Original Paper

The Relationship Between Population-Level SARS-CoV-2 Cycle Threshold Values and Trend of COVID-19 Infection: Longitudinal Study

Paria Dehesh[1], DVM, PhD; Hamid Reza Baradaran[1,2*], MD, PhD; Babak Eshrati[1,3*], MD, PhD; Seyed Abbas Motevalian[1], MD, PhD; Masoud Salehi[4], PhD; Tahereh Donyavi[5], MD, PhD

[1]Department of Epidemiology, School of Public Health, Iran University of Medical Sciences, Tehran, Iran

[2]Ageing Clinical and Experimental Research Team, Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, United Kingdom

[3]Preventive Medicine and Public Health Research Center, Tehran, Iran

[4]Department of Biostatistics, School of Public Health, Iran University of Medical Sciences, Tehran, Iran

[5]Department of Biotechnology, School of Allied Medical Sciences, Iran University of Medical Sciences, Tehran, Iran

[*]these authors contributed equally

Corresponding Author:
Babak Eshrati, MD, PhD
Department of Epidemiology
School of Public Health
Iran University of Medical Sciences
Hemmat Highway
Tehran, 1449614535
Iran
Phone: 98 9183616737
Email: babak.eshrati@gmail.com

Abstract

Background: The distribution of population-level real-time reverse transcription-polymerase chain reaction (RT-PCR) cycle threshold (Ct) values as a proxy of viral load may be a useful indicator for predicting COVID-19 dynamics.

Objective: The aim of this study was to determine the relationship between the daily trend of average Ct values and COVID-19 dynamics, calculated as the daily number of hospitalized patients with COVID-19, daily number of new positive tests, daily number of COVID-19 deaths, and number of hospitalized patients with COVID-19 by age. We further sought to determine the lag between these data series.

Methods: The samples included in this study were collected from March 21, 2021, to December 1, 2021. Daily Ct values of all patients who were referred to the Molecular Diagnostic Laboratory of Iran University of Medical Sciences in Tehran, Iran, for RT-PCR tests were recorded. The daily number of positive tests and the number of hospitalized patients by age group were extracted from the COVID-19 patient information registration system in Tehran province, Iran. An autoregressive integrated moving average (ARIMA) model was constructed for the time series of variables. Cross-correlation analysis was then performed to determine the best lag and correlations between the average daily Ct value and other COVID-19 dynamics–related variables. Finally, the best-selected lag of Ct identified through cross-correlation was incorporated as a covariate into the autoregressive integrated moving average with exogenous variables (ARIMAX) model to calculate the coefficients.

Results: Daily average Ct values showed a significant negative correlation (23-day time delay) with the daily number of newly hospitalized patients (P=.02), 30-day time delay with the daily number of new positive tests (P=.02), and daily number of COVID-19 deaths (P=.02). The daily average Ct value with a 30-day delay could impact the daily number of positive tests for COVID-19 (β=–16.87, P<.001) and the daily number of deaths from COVID-19 (β=–1.52, P=.03). There was a significant association between Ct lag (23 days) and the number of COVID-19 hospitalizations (β=–24.12, P=.005). Cross-correlation analysis showed significant time delays in the average Ct values and daily hospitalized patients between 18-59 years (23-day time delay, P=.02) and in patients over 60 years old (23-day time delay, P<.001). No statistically significant relation was detected in the number of daily hospitalized patients under 5 years old (9-day time delay, P=.27) and aged 5-17 years (13-day time delay, P=.39).

Conclusions: It is important for surveillance of COVID-19 to find a good indicator that can predict epidemic surges in the community. Our results suggest that the average daily Ct value with a 30-day delay can predict increases in the number of positive confirmed COVID-19 cases, which may be a useful indicator for the health system.

(JMIR Public Health Surveill 2022;8(11):e36424) doi: 10.2196/36424

KEYWORDS

cycle threshold value; COVID-19; trend; surveillance; epidemiology; disease surveillance; surveillance; digital surveillance; prediction model; epidemic modeling; health system; infectious disease

Introduction

Coronaviruses are zoonotic pathogens that can be transmitted to humans after acquiring particular mutations [1]. SARS-CoV-2, which causes COVID-19, is mainly transmitted via airborne respiratory droplets. Although ocular secretions and oral-fecal transmission have also been indicated, these transmission methods remain uncertain [2,3].

A real-time reverse transcription-polymerase chain reaction (RT-PCR) test is used for detecting SARS-CoV-2 in respiratory samples as routine surveillance worldwide. The RT-PCR test has high sensitivity and specificity for diagnosing COVID-19 and offers faster turnaround times than the viral culture method; thus, this test has become the main method for diagnosing COVID-19. RT-PCR presents both qualitative and quantitative results with respect to the viral load [4]. The RT-PCR cycle threshold (Ct) value is identified as the number of amplification cycles needed to detect the target gene in samples [5]. The Ct value is a semiquantitative result of RT-PCR that reflects the amount of viral nucleic acids in a sample, and can thus be used as a proxy for viral load and may help decision-making in epidemic control. The Ct value has a reverse relationship with viral load so that each 3.3 increase in Ct value causes a 10-fold decrease in viral load [6]; the highest viral burden is on the first day of disease symptoms onset [7]. The positive result of COVID-19 RT-PCR tests has a lower Ct value than the recommended cutoff. In the United States, the Food and Drug Administration considers a Ct value <37 as the cutoff for a positive result of COVID-19 [8]. In more than 70% of samples with a Ct value <25, SARS-CoV-2 may be cultured, whereas only 3% of samples with a Ct value >35 can be cultured [9]. Several studies have reported that the Ct value also has an association with disease severity and mortality, and that the Ct values in patients who have more severe symptoms are low [5,10-12]. In addition, hospitalized patients who died from COVID-19 had lower Ct values [13]. A systematic review showed a significant correlation between Ct value and disease severity in hospitalized patients but not in nonhospitalized COVID-19 patients [5]. There is controversy among studies on the use of Ct values at an individual level for the prognosis of the disease or treatment planning. The Ct value may vary due to the collection method among laboratories [14] or the target gene selected for RT-PCR [15]. Moreover, the RT-PCR test can detect any viral material and does not distinguish between live viruses and viral debris, which may persist for a long time beyond the point of infectiousness [12].

To the best of our knowledge, few studies have examined the use of population-level Ct values as a measure of COVID-19 dynamics in communities. As Ct values have a significant relationship with disease severity and infectivity, a higher average Ct value in daily testing samples from a population may predict epidemic growth in a community. Hay et al [16] analyzed simulation and surveillance data and found that decreases in the proportion of Ct values in a population may cause a local increase in transmission or a new number of patients [16]. In addition, the median Ct value may be an effective measure for forecasting a pandemic surge.

To resolve these issues, the aims of this study were to determine the relationships between the daily trend of average Ct value and COVID-19 dynamics, including the daily number of hospitalized patients with COVID-19, daily number of new positive tests, daily number of COVID-19 deaths, and number of hospitalized patients with COVID-19 by age. We further aimed to determine the lag between these series.

Methods

Samples and RT-PCR

The samples included in this study were collected from March 21, 2021, to December 1, 2021. Inclusion criteria were samples obtained from individuals suspected of having COVID-19 and were referred to a laboratory in Tehran, Iran, to confirm the diagnosis. Daily results of Ct values of all patients referred to the laboratory for RT-PCR tests were recorded. The daily number of positive cases and the number of hospitalized people by age group for 9 months were extracted from the COVID-19 patient information registration system in Tehran province, Iran.

This study included samples of the upper respiratory tract (both nasopharyngeal and anterior nares swab samples) taken using a sterile Dacron thin swab with a plastic or aluminum handle as the main test specimen. The samples were collected by a physician, nurse, laboratory expert, and other staff with sufficient training and experience. All biological samples were sent to the Molecular Diagnostic Laboratory of Iran University of Medical Sciences in Tehran, Iran. All samples were analyzed using the Pishtazteb One-step RT-PCR COVID-19 Kit (dual-target gene diagnosis), and RNA extraction was performed using a Zybio nucleic acid extraction kit (magnetic bead method). To confirm the diagnosis, the target genes were the SARS-CoV-2 nucleocapsid gene and RdRp gene [17]. For each sample, the Ct value was recorded. The samples that produced a positive result in the RT-PCR test and had a Ct value ≤37 were recorded to determine the daily average Ct values.

Statistical Analysis

Overview

The daily median Ct value among all patients referred to the laboratory and the daily number of hospitalized patients with COVID-19 by age group were plotted over time. The autoregressive integrated moving average (ARIMA) and autoregressive integrated moving average with exogenous variables (ARIMAX) models were used to determine significant associations between the daily average Ct value and the daily number of COVID-19 hospitalizations by age, daily number of COVID-19 deaths, and daily number of positive tests in Tehran province, Iran.

ARIMA Model

Time-series analyses are appropriate when dealing with a set of data that has a time trend [18]. The Box-Jenkins time-series approach, especially the ARIMA model, is one of the best methods in time-series analysis of autocorrelated data [19], such as the daily average Ct value. In autoregressive models, the outcome (Y_t) is a linear function of the previous values and a random component. Nonseasonal ARIMA model parameters are (*p*, *d*, *q*) overall, where *p* is the order of autoregression (AR), *d* is the degree of trend difference, and *q* is the order of moving average (MA). To perform time-series analysis, it is first necessary to check the stability of the mean and variance. For this purpose, the augmented Dickey-Fuller (ADF) test is used [20] for checking the stability of the mean and the Box-Cox test is used to check the stability of the variance. Logarithm transformation and differentiation were used to establish stability in the variance and mean, respectively. The first-time differences can be expressed as:

$$Y'_t=Y_t-Y_{t-1} \quad \textbf{(1)}$$

Where Y_t represents nonstationary time-series data and Y'_t is the time series after the first-time differences. If the time series has a seasonal trend, seasonal differences are used to stabilize the series. The AR parameter *p* represents the linear correlation of the current value of the time series Y_t with the previous values Y_{t-1}, Y_{t-2},... and current residuals ε_t [21]. The MA parameter *q* shows the linear correlation of the current value of the time series Y_t with the current and previous residuals of the time series ε_t, ε_{t-1},… [22]. The general formula of AR (*p*) and MA (*q*) models are represented in equations (2) and (3), respectively:

$$Yt=C+\beta_1 Y_{t-1}+\beta_2 Y_{t-2}+\ldots+\beta pYt-n+\varepsilon_t \quad \textbf{(2)}$$

$$Yt=C+\varepsilon_t-\varphi_1\varepsilon_{t-1}-\varphi_2\varepsilon_{t-2}\ldots-\varphi_q\varepsilon_{t-q} \quad \textbf{(3)}$$

where C is a constant; β_1, β_2,…, β_p are AR model terms; and φ_1, φ_2,…, φ_q are MA model terms. The number of AR and MA parameters was determined by the autocorrelation function and partial autocorrelation function.

The general form of the ARIMA model can be written as:

$$Y'_t=C+\beta_1 Y_{t-1}+\beta_2 Y_{t-2}+\ldots+\beta_p Y_{t-p}+\varphi_1\varepsilon_{t-1}+\varphi_2\varepsilon_{t-2}+\ldots+\varphi_q\varepsilon_{t-q}+\varepsilon_t\ldots \quad \textbf{(4)}$$

Four main steps for the development of the ARIMA model include checking mean and variance stability (see Table S1 in Multimedia Appendix 1), and identifying *p* and *q* terms (see Figure S1 in Multimedia Appendix 1).

Model Parameter Estimation

The maximum-likelihood approach was used for the model parameters. To determine the best ARIMA model, among the models that passed the residual test (normality and stability in the variance), the model with the lowest Bayesian information criterion (BIC) and Akaike information criterion (AIC) was selected as the final model. The BIC and AIC formulae are represented as follows:

$$BIC=-2.\ln(L)+k.\ln(m) \quad \textbf{(5)}$$

$$AIC=2k-2\ln(L) \quad \textbf{(6)}$$

Where m is the number of observations, k is the total number of parameters in the model, and ln(*L*) is the likelihood function.

The ARIMA model was developed to the time series of the daily average Ct value, daily number of hospitalized patients with COVID-19, new number of daily positive tests, daily number of COVID-19 deaths, and number of hospitalized patients with COVID-19. The detailed method for derivation of the ARIMA model is described in Multimedia Appendix 1.

Cross-correlation Function

To evaluate the time delay between the daily average Ct value and the daily number of hospitalized patients with COVID-19, daily number of new positive tests, daily number of COVID-19 deaths, and number of hospitalized patients with COVID-19 by age, the cross-correlation function was used. The independent (daily average Ct value) and dependent variables (daily number of hospitalized patients with COVID-19, new number of daily positive tests, daily number of COVID-19 deaths, and number of hospitalized patients with COVID-19 by age) were preprocessed by the previously fit ARIMA models. The cross-correlation coefficient is mathematically represented as follows:

$$r_{\alpha\beta(k)}=C\alpha\beta(k)/S_\alpha S_\beta \quad \textbf{(7)}$$

where $C\alpha\beta(k)$ is the value of covariance between the preprocessed input time series and preprocessed output time series at the lag *k*, is the value of the standard deviation of the preprocessing input time series, and is the value of the standard deviation of the preprocessing output time series [23]. Three indicators, Schwarz Bayesian information criterion (SBIC), Hannan-Quinn information criterion (HQIC), and AIC, were used to select the best lag.

$$SBIC=\log(n)k-2\log(L(\hat{\theta})) \quad \textbf{(8)}$$

$$HQIC=-2\ln(L(\hat{\theta}))+2k\log(\log n) \quad \textbf{(9)}$$

In equations (8) and (9), n is the sample size, k is the number of estimated parameters, θ is the set of all parameter values, and $L(\hat{\theta})$ is the likelihood of the model.

ARIMAX Model

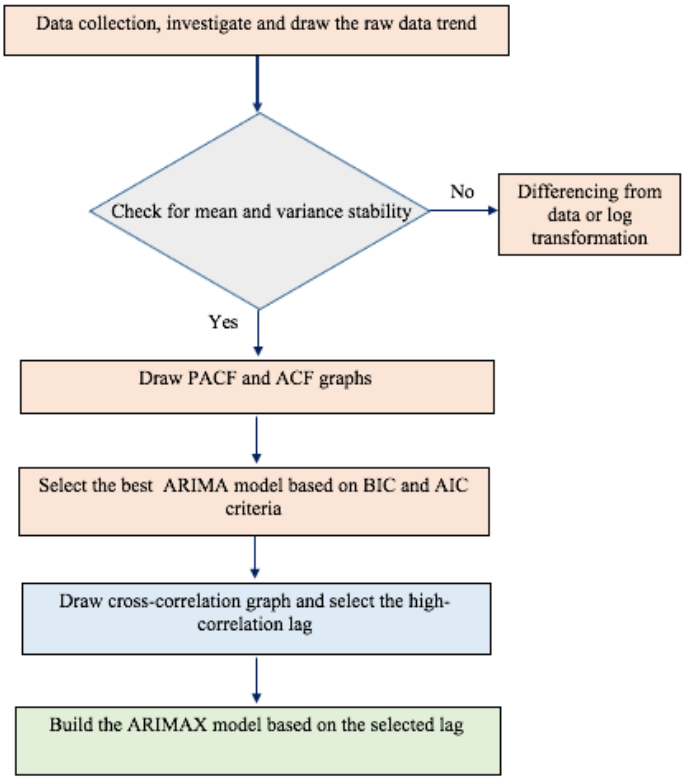
The ARIMAX model is an expansion of the ARIMA model by adding an explanatory independent variable. The ARIMAX model is the combination of multiple regression analysis and time-series analysis; therefore, it can determine the impact factor

of the relationship between different lags of Ct values and other study variables. The ARIMAX model formula is as follows:

$$Y_t = \beta x(t) + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ldots + \alpha_p Y_{t-p} + \varepsilon_t - \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \ldots + \varphi_q \varepsilon_{t-q} + \varepsilon_t \ldots \quad \textbf{(10)}$$

where x(t) is an independent variable at time t and β is its associated coefficient. $Y_{t-1)} \ldots Y_{t-p}$ is the previous value of a dependent variable, and $\varepsilon_t \ldots \varepsilon_{t-q}$ is the residual of the time series. To determine the association and coefficient of the association between the lags of the x_{t+m} time series and series Y_t, the ARIMAX model was used. The cross-correlation function was used to find the linear correlation between x_{t+m} and Y_t for different lags, which can help to find the best lags of the independent variable that might be used to predict the dependent variable [24]. The lags of Ct values that were selected through the correlation function were incorporated as covariates into the ARIMAX model with other dependent variables such as the daily number of hospitalized patients with COVID-19, number of new daily positive cases, daily number of COVID-19 deaths, and number of hospitalized patients with COVID-19 by age. The maximum-likelihood method was used for estimation of the parameters. The Ljung-Box Q test was applied to evaluate white noise for the residual series. Data were analyzed by Stata software version 14. Figure 1 shows the steps of building the best ARIMAX model.

Figure 1. Steps of building the best ARIMAX model. ACF: autocorrelation function; AIC: Akaike information criterion; ARIMA: autoregressive integrated moving average; ARIMAX: autoregressive integrated moving average with exogenous variables; BIC: Bayesian information criterion; PACF: partial autocorrelation function.



Ethics Considerations

Since individual data were not used in this study, no formal ethical assessment or informed consent was required. This study was approved by the Ethics Committee of Iran University of Medical Sciences (ethical code: IR.IUMS.REC.1400.799).

Results

Evaluation Outcomes

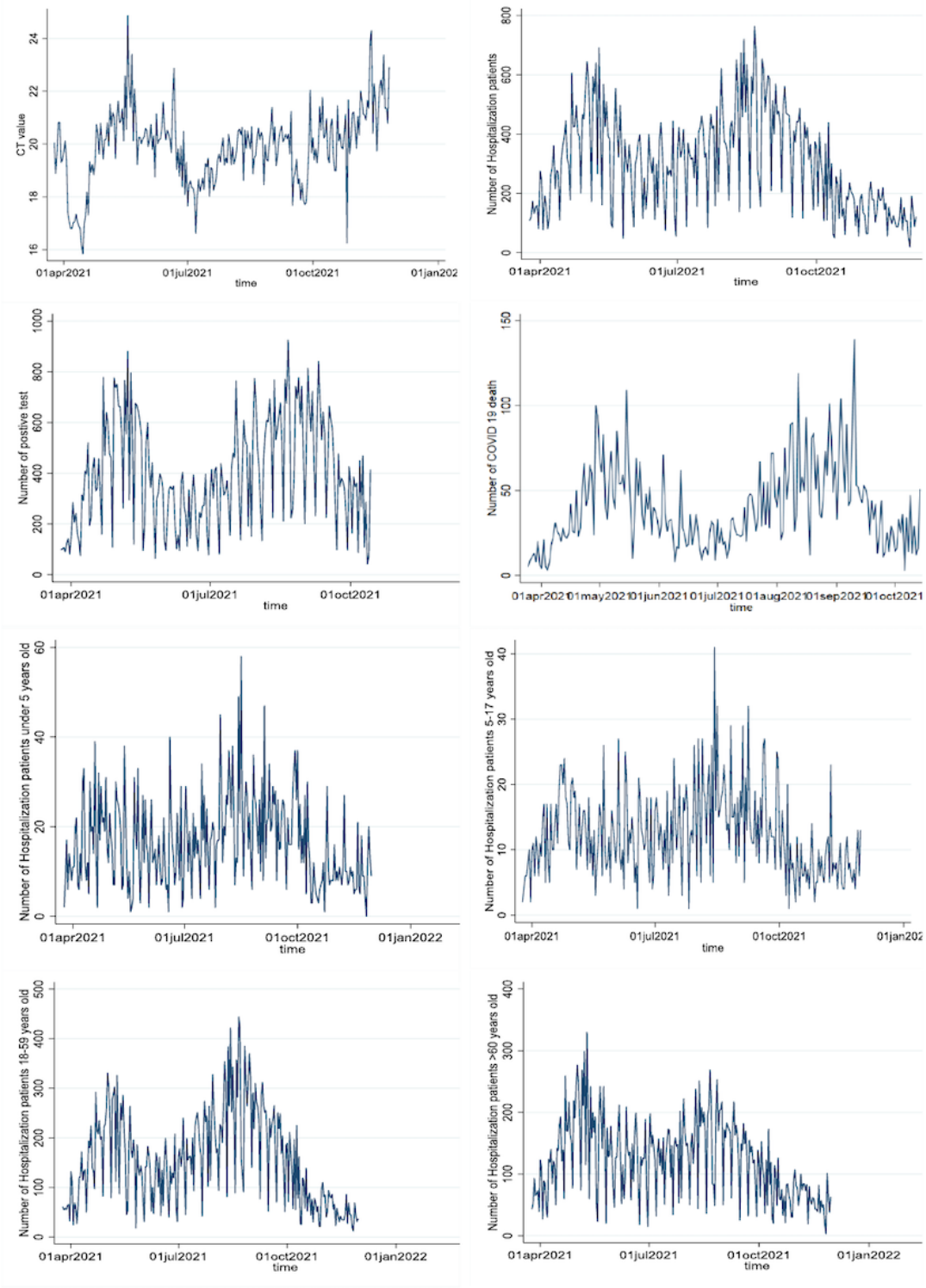
Table 1 shows descriptive statistics of the study variables that were included in the analysis. The minimum value of Ct was related to April 11, 2021, and the maximum frequency of hospitalized patients was related to August 23, 2021. Over 9 months, 80,882 positive COVID-19 tests were referred to the Molecular Diagnostic Laboratory of Iran University of Medical Sciences in Tehran, Iran.

Figure 2 shows the time trend of Ct values, along with the trends of the number of hospitalized patients, number of positive tests, number of COVID-19 deaths, number of hospitalized patients under 5 years old, number of hospitalized patients aged 5-17 years old, number of hospitalized patients aged 18-59 years old, and number of hospitalized patients over 60 years old over the 9 months. Similar to the a priori hypothesis, the daily average Ct value was negatively correlated with the daily number of hospitalized patients, daily count of positive COVID-19 tests (with a time delay), daily number of COVID-19 deaths, and daily number of hospitalized patients by age group. As shown in Figure 2, there was a time delay of approximately 28-32 days between the average daily Ct value and the daily number of hospitalized patients with COVID-19, daily count of positive COVID-19 tests, and daily number of COVID-19 deaths.

Table 1. Descriptive statistics of the study variables.

Variables	Maximum	Minimum	Mean (SD)
Dependent variable: cycle threshold value	24.87	15.83	19.89 (1.33)
Independent variables			
Number of hospitalized patients	763	47	310.65 (260.259)
Number of positive tests	925	42	396.48 (211.05)
Number of COVID-19 deaths	72	0	15.98 (24.57)
Number of hospitalized patients under 5 years old	58	0	16.514 (10.23)
Number of hospitalized patients aged 5-17 years	41	1	12.35 (6.78)
Number of hospitalized patients aged 18-59 years	444	12	155.94 (91.61)
Number of hospitalized patients over 60 years old	330	3	123.58 (63.37)

Figure 2. Trends of cycle threshold (Ct) values and other study variables over 9 months.



ARIMA Model for Study Variables

Table 2 shows the best ARIMA models for the study variables. The ARIMA (1,0,1) model was the best model for the daily average Ct value in comparison with other models, having the lowest BIC value, daily number of the hospitalized patients, and daily count of positive COVID-19 tests. The ARIMA (1,0,2) model was the best model for the daily number of COVID-19 deaths. All models had the lowest number of significant estimated parameters, and the residual analysis showed a good fit (normality and stability in the variance) for the selected ARIMA models using the AIC. There was no seasonal pattern in the study variables. The ADF test was used for evaluating stability in the mean and the Box-Cox test was used to test the time-series stability in the variance. The time series of the daily number of hospitalized patients by age did not show stability for the variance, and therefore log transformation was applied to this variable.

Table 2. The best selected autoregressive integrated moving average (ARIMA) models using the Bayesian information criterion (BIC) and Akaike information criterion (AIC).

Variable	ARIMA	Log likelihood	AIC	BIC
Cycle threshold value	(1,0,1)[a]	–355.99	702.38	716.42
Number of hospitalized patients	(1,0,1)	–1553.31	–1231.24	–1220.61
Number of positive COVID-19 tests	(1,0,1)	–1198.16	2494.99	2827.07
Number of COVID-19 deaths	(1,0,2)	–933.16	1876.33	1893.88
Number of hospitalized patients under 5 years old	(1,0,1)	–905.64	481.26	494.797
Number of hospitalized patients aged 5-17 years	(1,0,1)	–819.80	393.80	407.322
Number of hospitalized patients aged 18-59 years	(1,0,1)	–1401.00	374.58	384.72
Number of hospitalized patients over 60 years old	(1,0,1)	–919.60	397.41	407.55

[a]The numbers in parentheses represent the parameters (*p*, *d*, *q*) of the model, where *p* is the order of autoregression, *d* is the degree of trend difference, and *q* is the order of moving average.

Cross-correlation Analysis

Figure 3 shows the cross-correlations between the study variables and Ct value. In this figure, negative lags would not be considered because the negative lag indicates that the study variables could affect the average Ct value in a certain period at a later point in time; therefore, the positive lag was used to show the effect of the Ct value on the study variables in the future. A cross-correlation function was performed between the preprocessed input and output series. Table 3 shows the best lag difference between the Ct value and the study variables. Indicators such as AIC, SBIC, and HQIC were used to examine the selected lag. There was no statistically significant (all $P>.05$) lag (time delays) between the average Ct value and the daily number of hospitalized patients under 5 years old and the number of hospitalized patients aged 5-17 years. However, a significant 23-day lag was found between the average Ct value and number of hospitalized patients. The daily count of positive COVID-19 tests as well as the daily number of COVID-19 deaths had a significant 30-day lag with the average Ct value.

Figure 3. Cross-correlations (y-axes) between cycle threshold (Ct) values and other study variables.

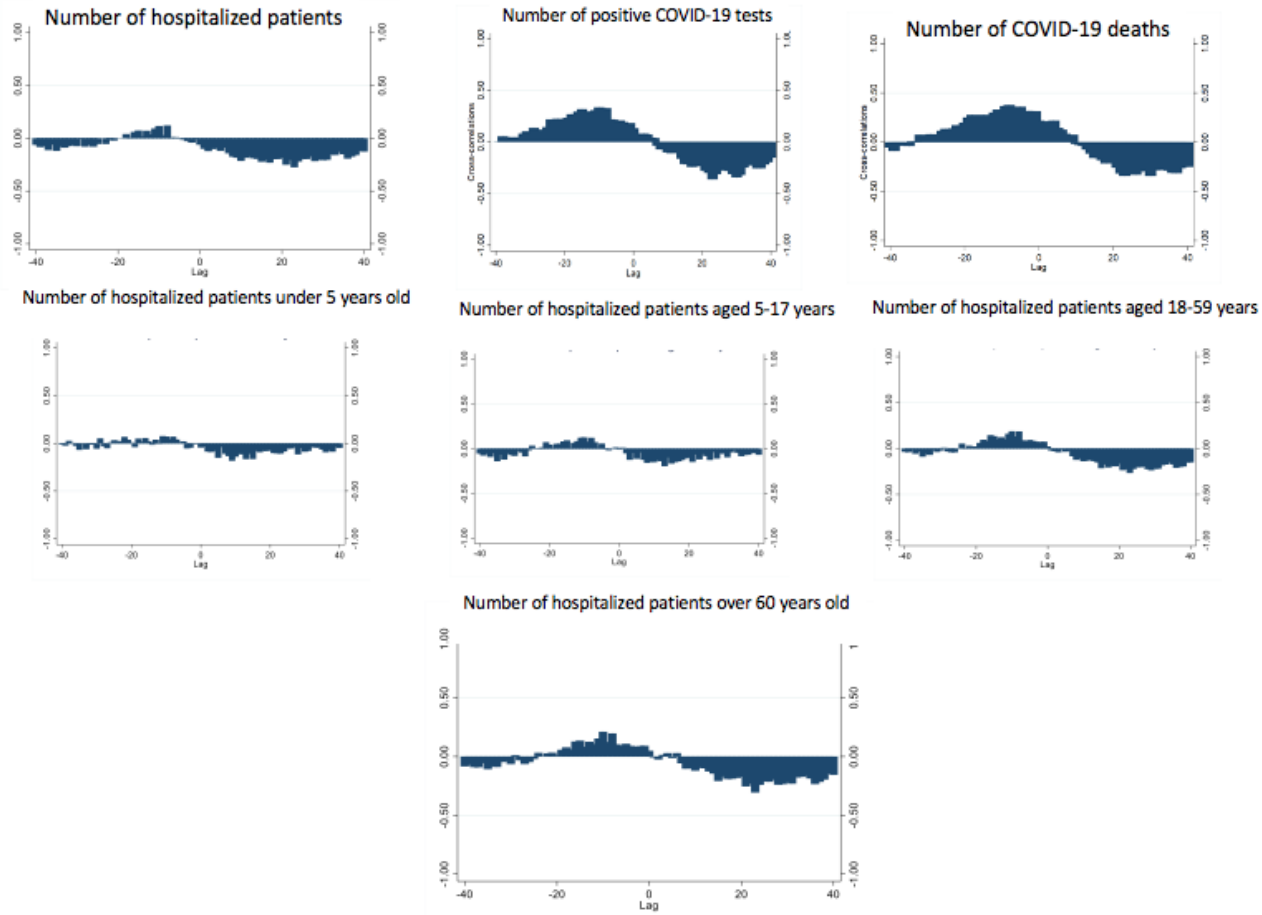


Table 3. High-correlation lags between the cycle threshold value and other study variables.

Variable	Lag	*r*	*P* value	AIC[a]	HQIC[b]	SBIC[c]
Number of hospitalized patients	23	–0.25	.02	15.34	15.95	16.85
Number of positive tests	30	–0.34	.02	15.83	16.83	8.24
Number of COVID-19 deaths	30	–0.26	.02	10.90	11.69	12.86
Number of hospitalized patients under 5 years old	9	–0.22	.76	10.38	10.73	11.25
Number of hospitalized patients aged 5-17 years	13	–0.23	.29	9.54	9.89	10.41
Number of hospitalized patients aged 18-59 years	23	–0.27	.04	14.24	14.85	15.75
Number of hospitalized patients over 60 years old	23	–0.30	.07	13.54	14.15	15.05

[a]AIC: Akaike information criterion.

[b]HQIC: Hannan-Quinn information criterion.

[c]SBIC: Schwarz Bayesian criterion.

Impact of the Ct Value on Study Variables (ARIMAX Model)

After obtaining the best lag between the daily Ct value and other variables using cross-correlation analysis (Table 3), ARIMAX was used to calculate the impact coefficients of the selected lags. Table 4 shows that a Ct value with a 30-day delay could affect the daily number of positive COVID-19 tests and the daily number of deaths from COVID-19. Specifically, a decrease in Ct value may cause an increase of approximately 16.87 times in the average number of new positive tests for COVID-19 after 30 days. In addition, the daily number of deaths from COVID-19 will increase by approximately 1.52 times after 30 days with a decrease in the Ct value. There was a significant coefficient between Ct lag (23 days) and the number of COVID-19 hospitalizations. There was also a significant association of the Ct value with a 23-day delay and the number of COVID-19 hospitalizations for patients aged 18-59 years and patients aged more than 60 years.

Table 4. Estimated coefficients obtained using autoregressive integrated moving average with exogenous variables models.

Variables and parameters	Coefficient (β)	95% CI	*P* value
Number of hospitalized patients; best model: (1,0,1)			
Ct[a] (23)[b]	–24.12	–41.08 to –7.16	.005
AR[c] (1)	.99	.95 to 1.02	<.001
MA[d] (1)	–.87	–.96 to –.78	<.001
Number of COVID-19 deaths; best model: (1,0,2)			
Ct (30)	–1.52	–2.86 to –.18	.03
AR (1)	.96	.89 to 1.03	<.001
MA (1)	–1.07	–1.22 to –.92	<.001
MA (2)	.21	.09 to .34	.001
Number of positive tests; best model: (1,0,1)			
Ct (30)	–16.87	–28.93 to –4.82	<.001
AR (1)	.96	.84 to 1.07	<.001
MA (1)	–.89	–1.06 to –.71	<.001
Number of hospitalized patients under 5 years old; best model: (1,0,1)			
Ct (9)	–.60	–1.68 to .47	.27
AR (1)	.96	.84 to 1.07	<.001
MA (1)	–.89	–1.06 to –.71	<.001
Number of hospitalized patients aged 5-17 years (1,0,1)			
Ct (13)	–.40	–1.30 to .50	.39
AR (1)	.97	.92 to 1.03	<.001
MA (1)	–.89	–.99 to –.79	<.001
Number of hospitalized patients aged 18-59 years; best model: (1,0,1)			
Ct (23)	–11.87	–21.81 to –1.94	.02
AR (1)	.99	.95 to 1.02	<.001
MA (1)	–.85	–.94 to –.76	<.001
Number of hospitalized patients over 60 years old; best model: (1,0,1)			
Ct (23)	–11.44	–17.82 to –5.07	<.001
AR (1)	.99	.96 to 1.02	<.001
MA (1)	–.90	–.98 to –.81	<.001

[a]Ct: cycle threshold.

[b]The numbers in parentheses indicate the lag in days.

[c]AR: autoregressive.

[d]MA: moving average.

Discussion

Principal Findings

The Ct value is a good proxy for viral load, which can offer the possibility of isolating people who have a higher viral load (lower Ct value) and those who have been in contact with these people for the past 5 days to reduce the transmission rate [11]. Therefore, the Ct value can be a good indicator for predicting the state of the disease process in the future. This study investigated the relationship between the population distribution of Ct values obtained from SARS-CoV-2–positive RT-PCR tests and COVID-19 dynamics. The results showed that the daily average Ct value has a significant negative relationship with three study variables of COVID-19 dynamics: daily number of hospitalized patients, daily count of positive COVID-19 tests, and daily COVID-19 deaths. The Ct value can predict the peak of the epidemic curve of the number of new positive COVID-19 patients with an interval of 30 days earlier.

Comparison With Prior Work

This result is consistent with the results of a study by Walker et al [21] showing that a declining population-level Ct value

preceded increases in SARS-CoV-2 positivity tests. Another study showed a negative association between individual Ct values and severity of symptoms of COVID-19 [25]. A few studies have focused on the effect of the population-level Ct value as an indicator for predicting pandemic surges. Consistent with this study, Tso et al [26] showed that daily median Ct values have a negative correlation with the daily count of positive tests, daily transmission rates, and daily number of COVID-19 hospitalizations in the greater El Paso area; they also showed a significant 33-day time delay between daily median Ct values and the daily number of COVID-19 hospitalizations. In this study, we found a significant 23-day time delay between the daily average Ct value and the number of hospitalized COVID-19 patients aged 18-59 years and aged more than 60 years. The former age group represents the major workforce, and are thus more likely to be exposed and become infected with the SARS-CoV-2 virus. Buchan et al [27] showed that the average Ct values were statistically similar among age groups, but patients in the age group of 80-89 years had slightly lower Ct values. According to an epidemiology study in Iran, the majority of hospitalized COVID-19 patients were in the age group of 50-60 years [28]. The relationship between the daily average Ct value and the number of COVID-19 patients aged under 5 years was not significant in this study.

Hay et al [16] estimated the epidemic trajectory in Massachusetts, United States, using a mathematical model for population-level Ct values, and also found that an increasing epidemic wave will be accompanied by a high frequency of recently infected patients with high viral loads (lower Ct values), whereas a declining epidemic wave occurs when the number of patients with older infections is high. Therefore, Ct values obtained from the disease care system during the epidemic of SARS-CoV-2 can determine the course of the epidemic process at short intervals [16]. In this study, the ARIMAX model was used to find the effect of Ct value delay time on the number of positive COVID-19 tests, and a 30-day delay was found between the average population-level Ct value and the number of positive COVID-19 cases.

Limitations

Differences in how measurements of Ct value or assurance about the quality of the data sets that are used to measure population-level Ct values in different geographical areas may affect the power of the Ct value for predicting local COVID-19 epidemic waves. Previous studies have indicated that changes in the population-level Ct values of surveillance samples may lead to a disease outbreak [16,29]. There is a hypothesis that if only patients with clinical symptoms who had positive tests were used to calculate the daily average Ct value, the association between the daily Ct value and COVID-19 cases would be more readily detected; thus, a decrease in Ct values may be more closely associated with the increasing number of COVID-19 patients. To investigate this hypothesis, only the Ct value of patients with symptoms was used to calculate the daily average Ct value in this study.

Conclusions

The daily average population-level Ct value has a relationship with the number of positive SARS-CoV-2 tests and time delay. Thirty days after reducing the daily average Ct value, the number of new COVID-19 cases is expected to increase. It is important to find a good indicator that can predict epidemic surges in the community for improved COVID-19 surveillance. Faster prediction of a new wave of disease will help health policymakers to initiate appropriate public health policies such as lockdowns for decreasing an anticipated pandemic surge, and will provide health systems an opportunity to meet the needs of medicine and facilities to support additional patients.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Additional data related to the statistical analysis steps and detailed description of the ARIMA model.
[DOCX File , 722 KB-Multimedia Appendix 1]

References

1. Woo PCY, Lau SKP, Huang Y, Yuen K. Coronavirus diversity, phylogeny and interspecies jumping. Exp Biol Med 2009 Oct;234(10):1117-1127. [doi: 10.3181/0903-MR-94] [Medline: 19546349]
2. Seah IYJ, Anderson DE, Kang AEZ, Wang L, Rao P, Young BE, et al. Assessing viral shedding and infectivity of tears in coronavirus disease 2019 (COVID-19) patients. Ophthalmology 2020 Jul;127(7):977-979 [FREE Full text] [doi: 10.1016/j.ophtha.2020.03.026] [Medline: 32291098]
3. Liang L, Wu P. There may be virus in conjunctival secretion of patients with COVID-19. Acta Ophthalmol 2020 May 18;98(3):223-223 [FREE Full text] [doi: 10.1111/aos.14413] [Medline: 32189460]
4. Cheng H, Jian S, Liu D, Ng T, Huang W, Lin H, Taiwan COVID-19 Outbreak Investigation Team. Contact tracing assessment of COVID-19 transmission dynamics in Taiwan and risk at different exposure periods before and after symptom onset. JAMA Intern Med 2020 Sep 01;180(9):1156-1163 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2020] [Medline: 32356867]
5. Rao S, Manissero D, Steele V, Pareja J. A systematic review of the clinical utility of cycle threshold values in the context of COVID-19. Infect Dis Ther 2020 Sep;9(3):573-586 [FREE Full text] [doi: 10.1007/s40121-020-00324-3] [Medline: 32725536]

6. Tom M, Mina MJ. To interpret the SARS-CoV-2 test, consider the cycle threshold value. Clin Infect Dis 2020 Nov 19;71(16):2252-2254 [FREE Full text] [doi: 10.1093/cid/ciaa619] [Medline: 32435816]
7. Walsh KA, Jordan K, Clyne B, Rohde D, Drummond L, Byrne P, et al. SARS-CoV-2 detection, viral load and infectivity over the course of an infection. J Infect 2020 Sep;81(3):357-371 [FREE Full text] [doi: 10.1016/j.jinf.2020.06.067] [Medline: 32615199]
8. Emergency use authorization (EUA) summary COVID-19 RT-PCR test (Laboratory Corporation of America). US Food and Drug Administration. 2020. URL: https://www.fda.gov/media/136151/download [accessed 2022-10-23]
9. Jaafar R, Aherfi S, Wurtz N, Grimaldier C, Van Hoang T, Colson P, et al. Correlation between 3790 quantitative polymerase chain reaction-positives samples and positive cell cultures, including 1941 severe acute respiratory syndrome coronavirus 2 isolates. Clin Infect Dis 2021 Jun 01;72(11):e921-e921 [FREE Full text] [doi: 10.1093/cid/ciaa1491] [Medline: 32986798]
10. La Scola B, Le Bideau M, Andreani J, Hoang VT, Grimaldier C, Colson P, et al. Viral RNA load as determined by cell culture as a management tool for discharge of SARS-CoV-2 patients from infectious disease wards. Eur J Clin Microbiol Infect Dis 2020 Jun 27;39(6):1059-1061 [FREE Full text] [doi: 10.1007/s10096-020-03913-9] [Medline: 32342252]
11. Sarkar B, Sinha R, Sarkar K. Initial viral load of a COVID-19-infected case indicated by its cycle threshold value of polymerase chain reaction could be used as a predictor of its transmissibility - An experience from Gujarat, India. Indian J Community Med 2020;45(3):278. [doi: 10.4103/ijcm.ijcm_593_20]
12. Zhang X, Lu S, Li H, Wang Y, Lu Z, Liu Z, et al. Viral and antibody kinetics of COVID-19 patients with different disease severities in acute and convalescent phases: a 6-month follow-up study. Virol Sin 2020 Dec 22;35(6):820-829 [FREE Full text] [doi: 10.1007/s12250-020-00329-9] [Medline: 33351168]
13. Choudhuri J, Carter J, Nelson R, Skalina K, Osterbur-Badhey M, Johnston A, et al. SARS-CoV-2 PCR cycle threshold at hospital admission associated with patient mortality. PLoS One 2020 Dec 31;15(12):e0244777 [FREE Full text] [doi: 10.1371/journal.pone.0244777] [Medline: 33382805]
14. Miranda R, Guterres A, de Azeredo Lima CH, Filho PN, Gadelha MR. Misinterpretation of viral load in COVID-19 clinical outcomes. Virus Res 2021 Apr 15;296:198340 [FREE Full text] [doi: 10.1016/j.virusres.2021.198340] [Medline: 33592214]
15. Binnicker MJ. Challenges and Controversies to Testing for COVID-19. J Clin Microbiol 2020 Oct 21;58(11):e01695-20 [FREE Full text] [doi: 10.1128/JCM.01695-20] [Medline: 32817231]
16. Hay J, Kennedy-Shaffer L, Kanjilal S, Lennon N, Gabriel S, Lipsitch M, et al. Estimating epidemiologic dynamics from cross-sectional viral load distributions. Science 2021 Jul 16;373(6552):2021 [FREE Full text] [doi: 10.1126/science.abh0635] [Medline: 34083451]
17. Sheikhzadeh E, Eissa S, Ismail A, Zourob M. Diagnostic techniques for COVID-19 and new developments. Talanta 2020 Dec 01;220:121392 [FREE Full text] [doi: 10.1016/j.talanta.2020.121392] [Medline: 32928412]
18. Brockwell P, Brockwell P, Davis R. Introduction to time series and forecasting. New York, NY: Springer; 2016.
19. Malki Z, Atlam E, Ewis A, Dagnew G, Alzighaibi AR, ELmarhomy G, et al. ARIMA models for predicting the end of COVID-19 pandemic and the risk of second rebound. Neural Comput Appl 2021 Oct 23;33(7):2929-2948 [FREE Full text] [doi: 10.1007/s00521-020-05434-0] [Medline: 33132535]
20. Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos Solitons Fractals 2020 Jun;135:109864 [FREE Full text] [doi: 10.1016/j.chaos.2020.109864] [Medline: 32390691]
21. Diop ML, Kengne W. Piecewise autoregression for general integer-valued time series. J Stat Plan Inference 2021 Mar;211:271-286. [doi: 10.1016/j.jspi.2020.07.003]
22. Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. Sci Total Environ 2020 Aug 10;729:138817 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.138817] [Medline: 32360907]
23. Agustina D, Yosmar S, Rizal J. The identification of tsunami height correlation model with earthquake parameters. 2017 Presented at: IOP Conference Series: Earth and Environmental Science, The 4th International Seminar on Sciences; October 19-20, 2017; Bogor, Indonesia. [doi: 10.1088/1755-1315/187/1/012076]
24. Abolmaali S, Shirzaei S. A comparative study of SIR model, linear regression, logistic function and ARIMA model for forecasting COVID-19 cases. AIMS Public Health 2021;8(4):598-613 [FREE Full text] [doi: 10.3934/publichealth.2021048] [Medline: 34786422]
25. Rabaan AA, Tirupathi R, Sule AA, Aldali J, Mutair AA, Alhumaid S, Muzaheed, et al. Viral dynamics and real-time RT-PCR Ct values correlation with disease severity in COVID-19. Diagnostics 2021 Jun 15;11(6):1091 [FREE Full text] [doi: 10.3390/diagnostics11061091] [Medline: 34203738]
26. Tso CF, Garikipati A, Green-Saxena A, Mao Q, Das R. Correlation of population SARS-CoV-2 cycle threshold values to local disease dynamics: exploratory observational study. JMIR Public Health Surveill 2021 Jun 03;7(6):e28265 [FREE Full text] [doi: 10.2196/28265] [Medline: 33999831]
27. Buchan B, Hoff J, Gmehlin C, Perez A, Faron M, Munoz-Price L, et al. Distribution of SARS-CoV-2 PCR cycle threshold values provide practical insight into overall and target-specific sensitivity among symptomatic patients. Am J Clin Pathol 2020 Sep 08;154(4):479-485 [FREE Full text] [doi: 10.1093/ajcp/aqaa133] [Medline: 32687186]
28. Nikpouraghdam M, Jalali Farahani A, Alishiri G, Heydari S, Ebrahimnia M, Samadinia H, et al. Epidemiological characteristics of coronavirus disease 2019 (COVID-19) patients in IRAN: a single center study. J Clin Virol 2020 Jun;127:104378 [FREE Full text] [doi: 10.1016/j.jcv.2020.104378] [Medline: 32353762]

29. Walker A, Pritchard E, House T, Robotham J, Birrell P, Bell I, COVID-19 Infection Survey team. Ct threshold values, a proxy for viral load in community SARS-CoV-2 cases, demonstrate wide variation across populations and over time. Elife 2021 Jul 12;10:e64683. [doi: 10.7554/eLife.64683] [Medline: 34250907]

Abbreviations

ADF: augmented Dickey-Fuller
AIC: Akaike information criterion
AR: autoregressive
ARIMA: autoregressive integrated moving average
ARIMAX: autoregressive integrated moving average with exogenous variables
BIC: Bayesian information criterion
Ct: cycle threshold
HQIC: Hannan-Quinn information criterion
MA: moving average
RT-PCR: reverse transcription-polymerase chain reaction
SBIC: Schwarz Bayesian information criterion

Edited by A Mavragani, G Eysenbach; submitted 14.01.22; peer-reviewed by M Yousefi, M Mahmoodi, M Mahmoodi; comments to author 18.05.22; revised version received 30.05.22; accepted 11.10.22; published 08.11.22

Please cite as:
Dehesh P, Baradaran HR, Eshrati B, Motevalian SA, Salehi M, Donyavi T
The Relationship Between Population-Level SARS-CoV-2 Cycle Threshold Values and Trend of COVID-19 Infection: Longitudinal Study
JMIR Public Health Surveill 2022;8(11):e36424
URL: https://publichealth.jmir.org/2022/11/e36424
doi: 10.2196/36424
PMID: 36240022