S, Littlehale N, Cheng A, Kupczewski K, McVay D, Riedel S, Kirby JE, Arnaout R**. Sep 2021. Saliva Is Comparable to Nasopharyngeal Swabs for Molecular Detection of SARS-CoV-2. Microbiol Spectr 9 (1):e0016221. doi:

10.1128/Spectrum.00162-21.

10. **CDC**. Jul. 2021. CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel Instructions for Use. CDC.
11. **Platts AE, Johnson GD, Linnemann AK, Krawetz SA**. Sep 2008. Real-Time PCR Quantification Using a Variable Reaction Efficiency Model. Anal Biochem 380 (2):315–322. doi:10.1016/j.ab.2008.05.048.
12. **Arnaout R, Lee RA, Lee GR, Callahan C, Cheng A, Yen CF, Smith KP, Arora R, Kirby JE**. Feb 2021. The Limit of Detection Matters: The Case for Benchmarking Severe Acute Respiratory Syndrome Coronavirus 2 Testing. Clin Infect Dis doi:10.1093/cid/ciaa1382.
13. **Callahan C, Lee RA, Lee GR, Zulauf K, Kirby JE, Arnaout R**. Jun 2020. Nasal-Swab Testing Misses Patients with Low SARS-CoV-2 Viral Loads. medRxiv p 2020.06.12.20128736. doi:10.1101/2020.06.12.20128736. Publisher: Cold Spring Harbor Laboratory Press.
14. **Lievens A, Van Aelst S, Van den Bulcke M, Goetghebeur E**. Jan 2012. Enhanced Analysis of Real-Time PCR Data by Using a Variable Efficiency Model: FPK-PCR. Nucleic Acids Res 40 (2):e10. doi: 10.1093/nar/gkr775.
15. **Zhao S, Fernald RD**. Oct 2005. Comprehensive Algorithm for Quantitative Real-Time Polymerase Chain Reaction. J Comput Biol 12 (8):1047–1064. doi:10.1089/cmb.2005.12.1047.
16. **Arnaout R, Arnaout R**. 2022. Visualizing omicron: COVID-19 deaths vs. cases over time. Plos one 17 (4):e0265233.