Heliyon 10 (2024) e25090

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

COVID-19 pandemic waves: Identification and interpretation of global data

Ranjula Bali Swain [a,b,*], Xiang Lin [a], Fan Yang Wallentin [c]

[a] Department of Economics, Södertörn University, 141 89 Huddinge, Stockholm, Sweden
[b] Center for Sustainability Research (SIR), Stockholm School of Economics, Box 6501, SE-11383, Stockholm, Sweden
[c] Department of Statistics, Uppsala University, Uppsala, Sweden

ABSTRACT

The mention of the COVID-19 waves is as prevalent as the pandemic itself. Identifying the beginning and end of the wave is critical to evaluating the impact of various COVID-19 variants and the different pharmaceutical and non-pharmaceutical (including economic, health and social, etc.) interventions. We demonstrate a scientifically robust method to identify COVID-19 waves and the breaking points at which they begin and end from January 2020 to June 2021. Employing the Break Least Square method, we determine the significance of COVID-19 waves for global-, regional-, and country-level data. The results show that the method works efficiently in detecting different breaking points. Identifying these breaking points is critical for evaluating the impact of the economic, health, social and other welfare interventions implemented during the pandemic crisis. Employing our method with high frequency data effectively determines the start and end points of the COVID-19 wave(s). Identifying waves at the country level is more relevant than at the global or regional levels. Our research results evidenced that the COVID-19 wave takes about 48 days on average to subside once it begins, irrespective of the circumstances.

1. Introduction

Since the beginning, humanity has remained preoccupied with the three challenges of famine, war, and plague [1,2]. More people have died from diseases than war and famines put together. Historically, populations were kept low and dispersed by a high death rate driven by infection [3]. The pandemic threat has multiplied with greater globalization, urbanization, and connectivity [4]. Plague, for example, has been described as a disease of trade [5]. By mid-January 2022, COVID-19 had infected 313 million, resulting in over 5.5 million reported deaths around the globe.

Determining the exact date of the start and end of the pandemic wave is critical to evaluate the impact of the economic, health, social and other welfare interventions implemented during the pandemic crisis [6–8]. Furthermore, instead of selecting ad hoc time periods, knowledge of the breaking points is essential for evaluating the effectiveness of policy measures implemented to limit the spread of COVID-19, such as school closings, travel restrictions, bans on public gatherings, emergency investments in healthcare facilities, social welfare provision, contact tracing, etc. [7,9–13].

In this paper, we scientifically identify the COVID-19 wave(s) and the breaking point at which each COVID-19 wave begins and ends from the January 23, 2020 till June 3, 2021. Unlike a ''spike", which is a temporary increase in new COVID infections, the "wave" may be calibrated as a ''sustained" period of the upward and downward period [7]. Employing Break Least Square regression (BLS) on the Johns Hopkins Center for Systems Science and Engineering (CSSE) (2019) data [14] from January 23, 2020 to June 3, 2021, we perform global, regional, and country level analysis.

* Corresponding author. Center for Sustainability Research (SIR), Stockholm School of Economics, Box 6501, SE-11383, Stockholm, Sweden.
E-mail addresses: Ranjula.Bali@sh.se (R. Bali Swain), Xiang.Lin@sh.se (X. Lin), Fan.Yang@statistics.uu.se (F.Y. Wallentin).

https://doi.org/10.1016/j.heliyon.2024.e25090
Received 1 May 2023; Received in revised form 4 January 2024; Accepted 19 January 2024
Available online 27 January 2024
2405-8440/

After testing the data for stationarity (Supplementary Information (SI) Tables 1a and 1b), the BLS is estimated for identifying the structural breaks and associated segments on the unadjusted and adjusted effective reproduction rate (where we replace R_t equal to 1 when $R_t < 1$). The data is adjusted to avoid possible misidentification of the start date of a wave where there is a switch from low R_t ($R_t < 1$) values, but this could result in a longer COVID wave duration. We choose 1.1, which is the mean of R_t as the cut-off value (refer to Table 1a in SI), to identify the segments of the wave. For further details on the method, refer to SI.

The analysis is conducted at the global, continent, and national levels. In addition, a subset of countries, including the United States; Germany, Italy, France, Netherlands, UK, Spain, and Sweden in Europe; India, South Korea, Japan, and Singapore in Asia; South Africa and Nigeria in Africa; and Brazil and Peru South America are investigated.

We find that COVID-19 waves typically take around 48 days to subside, regardless of the initial conditions. This finding provides crucial guidance for policymakers and healthcare authorities, enabling them to allocate resources more efficiently and plan interventions more precisely. The outcomes of this study offer invaluable insights into the temporal patterns of COVID-19 waves, which will prove particularly beneficial in devising more effective mitigation strategies and optimizing resource distribution. The primary audience for this research encompasses a wide spectrum of professionals, including public health officials, epidemiologists, policymakers, and researchers specializing in pandemic preparedness and response. This discussion underscores the broad relevance and potential impact of our findings in advancing the collective efforts to combat COVID-19. The following section carries out the literature review. Section 3 explains the Break Least Square Method, which identifies the start and end points of the COVID-19 wave. Section 4 presents results, and section 5 carries out the discussion. The final section provides conclusive remarks.

2. Literature review

Like the Spanish flu pandemic of 1918 [15], the COVID-19 pandemic also occurred in waves. However, evidence is limited on whether these waves originate from the nature of the disease, pharmaceutical, economic and other non-pharmaceutical interventions, or human behaviour [16,17]. The COVID-19 wave is a culmination and interactive effect of all these factors. Second, the COVID-19 wave does not have a standard definition [6] and the spread of the pandemic is unsynchronised and regionally divergent [18]. Third, when exposed to an unknown virus, the governments struggled with pandemic related uncertainty and made crucial decisions on the timing, duration and intensity of interventions [19,20].

The standard model in epidemiology is the Susceptible Infected and Recovered (SIR) model [21] or complex networks [22–24], which are commonly used to describe the exponential spread of the disease. The spread of the infection in our analysis is measured using the effective reproduction rate (R_t, hereafter reproduction rate) as estimated by Ref [25]. They combine the basic epidemiological theory with standard time-series filtering techniques (Kalman) to estimate a transparent closed-form estimator. R_t measures the real-time average number of secondary cases produced by a primary case, as adjusted for the depletion of susceptible individuals and changes in control measures, contact rates, and climatic conditions. For example, in January 2020, the basic reproduction rate (R_0) in Wuhan, China, was calculated to be between two and three; however, after the lockdown, estimates put the R_t at just over one [26]. R_t is critical for evaluating public policy decisions during a pandemic [27,28]. Some social scientists argue that $R_t < 1$ should be taken as a sign for restricting the use of public policy during the pandemic [29]. Recent literature empirically tries to identify COVID-19 waves [30,31] based on the state of R being larger/smaller than 1 and the minimum length of persistent developments. These studies find that, on average, the duration of waves is about 61.7 days.

3. Method

3.1. Break Least Square method (BLS)

The model: The mean of the reproduction rate R_t is modeled as the mean c with the error term ε_t:

$$R_t = c + \varepsilon_t. \tag{1}$$

In the current study, we employ the Break Least Squares (BLS) estimation, which allows the mean to be time-varying with different values, c_j, where $j = 1, \cdots, M$. M is the number of segments in terms of distinguishing values of the mean of R_t. There are M-1 corresponding breaks at dates, $T_1, T_2, \cdots, T_{M-1}$. Note that $T_0 = 1$ and $T_M = T$ are the dates of initial and final observations of the series of R_t, respectively. equation (1) can be modified as

$$R_t = c_j + \varepsilon_t \quad \text{for } j = 1, \cdots, M \text{ and } j - 1 \leq t \leq j. \tag{2}$$

Equation (2) is a pure structural change model, a simplified version of the original multivariant model, which also permits coefficients to be irrelevant to the breaks [32]. Given the specified dates T_j associated with the number of segments, M (M-1 breaks) [32], the sum-of-squared residuals (RSS) is defined as

$$RSS(c_j) = \sum_{j=1}^{M} \left\{ \sum_{t=T_{j-1}}^{T_j} (R_t - c_j) \right\}^2. \tag{3}$$

The least square method by minimizing (3) can be employed to obtain the best-fitted values $\widehat{c}_j$ in each segment with the known dates of breaks.

Generally, the number of breaks is unknown. Since the main objective of this paper is to identify the start and end date of the COVID waves and their length, the crucial task is to identify the number of breaks and associated dates.

This paper adopts the global optimizing procedure for identifying multiple breaks developed by Refs [32,33]. In comparison with the sequential procedure, the global procedure has the advantage of identifying any given number of breaks. It minimizes (3) subject to T_j when M is given. As pointed out by Ref [34], the sequential approach may sometimes fail to identify the breaks when the true number of breaks is unknown.

Nevertheless, there is no guarantee that introducing an additional break could significantly reduce the RSS. This is important to determine the optimal number of breaks when the true breaks are unknown. The global optimizing procedure is, thus, also referred to as a double maximization, reflecting the fact that the algorithm for searching optimal numbers for 'breaks' is built on two steps. The first step contains a series of "zero versus l breaks" tests, which maximize the given breaks. The second step is the integrated tests for identifying the number of breaks built on maximizing the statistics from the tests in step one.

The "zero versus l breaks" test is a multiple breaks test, an extension of the single-break approach developed in Refs [35,36]. When the number of breaks, l, is pre-specified for the equality of $\hat{c}_i$ across the segments, the hypothesis $H_0 : \hat{c}_1 = \cdots = \hat{c}_l$, against $H_A : \hat{c}_i \neq \hat{c}_{i+1}$ for some i, is tested. This test is based on F-test in which the F-ratio, denoted as $F(l)$, is defined as the unrestricted SSR against the restricted SSR with the alternative. The higher the F-ratio, the more likely to reject the null of equality.

With an unknown number of breaks, the "zero versus l breaks" tests are carried out sequentially from l=1 to the assigned maximum break l^{max} or until the breaks remain significant, m^*, when $F(m^*)$ is significant, and $F(m^*+1)$ is insignificant. Let $m^{**} = \min(l^{max}, m^*)$. Refs [33,37] propose two statistics to determine the optimal number of breaks, UD_{MAX} and WD_{MAX}. We denote $D_{MAX} = \max_{l=1,\cdots,m^{**}} w_l F(l)$. Note that the definition UD_{MAX} is D_{MAX} when $w_l = 1$. And WD_{MAX} when w_l is a function of asymptotic critical values to make the same implied p-values. These statistics do not follow ordinary F-distribution due to the non-nest features. Ref [37] provide the critical values for testing the significance of the number of breaks. Ref [34] made the critical values available for small samples (less than 50 observations).

To identify the breaking points of COVID-19 waves, a maximum number of breaks, l^{max}, is set at 8, which corresponds 10 % trimming percentage. The maximum number of breaks sometimes reduce to 6 for the adjusted data. When the optimal number of breaks is identified according to the global optimizing procedure, the mean values in all segments are estimated. Due to the estimating errors and using the average values of reproduction rate, a critical value of 1.1 (instead of 1) has been adopted to classify whether a segment is a part of the COVID wave. The mean value is above 1.1, indicating the segment is part of the COVID wave. If several allied segments simultaneously exceed the critical value, they are classified as a single wave. The duration of a wave is the sum of the durations of all involved segments. The values of the means are reported separately.

BLS has a few advantages compared to alternative procedures. First, the BLS does not rely on the normality assumption. Other procedures based on the likelihood function require the underlying distributions to be normal, for instance, the Markov Regime Switching model [38,39]. In addition, the Markov Regime Switching model can only identify two or three regimes (convergence in the higher order of the Markov Switching model is problematic). Thus, it is useful in identifying extreme situations like a higher (est) effective reproduction rate R_t.

The BLS can also deal with possible specification errors, such as heterogeneity and autocorrelation with the common robust stand errors. Another alternative method, the change point analysis [40,41] requires the underlying series to be free from these errors. BLS can deal with autocorrelation directly. This can greatly improve the precision of the estimation; however, it makes it hard to identify the factor causing the differences across the segments due to the lag of the series. Determination of the COVID-19 wave involves a comparison between the estimated values of the mean and the cutting point of R_t, 1.1. However, making more precise forecasting is not the primary goal of this paper. Thus, the paper adopts a standard BLS estimator with robust standard errors.

3.2. Stationarity and stationarity with breakpoint tests

As a requirement for using BLS, the underlying series needs to be stationary or at least stationary with a break. To ensure the stationarity of the underlying reproduction rate R_t series. The augmented Dickey-Fuller (ADF) [42] (Dickey and Fuller, 1979) and ADF with a break (BADF) [43] stationarity tests are implemented. The specification of ADF is with an intercept and a linear trend. The number of maximum lags is set at 17. Schwarz Information Criterion (SIC) is adopted to select the optimal lags. The BADF includes an intercept, and the breaking point is allowed in the intercept. The break type is an additive outlier. The number of breakpoints is determined by minimizing the Dickey-fuller t-statistic. The number of maximum lags is set at 17. The same SIC is adopted to select the optimal lags. Note that the current method can only identify one break point. Thus, identifying the waves that require multiple breaks may be inappropriate.

3.3. Data

The analysis in this study is based on the Johns Hopkins Center for Systems Science and Engineering (CSSE) (2019) data [14] from the beginning of the pandemic, January 23, 2020, to June 3, 2021. This dataset includes data from various aggregated data sources, namely, the World Health Organization (WHO), the European Centre for Disease Prevention and Control (ECDC), DXY.cn. Pneumonia. 2020, QQ News, US CDC, BNO, WorldoMeters, 1Point3Arces, COVID Tracking Project, Los Angeles Times, The Mercury News. For further details and other non-US data sources at the country/region and US data sources, refer to https://github.com/CSSEGISandData/COVID-19, as accessed on September 4, 2023.

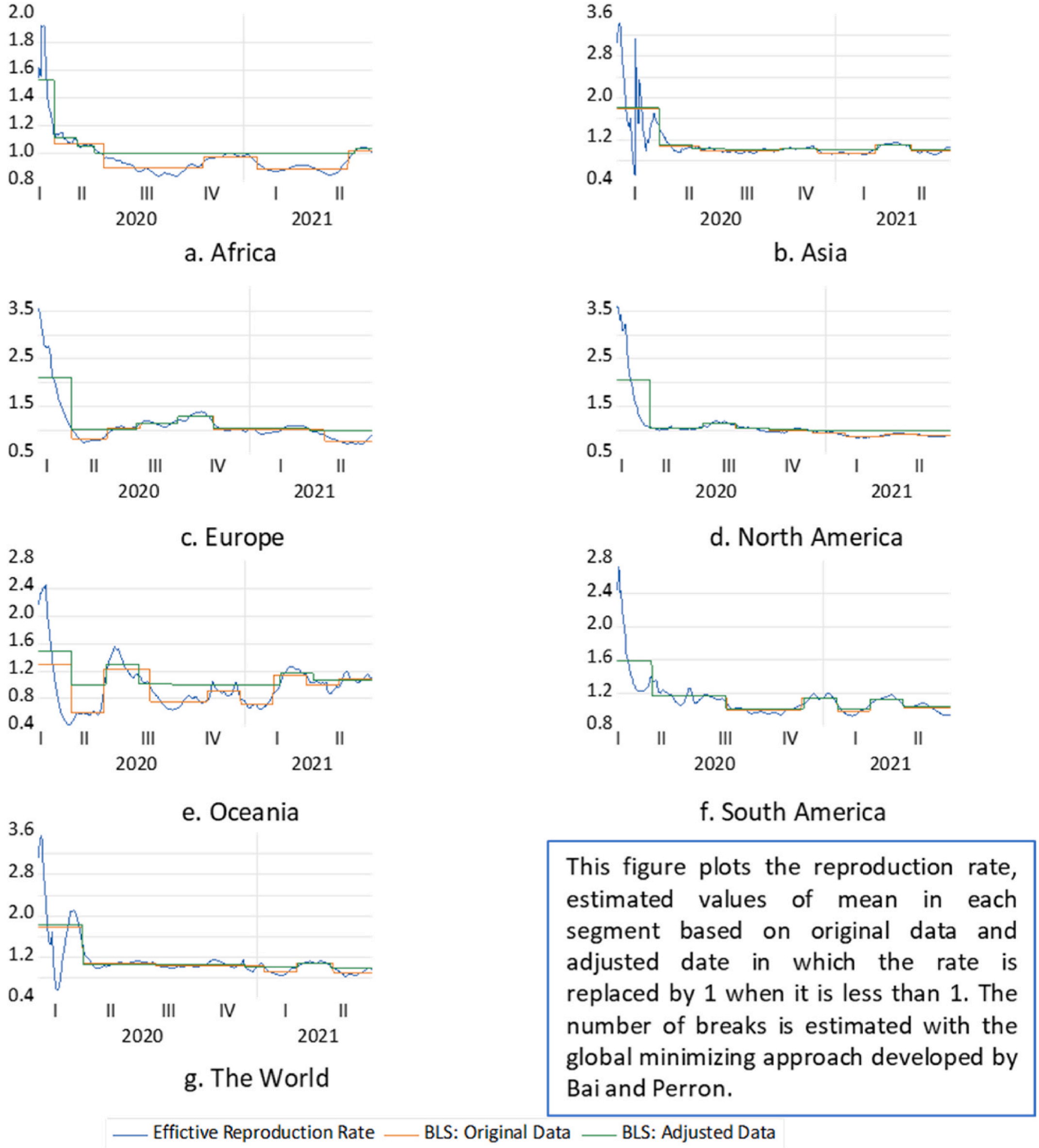

Fig. 1. Break Least Square, Adjusted Break Least Square and Reported data for Reproduction Rate for the different Regions/Continents and the World 2020.

Table 1
Time and Duration of the COVID-19 waves, as identified by Break Least Square, Adjust Break Least Square, and Reported data for Reproduction Rate for different Regions/Continents.

Continent	Break LS: original data	Break LS: adjusted data
ASIA		
First wave		
Mean value(s)	1.79	1.81
Start	2020-01-23	2020-01-23
End:	2020-03-29	2020-03-29
Duration:	67 days +	67 days +
EUROPE		
First wave		
Mean value(s)	2.11	2.11
Start	2020-02-24	2020-02-24
End:	2020-04-12	2020-04-12
Duration:	49 days +	49 days +
Second wave		
Mean value(s)	1.14/1.29	1.14/1.29
Start	2020-07-24	2020-07-18
End:	2020-11-08	2020-11-07
Duration:	106 days	113 days
NORTH AMERICA		
First wave		
Mean value(s)	2.07	2.07
Start	2020-03-05	2020-03-05
End:	2020-04-21	2020-04-21
Duration:	48 days +	48 days +
Second wave		
Mean value(s)	1.13	1.14
Start	2020-07-07	2020-07-07
End:	2020-08-23	2020-08-24
Duration:	48 days	49 days
OCEANIA		
First wave		
Mean value(s)	1.29	1.49/1.26
Start	2020-03-11	2020-03-11
End:	2020-04-26	2020-04-26
Duration:	47 days +	47 days +
Second wave		
Mean value(s)	1.23	1.30
Start	2020-06-13	2020-06-16
End:	2020-08-16	2020-08-01
Duration:	65 days	47 days
Second wave		
Mean value(s)	1.14	1.17
Start	2021-02-10	2021-02-20
End:	2021-03-28	2021-04-07
Duration:	47 days	47 days
AFRICA		
First wave		
Mean value(s)	1.52	1.52/1.11
Start	2020-03-15	2020-03-15
End:	2020-04-06	2020-05-08
Duration:	23 days +	55 days +
SOUTH AMERICA		
First wave		
Mean value(s)	1.58/1.17	1.58/1.17
Start	2020-03-14	2020-03-14
End:	2020-08-16	2020-08-15
Duration:	156 days +	155 days +
Second wave		
Mean value(s)	1.14	1.15
Start	2020-12-01	2020-12-05
End:	2021-03-08	2021-01-20
Duration:	47 days	47 days
Third wave		
Mean value(s)	1.13	1.13
Start	2021-03-09	2021-03-09
End:	2021-04-24	2021-04-24
Duration:	47 days	47 days

Notes: Mean values provide the mean effective reproduction rate mean value.

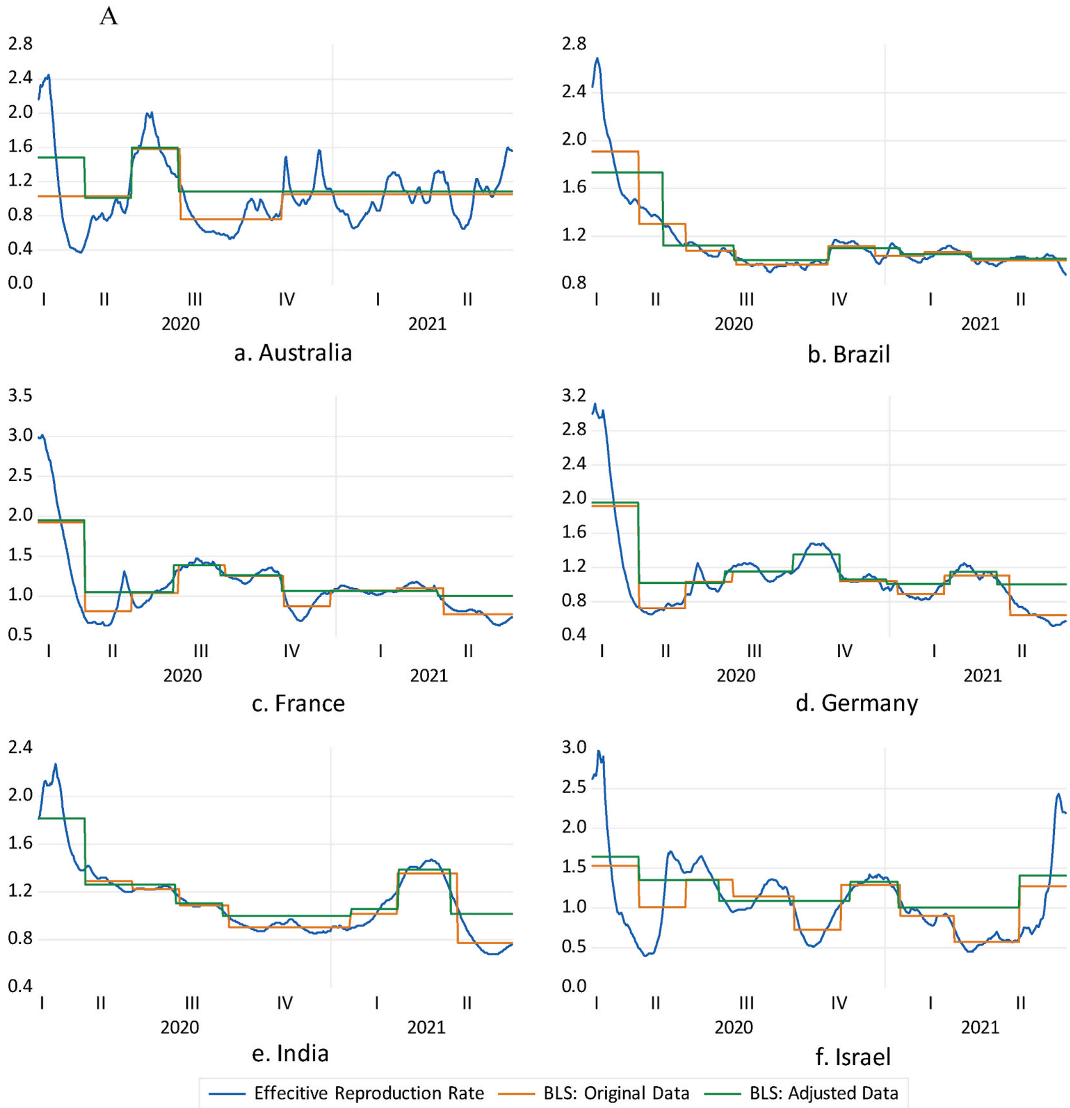

Fig. 2. Break Least Square, Adjusted Break Least Square and Reported data for Reproduction Rate for Selected Countries Fig. 2A–. 2B, C

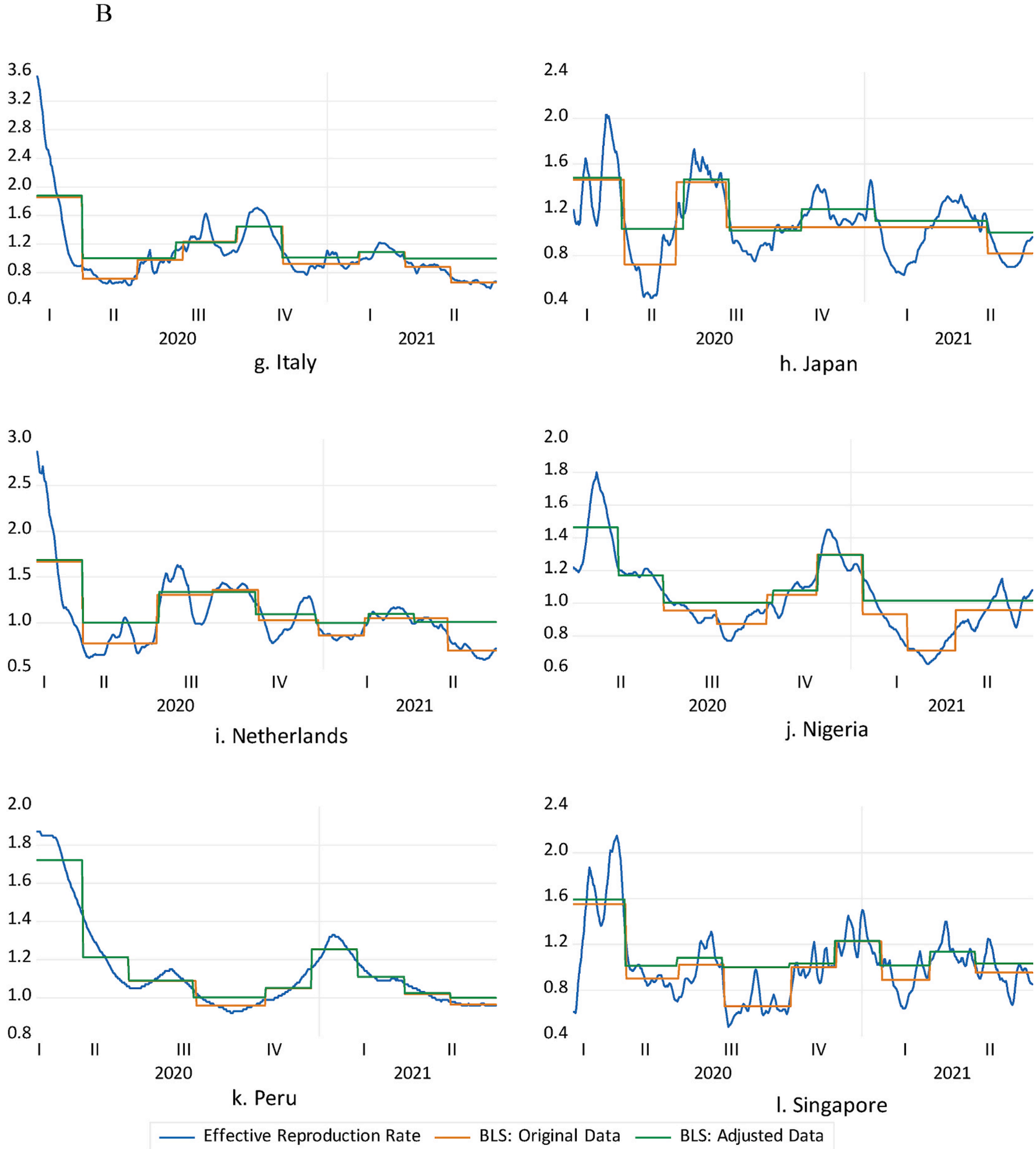

Fig. 2. (*continued*).

C

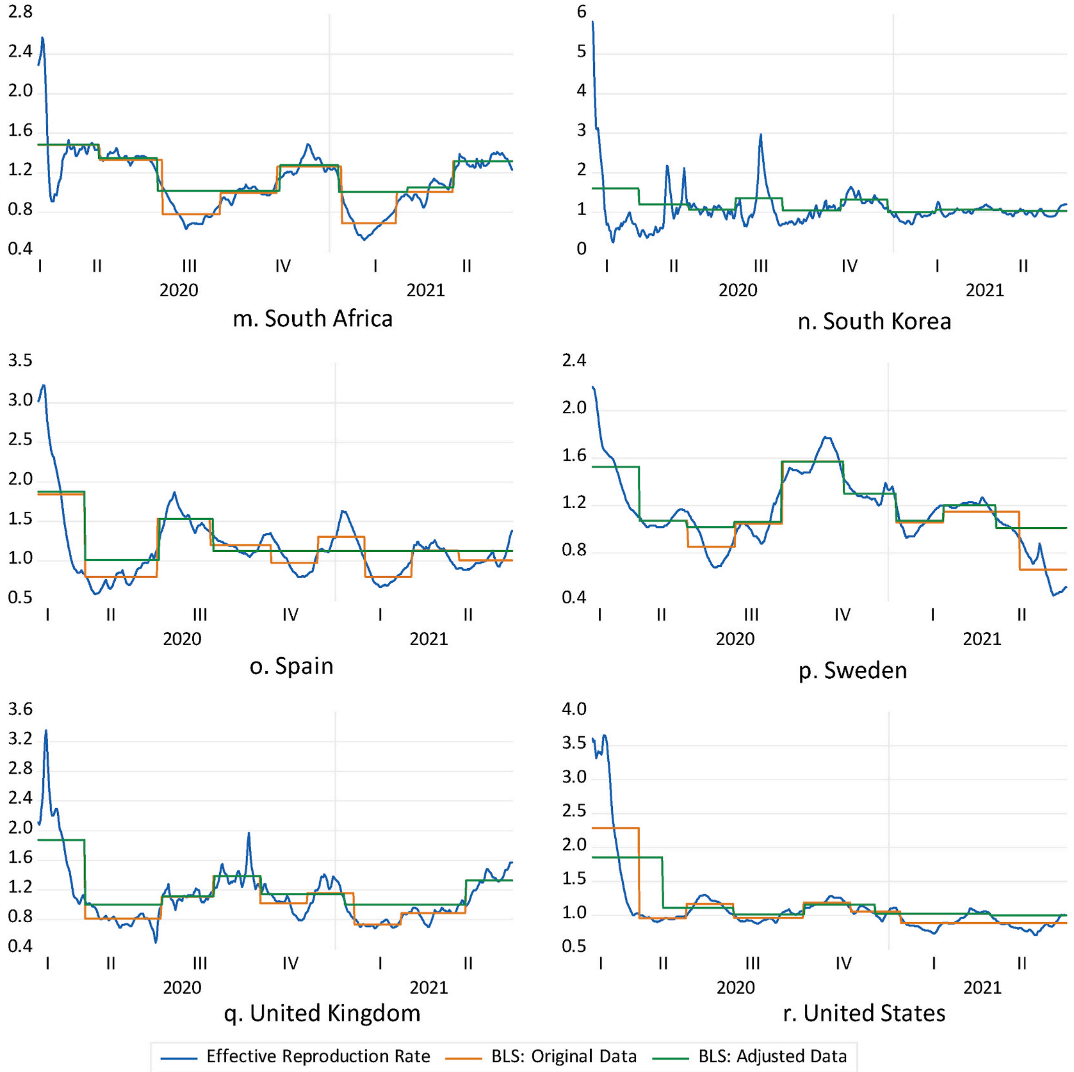

Fig. 2. (*continued*).

Table 2A
Time and Duration of the COVID-19 waves, as identified by Break Least Square, Adjust Break Least Square, and Reported data for Reproduction Rate for Selected Countries.

Country	Break LS: original data	Break LS: adjusted data
Australia		
First wave		
Mean value(s)	1.58	1.48
Start	2020-08-01	2020-03-11
End:	2020-10-11	2020-04-26
Duration:	102 days +	47 days +
Second wave		
Mean value(s)		1.60
Start		2020-06-13
End:		2020-07-29
Duration:		47 days
Brazil		
First wave		
Mean value(s)	1.91/1.30	1.73/1.12
Start	2020-03-14	2020-03-14
End:	2020-06-15	2020-08-02
Duration:	94 days +	142 days +
Second wave		
Mean value(s)	1.12	
Start	2020-11-05	
End:	2020-12-21	
Duration:	47 days	
France		
First wave		
Mean value(s)	1.93	1.95
Start	2020-03-01	2020-03-01
End:	2020-04-17	2020-04-17
Duration:	48 days +	48 days +
Second wave		
Mean value(s)	1.39/1.25	1.39/1.26
Start	2020-07-23	2020-07-18
End:	2020-11-07	2020-11-05
Duration:	108 days	111 days
Third wave		
Mean value(s)	1.10	
Start	2021-03-04	
End:	2021-04-20	
Duration:	48 days	
Germany		
First wave		
Mean value(s)	1.92	1.96
Start	2020-03-03	2020-03-02
End:	2020-04-18	2020-04-18
Duration:	48 days +	48 days +
Second wave		
Mean value(s)	1.16/1.35	1.16/1.35
Start	2020-07-24	2020-07-16
End:	2020-11-10	2020-11-10
Duration:	110 days	118 days
Third wave		
Mean value(s)	1.10	1.32
Start	2021-02-26	2021-03-04
End:	2021-05-03	2021-04-20
Duration:	67 days	48 days +
India		
First wave		
Mean value(s)	1.82/1.29/1.22	1.81/1.26
Start	2020-03-15	2020-03-15
End:	2020-08-02	2020-07-29
Duration:	141 days +	137 days +
Second wave		
Mean value(s)	1.35	1.39
Start	2021-03-08	2021-03-09
End:	2021-05-06	2021-04-29
Duration:	60 days	52 days
Israel		
First wave		

(*continued on next page*)

Table 2A (*continued*)

Country	Break LS: original data	Break LS: adjusted data
Mean value(s)	1.53	1.64/1.35
Start	2020-03-14	2020-03-14
End:	2020-04-29	2020-07-18
Duration:	47 days +	127 days +
Second wave		
Mean value(s)	1.35/1.15	1.33
Start	2020-06-16	2020-11-12
End:	2020-10-01	2021-01-13
Duration:	108 days	48 days
Third wave		
Mean value(s)	1.29	1.41
Start	2020-11-18	2021-05-15
End:	2021-01-15	2021-06-30
Duration:	59 days	47 days
Fourth wave		
Mean value(s)	1.27	
Start	2021-05-15	
End:	2021-06-30	
Duration:	47 days +	

4. Results

4.1. Global and regional COVID-19 wave

Declaring global pandemic waves and prescribing policy solutions at the aggregate level is problematic. Fig. 1g presents the BLS and adjusted BLS results for the World and captures a single long wave over the study period. The regional (Fig. 1 and Table 1) and country level results (Fig. 2 and Table 2) show evidence for the greater number of waves. BLS identifies one initial wave of infection in Asia (Fig. 1b) for 67 days over the period 23 January to March 29, 2020, with a 1.81 reproduction rate. Similarly, a single wave is identified for the African continent (Fig. 1a). These overlap over the same period, though their duration varies. The adjusted BLS identifies that the African wave begins slightly later (March 15, 2020), with a mean reproduction rate of 1.52. The length of this first infectious wave is 55 days.

In contrast, two COVID waves are identified for Europe (Fig. 1c). The first wave exhibits a high reproduction rate of 2.11 and is 49 days long, extending from February 24, 2020 to April 12, 2020. The second COVID wave is much longer (113 days), plateaus, and is relatively less severe. It has a lower mean reproduction rate of 1.29 and extends from 18 July to 7 November.

Evidence shows North America underwent two COVID waves (Fig. 1d). The first wave began around March 5, 2020 and ended 48 days later on April 21, 2020. The mean reproduction rate during the first wave was high (2.07). The second wave was initiated on July 7, 2020 with a mean reproduction rate of 1.14. This wave was over 49 days and ended on August 24, 2020.

Oceania and the South American continent witnessed three COVID waves (Fig. 1e and f). According to the adjusted BLS, Oceania's mean reproduction rate during the first, second and third waves are 1.49, 1.30, and 1.17, respectively. Each of these waves is roughly 48 days long. The first wave begins on March 11, 2020 and ends on April 26, 2020. The second and third waves started (ended) on June 16, 2020 (August 1, 2020) and February 20, 2021 (April 7, 2021), respectively. The variation in the global and regional/continent results demonstrates that it is imperative to acknowledge the limitations of aggregate data for identifying pandemic waves.

At the aggregate level (global, regional (continent), or country), much information and variation in the data are lost. Second, taking the mean value of the reproduction rate does not make sense for large countries or countries where the outbreaks are limited to a few localized areas. Cities with better connectivity and high population density profiles are more susceptible to the initial spread of the virus. At the same time, rural and isolated areas catch the infection slowly. Third, the aggregate shocks due to emergent new variants of concern also introduce regional variation in infections [44]. Fourth, data measurement is fraught with reporting errors and lags due to case detection, definitions, testing strategies, reporting practice, and lag times that differ between countries/territories/areas [45]. Fifth, behavioral changes and policy interventions (both non-pharmaceutical and pharmaceutical) impact when and for how long the COVID waves occur.

4.2. COVID waves for selected countries

First discovered in Wuhan, China, community transmission of COVID-19 was likely presented in Europe and the USA by January 2020, with international travel as the main driver and transmitted through events between December 2019 and January 2020 [46]. Fig. 2A, B, and 2C present the evidence from the BLS models for the selected countries. The corresponding breakpoints (start and end dates) of the COVID-19 waves and their length are presented in Table 2A, 2B, and 2C.

Most European countries have experienced three COVID waves till June 30, 2021. The first wave has high mean reproduction rates (ranging from 1.53 to 1.96) and is 48 days long on average. The initial date of spread is slightly early for Italy (February 24, 2020) as compared to France (March 1, 2020), the UK (March 3, 2020), Germany (March 2, 2020), Spain (March 3, 2020), Netherlands (March 7, 2020) and Sweden (March 7, 2020). The second wave has a lower mean reproduction rate but is much longer. While France,

Table 2B
Time and Duration of the COVID-19 waves, as identified by Break Least Square, Adjust Break Least Square, and Reported data for Reproduction Rate for Selected Countries.

Italy		
First wave		
Mean value(s)	1.85	1.88
Start	2020-02-24	2020-02-24
End:	2020-04-12	2020-04-12
Duration:	49 days +	49 days +
Second wave		
Mean value(s)	1.24/1.45	1.23/1.45
Start	2020-07-30	2020-07-22
End:	2020-11-13	2020-11-12
Duration:	107 days	114 days
Japan		
First wave		
Mean value(s)	1.46	1.48
Start	2020-02-22	2020-02-22
End:	2020-04-16	2020-04-13
Duration:	55 days +	52 days +
Second wave		
Mean value(s)	1.44	1.46
Start	2020-06-12	2020-06-20
End:	2020-08-04	2020-10-24
Duration:	54 days	78 days
Netherlands		
First wave		
Mean value(s)	1.67	1.69
Start	2020-03-07	2020-03-07
End:	2020-04-23	2020-04-23
Duration:	48 days +	48 days +
Second wave		
Mean value(s)	1.30/1.36	1.34/1.26
Start	2020-07-11	2020-07-13
End:	2020-10-24	2020-12-23
Duration:	106 days	163 days
Nigeria		
First wave		
Mean value(s)	1.46/1.17	1.46/1.17
Start	2020-03-30	2020-03-30
End:	2020-06-27	2020-06-27
Duration:	90 days +	90 days +
Second wave		
Mean value(s)	1.30	1.30
Start	2020-11-28	2020-11-29
End:	2021-02-25	2021-01-12
Duration:	45 days	45 days
Peru		
First wave		
Mean value(s)	1.67	1.69
Start	2020-03-07	2020-03-07
End:	2020-04-23	2020-04-23
Duration:	48 days +	48 days +
Second wave		
Mean value(s)	1.30/1.36	1.34/1.26
Start	2020-07-11	2020-07-13
End:	2020-10-24	2020-12-23
Duration:	106 days	163 days
Third wave		
Mean value(s)		1.10
Start		2021-02-07
End:		2021-04-05
Duration:		48 days
Singapore		
First wave		
Mean value(s)	1.55	1.59
Start	2020-03-01	2020-03-01
End:	2020-04-25	2020-04-24
Duration:	56 days +	55 days +
Second wave		
Mean value(s)	1.23	1.23

(continued on next page)

Table 2B (*continued*)

Italy		
Start	2020-12-05	2020-12-03
End:	2021-01-21	2021-01-19
Duration:	48 days	48 days
Third wave		
Mean value(s)	1.14	1.14
Start	2021-03-14	2021-03-14
End:	2021-04-30	2021-04-30
Duration:	48 days	48 days

Germany, Italy, and Sweden witnessed a 113–118 days long wave, the duration of the second wave for the Netherlands and the UK is much longer at 163 and 186 days, respectively. For most European countries, the second wave begins around mid-July and ends around mid-November. Netherlands and UK are outliers, with the wave beginning on July 13, 2020 and July 8, 2020 and ending around December 23, 2020 and January 9, 2021, respectively. Spain has a long plateau that extends nearly one year from summer 2020 to 2021 (361 days).

The third wave is about 48 days long, with mean reproduction rates similar to the second wave. The start and end dates for the third wave vary substantially across countries. The third cycle begins (ends) on 4 March 21 (April 20, 2021) for Germany; February 7, 2021 (April 5, 2021) for the Netherlands; December 14, 2020 (January 30, 2021) for Spain; February 26, 2021 (April 19, 2021) for Sweden and May 14, 2021 (June 30, 2020) for the UK. The BLS unadjusted data shows evidence of a third wave for France (48 days, 4 March – April 20, 2021).

The developments of COVID-19 are accompanied with new variants. Table 3 summarizes the waves, if any, associated with new variants in the countries of emergence. The durations are estimated on the basis of our analysis. Note that Alpha and Zeta created relatively longer waves. Iota, on the other hand, does not seem to produce a wave in United States. Our analysis shows that other COVID-19 variants led to waves with durations of around 45–49 days.

5. Discussion

This study is not conclusive but indicative of the methods that may be employed to identify the COVID-19 waves. The method can be applied to longer time series to identify different COVID-19 variant waves, including the Omicron wave and/or the more recent COVID-19 explosion in China. In this study, we show that BLS may effectively detect breaking points (start and end) in the effective reproduction rate to identify COVID-19 pandemic waves and their length. We empirically demonstrate that identifying waves at the country or local level is more relevant for interventions than global or aggregate levels. Furthermore, irrespective of the circumstances, once the COVID wave begins, it takes about 48 days on average to subside. It is possible that this figure may change as new COVID variants emerge. Identifying the correct breaking points is critical to evaluate the impact of various COVID-19 variants and the different pharmaceutical and non-pharmaceutical interventions and to prescribe effective policy solutions.

Similar studies on COVID-19 waves also adopt the effective reproduction number R [30]. The identification of waves depends on R and the minimum length of persistent developments. The average duration of waves is about 61.7 days. In contrast, this paper imposes no such constraint and finds that the shortest duration is about 48 days. However, the results are not comparable as the samples are different. Nevertheless, the numbers of waves are not dramatically different. Another study [31] considers two other proxies, confirmed cases and death. An algorithm is developed to rule out the noise/temporal developments from the observed waves. Our results present the same conclusion regarding Italy and the US (two extreme cases) as [31].

6. Conclusion

The waves are persistent and substantial periods. In comparison to the temporal and short-lived random developments, waves demand more attention and resources of policy responses. Thus, identifying the waves is crucial for policy makers. This paper provides a simple framework for detecting the COVID waves. The information is useful for policy makers to plan specific interventions. Besides these policy implications, our study provides necessary information for evaluating the impacts of the COVID-19 pandemic. In contrast to the low frequency data and its analysis in other studies, our paper employs BLS method with high frequent data on COVID-19. However, BLS can only estimate the break points individually for each country. It is not possible to capture the spatial effects of other countries. Since contagion is an important feature of the pandemic, the analyses with aggregated and individual data might not be sufficient. The statistical methods that can identify spatial effects and common trends need to be considered in our future work. In addition, to identifying the waves, we have to set up a small number of maximum breaks. Future research may consider other statistical methods, such as the generalized autoregressive score model, to model the time varying parameters according to the data distribution.

CRediT authorship contribution statement

Ranjula Bali Swain: Writing – review & editing, Writing – original draft, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Xiang Lin:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Fan Yang Wallentin:** Writing – review & editing, Supervision, Investigation, Conceptualization.

Table 2C
Time and Duration of the COVID-19 waves, as identified by Break Least Square, Adjust Break Least Square, and Reported data for Reproduction Rate for Selected Countries.

South Africa		
First wave		
Mean value(s)	1.48/1.33	1.49/1.35
Start	2020-03-19	2020-03-19
End:	2020-07-19	2020-07-14
Duration:	123 days +	118 days +
Second wave		
Mean value(s)	1.26	1.28
Start	2020-11-10	2020-11-13
End:	2021-01-12	2021-01-09
Duration:	64 days	58 days
Third wave		
Mean value(s)	1.31	1.32
Start	2021-05-03	2021-05-04
End:	2021-06-30	2021-06-30
Duration:	59 days +	58 days +
South Korea		
First wave		
Mean value(s)	1.35	1.60/1.20
Start	2020-02-21	2020-02-21
End:	2020-04-09	2020-05-31
Duration:	49 days +	101 days +
Second wave		
Mean value(s)	1.30	1.35
Start	2020-07-12	2020-07-20
End:	2020-10-18	2020-09-06
Duration	50 days	61 days
Third wave		
Mean value(s)	1.27	1.32
Start	2020-10-19	2020-11-07
End:	2020-12-25	2020-12-25
Duration:	68 days	49 days
Spain		
First wave		
Mean value(s)	1.84	1.88
Start	2020-03-03	2020-03-03
End:	2020-04-19	2020-04-19
Duration:	48 days +	48 days +
Second wave		
Mean value(s)	1.53/1.20	1.53/1.12
Start	2020-07-03	2020-07-05
End:	2020-10-26	2021-06-30
Duration:	116 days	361 days
Third wave		
Mean value(s)	1.31	
Start	2020-12-14	
End:	2021-01-30	
Duration:	48 days	
Sweden		
First wave		
Mean value(s)	1.52/1.10	1.53
Start	2020-03-07	2020-03-07
End:	2020-04-23	2020-04-23
Duration:	48 days +	48 days +
Second wave		
Mean value(s)	1.57/1.30	1.57/1.30
Start	2020-09-16	2020-09-15
End:	2021-01-08	2021-01-08
Duration:	115 days	116 days
Third wave		
Mean value(s)	1.15	1.20
Start	2021-02-26	2021-02-26
End:	2021-05-13	2021-04-19
Duration:	77 days	53 days
UK		
First wave		
Mean value(s)	1.88	1.88
Start	2020-03-03	2020-03-03

(*continued on next page*)

Table 2C (*continued*)

South Africa		
End:	2020-04-19	2020-04-19
Duration:	48 days +	48 days +
Second wave		
Mean value(s)	1.11/1.39	1.11/1.39/1.14
Start	2020-08-29	2020-07-08
End:	2020-12-02	2021-01-09
Duration:	96 days	186 days
Third wave		
Mean value(s)	1.16	1.32
Start	2021-01-20	2021-05-14
End:	2021-03-08	2021-06-30
Duration:	48 days	48 days +
Fourth wave		
Mean value(s)	1.33	
Start	2021-05-14	
End:	2021-06-30	
Duration:	48 days +	
USA		
First wave		
Mean value(s)	2.28	1.86/1.11
Start	2020-03-05	2020-03-05
End:	2020-04-21	2020-07-26
Duration:	48 days +	144 days +
Second wave		
Mean value(s)	1.17	1.15
Start	2020-06-09	2020-10-07
End:	2020-07-26	2020-12-17
Duration:	48 days	72 days
Third wave		
Mean value(s)	1.19	
Start	2020-10-06	
End:	2020-11-22	
Duration:	48 days	

Table 3
The duration of waves associated with new variants.

Variant	Country of Emergence	Date of Emergence	Average Durations of Wave Subsidence (days)
Alpha	United Kingdom	September 2020	96 days
Beta	South Africa	May 2020	The wave started before the emergence of Beta. Ended on July 19, 2020. Duration: 123 days
Gamma	Brazil	November 2020	47 days
Delta	India	October 2020	No wave is identified
Epsilon	United States	May 2020	The corresponding wave started on 9 June 2020. Duration: 48 days
Zeta	Brazil	April 2020	The wave started on March 14, 2020. Duration: 94 days
Eta	Multiple countries	December 2020	Nigeria: 45 days Singapore: 48 days Spain: 48 days
Iota	United States	November 2020	No wave is identified

Source [47].

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Bali Swain received vice-chancellors grant from Södertörn University for research on COVID-19 and Sustainability. No conflict of interest to declare. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e25090.

References

[1] T. Malthus, An Essay on the Principle of Population as it Affects the Future Improvement of Society, with Remarks on the Speculations of Mr. Goodwin, M. Condorcet and Other Writers (1 ed.) London (1798).
[2] Y. Harari, Homo Deus: A Brief History of Tomorrow, Random House, 2016.
[3] E. Boserup, The Economics of Agrarian Change under Population Pressure, George Allen and Unwin, London, 1965.
[4] C. Kenny. The plague cycle, the unending war between humanity and infectious disease, Scribner, New York, 2021.
[5] W.A. Bernstein, Splendid Exchange How Trade Shaped the World, Grove/Atlantic, Inc., New York, 2009.
[6] A. Ayala, P. Villalobos, F. Dintrans, Elorrieta, C. Castillo, C. Vargas, M. Maddaleno, Identification of COVID-19 waves: considerations for research and policy, Int J Environ Res Public Health 18 (21) (2021 Oct 21) 11058. https://pubmed.ncbi.nlm.nih.gov/34769577/.
[7] S. Zhang, M. Arroyo, R. Gao, S. Wang, A second wave? What do people mean by COVID waves? – a working definition of epidemic waves. Risk management health care, Policy 14 (2021) 3775–3782. https://www.tandfonline.com/doi/full/10.2147/RMHP.S326051.
[8] M. Coccia, The impact of first and second wave of the COVID-19 pandemic in society: comparative analysis to support control measures to cope with negative effects of future infectious diseases, Environ. Res. 197 (2021) 111099. https://pubmed.ncbi.nlm.nih.gov/33819476/.
[9] K. Agyapon-Ntra, P.E. McSharry, A global analysis of the effectiveness of policy responses to COVID-19, Sci. Rep. 13 (2023) 5629, https://doi.org/10.1038/s41598-023-31709-2.
[10] A.M.S. Fakir, T. Bharati, Pandemic catch-22: the role of mobility restrictions and institutional inequalities in halting the spread of COVID-19, PLoS One 16 (6) (2021) e0253348, https://doi.org/10.1371/journal.pone.0253348.
[11] T. Hale, N. Angrist, R. Goldszmidt, B. Kira, A. Petherick, T. Phillips, S. Webster, E. Cameron-Blake, Laura Hallas, S. Majumdar, H. Tatlow, A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker), Nat. Human Behav. 5 (2021) 529–538, https://doi.org/10.1038/s41562-021-01079-8.
[12] C. Potter, Chronicle of influenza pandemics, in: K.G. Nicholson, R.G. Webster, A.J. Hay (Eds.), Textbook of Influenza, Blackwell Scientific Publications, London, 1998, pp. 3–18.
[13] T. Hale, N. Angrist, R. Goldszmidt, B. Kira, A. Petherick, T. hillips, S. Webster, E. Cameron-Blake, L. Hallas, S. Majumdar, H. Tatlow, A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker), Nat. Human Behav. 5 (2021) 529–538. https://www.nature.com/articles/s41562-021-01079-8.
[14] C.S.S.E. Johns Hopkins, 2019-nCoV Data Repository by, Johns Hopkins CSSE, 2019.
[15] J. Taubenberger, D. Morens, 1918 Influenza: the mother of all pandemics, Emerg. Infect. Dis. 12 (2006) 15–22. https://wwwnc.cdc.gov/eid/article/12/1/05-0979_article.
[16] G. Cacciapaglia, C. Cot, F. Sannino, Multiwave pandemic dynamics explained: how to tame the next wave of infectious diseases, Sci. Rep. 11 (2021) 6638. https://www.nature.com/articles/s41598-021-85875-2.
[17] E. Cameron-Blake, H. Tatlow, K. Green, M. Di Falco, T. Hale, T. Phillips, A. Sudarmawan, H. Zha, What have we learned from tracking every government policy on COVID-19 for the past two years? Blavatnik School of Government, University of Oxford. https://www.bsg.ox.ac.uk/research/publications/what-have-we-learned-tracking-every-government-policy-covid-19-past-two-years. (Accessed September 1 2022).
[18] G. Chowell, S. Dahal, A. Tariq, K. Roosa, J. HymanM, R. Luo, An ensemble n-sub-epidemic modeling framework for short-term forecasting epidemic trajectories: application to the COVID-19 pandemic in the USA, PLOS Comput. Biol. (2022). https://doi.org/10.1371/journal.pcbi.1010602.
[19] L. Berger, N. Berger, V. Bosetti, I. Gilboa, L.P. Hansen, C. Jarvis, M. Marinacci, R.D. Smith, Rational policymaking during a pandemic, Proc. Natl. Acad. Sci. USA 118 (2021). https://www.pnas.org/doi/10.1073/pnas.2012704118.
[20] M. Bruckner, R. Mollerus, UN/DESA Policy Brief 66: COVID-19 and the least developed countries, in: Development Policy Branch in the Economic Analysis and Policy Division of UN DESA, 2020.
[21] W. Kermack, A. McKendrick, G. Walker, Contribution to the mathematical theory of epidemics, Proceeding of the Royal Society A: Mathematics, Physical and Engineering Sciences 115 (1927) 700–721. https://royalsocietypublishing.org/doi/10.1098/rspa.1927.0118.
[22] Z. Wang, M. Andrews, Z. Wu, L. Wang, C.T. Bauch, Coupled disease–behavior dynamics on complex networks: a review, Phys. Life Rev. 15 (2015) 1–29. https://www.sciencedirect.com/science/article/pii/S1571064515001372.
[23] M. Perc, J.J. Jordan, D.G. Rand, Z. Wang, S. Boccaletti, A. Szolnoki, Statistical physics of human cooperation, Phys. Rep. 687 (2017) 1–51. https://www.sciencedirect.com/science/article/abs/pii/S0370157317301424.
[24] X. Zhan, C. Liu, G. Zhao, Z. Zhang, G. Sun, J.J.H. Zhu, Z. Jin, Coupling dynamics of epidemic spreading and information diffusion on complex networks, Appl. Math. Comput. 332 (2018) 437–448. https://www.sciencedirect.com/science/article/pii/S0096300318302236.
[25] F. Arroyo-Marioli, F. Bullano, S. Kucinskas, Tracking R of COVID-19: a new real-time estimation using the Kalman filter, PLoS One 16 (2021) e0244474. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0244474.
[26] A. Kucharski, T. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, Early dynamics of transmission and control of COVID-19: a mathematical modelling study, Lancet Infect. Dis. 20 (2020) 553–558. https://www.thelancet.com/article/S1473-3099(20)30144-4/fulltext.
[27] A. Atkeson, What will Be the economic impact of COVID-19 in the US? Rough estimates of disease scenarios, Available at: https://www.nber.org/papers/w26867, 2020.
[28] G. Leung, Opinion | Lockdown Can't Last Forever. Here's How to Lift it, The New York Times, 2020. https://www.nytimes.com/2020/04/06/opinion/coronavirus-end-social-distancing.html.
[29] E. Budish, Maximize utility subject to R ≤ 1: a simple price-theory approach to COVID-19 lockdown and reopening policy, Available at: https://papers.ssrn.com/abstract=3567068, , 2020.
[30] S.X. Zhange, F.A. Marioli, R. Cao, S. Wang, A second wave? What do people mean by COVID waves? – a working definition of epidemic waves, Risk Manag. Healthc. Pol. (2021) 3775–3782.
[31] J. Harvey, B. Chan, T. Srivastava, A.E. Zarebski, P. Dlotko, P. Blaszczyk, R.H. Parkinson, L.J. White, R. Aguas, A. Mahdi, Epidemiological waves - types, drivers and modulators in the COVID-19 pandemic, Heliyon 9 (2023).
[32] J. Bai, P. Perron, Estimating and testing linear models with multiple structural changes, Econometrica 66 (1998) 47–78. https://www.jstor.org/stable/2998540?origin=crossref.
[33] J. Bai, P. Perron, Critical values for multiple structural change tests, Econom. J. 6 (2003) 72–78. https://www.jstor.org/stable/23113649.
[34] S. Antoshin, A. Berg, M. Souto, Testing for structural breaks in small samples, IMF Working Paper WP/08/75, https://www.imf.org/en/Publications/WP/Issues/2016/12/31/Testing-for-Structural-Breaks-in-Small-Samples-21808, 2008.
[35] R.E. Quandt, Tests of the hypothesis that a linear regression obeys two separate regimes, J. Am. Stat. Assoc. 55 (1960) 324–330. https://www.tandfonline.com/doi/abs/10.1080/01621459.1960.10482067.
[36] D.W.K. Andrews, Tests for parameter instability and structural change with unknown change point, Econometrica 61 (1993) 821–856. https://www.jstor.org/stable/2951764.
[37] J. Bai, P. Perron, Computation and analysis of multiple structural change models, J. Appl. Econom. 6 (2003) 72–78. https://onlinelibrary.wiley.com/doi/full/10.1002/jae.659.
[38] J.D. Hamilton, A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica 57 (1989) 357–384. https://www.jstor.org/stable/1912559.
[39] J.D. Hamilton, Regime-switching models. The New Palgrave Dictionary of Economics, second ed., 2005.
[40] W. Taylor, Change-point Analyzer 2.0 Shareware Program, Taylor Enterprises, Libertyville, Illinois, 2000.
[41] W. Taylor, A Pattern Test for Distinguishing between Autoregressive and Mean-Shift Data, Taylor Enterprises, Libertyville, Illinois, 2000.

[42] D.A. Dickey, W.A. Fuller, Distribution of the estimators for autoregressive time series with a unit root, J. Am. Stat. Assoc. 74 (1979) 427–431. https://www.tandfonline.com/doi/abs/10.1080/01621459.1979.10482531.
[43] P. Perron, The great crash, the oil price shock and the unit root hypothesis, Econometrica 57 (1989) 1361–1401. https://www.jstor.org/stable/1913712.
[44] WHO, Tracking SARS-CoV-2 variants, Available at: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/, 2022. (Accessed 13 January 2022).
[45] WHO. Coronavirus disease (COVID-2019) situation reports. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports (2020) (accessed January 13 2022).
[46] J.T. Davis, M. Chinazzi, N. Perra, K. Mu, A.P. Piontti, M. Ajelli, N.E. Dean, C. Gioannini, M. Litvinova, S. Merler, L. Rossi, K. Sun, X. Xiong, I.M. Longini Jr., M. E. Halloran, C. Vicound, A. Vespignani, Cryptic transmission of SARS-CoV-2 and the first COVID-19 wave, Nature 600 (2021) 127–132. https://www.nature.com/articles/s41586-021-04130-w.
[47] A. Aleem, A.B. Akbar Samad, S. Vaqar, Emerging variants of SARS-CoV-2 and novel therapeutics against coronavirus (COVID-19), in: StatPearls [Internet]. Treasure Island (FL), StatPearls Publishing, 2023 May 8, 2023 Jan–. PMID: 34033342.