Supplemental Materials

Early detection of change patterns in COVID-19 incidence and the implementation of public health policies: a multi-national study

Steven S. Coughlin[a], PhD, Ayten Yiğiter[b], PhD, Hongyan Xu[c], PhD, Adam E. Berman[d,e]*, MD MSc MPH, Jie Chen[c]**, PhD

[a] Division of Epidemiology, Department of Population Health Sciences, Medical College of Georgia, Augusta University, Augusta, Georgia, USA
[b] Department of Statistics, Faculty of Science, Hacettepe University, Beytepe – Ankara, Turkey
[c] Division of Biostatistics and Data Science, Department of Population Health Sciences, Medical College of Georgia, Augusta University, Augusta, Georgia, USA
[d] Division of Health Economics and Modeling, Department of Population Health Sciences, Medical College of Georgia, Augusta University, Augusta, Georgia, USA
[e] Division of Cardiology, Departments of Medicine and Pediatrics. Medical College of Georgia, Augusta University, Augusta, GA. USA.

* Corresponding author, Division of Health Economics and Modeling, Department of Population Health Sciences, Medical College of Georgia, Augusta University, Augusta, Georgia, USA

** Corresponding author, Division of Biostatistics and Data Science, Department of Population Health Sciences, Medical College of Georgia, Augusta University, Augusta, Georgia, USA
E-mail: jiechen@augusta.edu (J. Chen)

More details on Methods: A main mathematical model used in studying COVID-19 data is the traditional mathematical infectious disease model known as the "susceptible (S), infected (I), and recovered (R)" model, or SIR model[1] and its variations, such as stochastic SIR models, for the purpose of determining the total numbered of infected, the reproductive number, R_0, etc. SIR models are powerful and used for implied purposes. We propose an innovative analytic approach by applying statistical change point models, which have been previously employed to model volatility in stock markets, to detect changes in genomic data and data dynamics in other scientific disciplines, to segment the transformed case data. This allowed us to identify possible change or turning points as indicated by the dynamics of daily COVID-19 incidences. Specifically, our analytic approach consisted of two parts. The first was to use a B-spline model[2] to estimate the trend of the original cases (say x_i cases at day i, $i = 1, \dots, n$, where n is the number of days

available in the particular country's data). The spline regression models are traditional nonparametric methods used in trend fitting of data. The B-splines are constructed from polynomial pieces joined by some points called knots (indicated by blue vertical lines in panel A of all figures). These knots are determined by a model selection criterion so that piecewise local polynomials between the knots can together give the best prediction of the overall trend of the incidences. These knots, however, are not necessarily the statistical change points of the trend.

With the daily case trend fitted by the splines, we hope to find where the potential changes are in the trend for each country. This question can be answered by using a statistical change point model,[3] which is the second part of our analytics. The daily incidences in many countries have shown different moving average with varying dispersion (volatility) over different period, indicating piecewise changing mean and variance. Therefore, a statistical change point model is a natural choice for identifying change points in the mean and variance of the data. The model assumes that the data are approximately normally distributed. To satisfy this assumption, we first transformed the number of cases x_i to a normally distributed number of cases (say y_i with $y_i = \sqrt{x_i + 3/8}$). While the probability distribution of the coronavirus cases (x_i's) can be assumed to be Poisson,[4] the transformed values y_i's are approximately normal.[5] The significant change points are detected, where the significance is determined if the p-value is small. The p-value is calculated according to:[6]

$$p - value = 1 - \exp\{-2\exp[2\ln(\ln(n)) + \ln(\ln(\ln(n))) - (2\ln(\ln(n)\lambda_n)^{1/2}]\},$$

where

$$\lambda_n = 2\ln(n) - \min_{2\le k\le n-2}\{k\ln[\textstyle\sum_{i=1}^{k}(y_i - \bar{y}_k)^2] + (n-k)\ln[\sum_{i=k+1}^{n}(y_i - \bar{y}_{n-k})^2] -$$

$$n\ln[\textstyle\sum_{i=1}^{n}(y_i - \bar{y})^2] + (2+n)\ln(n) - k\ln n - (n-k)\ln(n-k)\},$$

$\bar{y}_k = \sum_{i=1}^{k} y_i/k$, $\bar{y}_{n-k} = \sum_{i=k+1}^{n} y_i/(n-k)$, and $\bar{y} = \sum_{i=1}^{n} y_i/n$, for the change point position *k* in the range between 1 and n.

For situations when the mean and variance change point is not able to detect any change points due to the reasons such as the transformed data may actually deviate from the normality assumption or the changes in mean and variance were not exactly piecewise, we used a Bayesian online change point algorithm[7] that places a Markov chain assumption on the change point location and use a Bayesian approach to calculate the posterior probability that any day being a change point sequentially. This is an iterative computing process via a Bayesian framework by assuming the mean and variance follow certain prior distributions, respectively. A point (date index) is found to be a change point if its posterior probability of being a change point > 0.50. The posterior probability is calculated according to the algorithm developed.[7] The identified change points combined with the spline-fitted trend can provide interpretation of how the change points may have provided better or sooner mandate dates for the implementation of public restrictive interventions in each country. The proposed analytics were applied to the WHO daily new cases for the period from January 2020 through May 11, 2020. In addition, we computed a 95% a confidence interval (CI) for each country's unknown true average incidence number between the change dates using the updated incidence data from WHO until June 5th, where the 95% CI for the average before and after the change point location k is obtaind by $\left(\left[\bar{y}_k - \frac{1.96 s_k}{\sqrt{k-1}}\right]^2 - \frac{3}{8}, \left[\bar{y}_k + \frac{1.96 s_k}{\sqrt{k-1}}\right]^2 - \frac{3}{8}\right)$ and $\left(\left[\bar{y}_{n-k} - \frac{1.96 s_{n-k}}{\sqrt{n-k-1}}\right]^2 - \frac{3}{8}, \left[\bar{y}_{n-k} + \frac{1.96 s_{n-k}}{\sqrt{n-k-1}}\right]^2 - \frac{3}{8}\right)$, respectively, with $s_k = \sqrt{\sum_{i=1}^{k} (y_i - \bar{y}_k)^2/(k-1)}$, and $s_{n-k} = \sqrt{\sum_{i=k+1}^{n} (y_i - \bar{y}_{n-k})^2/(n-k-1)}$.

All computations are carried out using our own codes written in Matlab (R2019b) and in R software (version 3.6), and SAS JMP Pro 14.0 software.

Note that at the time of this analysis, the WHO data had some negative new cases reported for a few countries; those negative case numbers were replaced by 0.

More Results on other countries:

In the **European region**, Turkey had its first COVID-19 confirmed case on March 11, and in the next few days, the case numbers lingered in single digits, but picked up quickly thereafter. Turkey instituted travel restrictions and other measures (e.g., public education about hygienic measures such as hand washing) to control the spread of the epidemic.[8,9] The first change point occurred on March 26 followed by an increasing trend (Supplemental Figure 5). It was observed that on and after March 26, transportation by air, bus, train etc. were restricted by the Turkey government's policy.[10] The cases continued to increase until April 26 (the second change point in Supplemental Figure 5) where a slow downturn point was observed. The downturn was slow moving and continued to linger, evidenced by the 95% CI of (1182,1604) (Table 2), for Turkey from April 27 to June 5.

Our analysis of the COVID-19 case data of two other countries in **the Americas region**, Canada and Mexico, revealed that Canada saw a gradually flattened curve after three change points observed in March and April (Supplemental Figure 6) and Mexico still saw an increasing trend after May 11th (Supplemental Figure 7). The 95% CI for each of these two countries from the last detected change date (see Table 1) to June 5 shows the similar situation (see Table 2).

In the region of **Africa**, thirteen nations with close ties to China, including Kenya, South Africa, and Nigeria, were identified as high-risk priority zones for proactive surveillance, detection and containing the spread of COVID-19.[11] Kenya introduced mandatory screening at all ports of entry, and established isolation facilities and a rapid response team to handle suspected cases.[11] Analysis of Kenya's incidence data of Kenya demonstrated two spikes (after the first two change points, see Supplemental Figure 8) with the second spike higher than the first one and the cases were lingering around 20 beyond May 11th. South Africa assigned 300 health officials to ports of

entry and began screening all travelers from China. In addition, national and provincial response teams were set up in South Africa.[11] Nigeria showed a worse picture than Kenya and the cases beyond May 11th were still rising (see Supplemental Figure 9). Likewise, the cases in South Africa still demonstrated an upwards trend (see Supplemental Figure 10). The last CI for each of these three countries was wider in the last period from the last detected change date to June 5 than the previous period (see Table 2) indicates the upward trend for each of them and stricter policy made need to be implemented in order to flatten the curve in these countries.

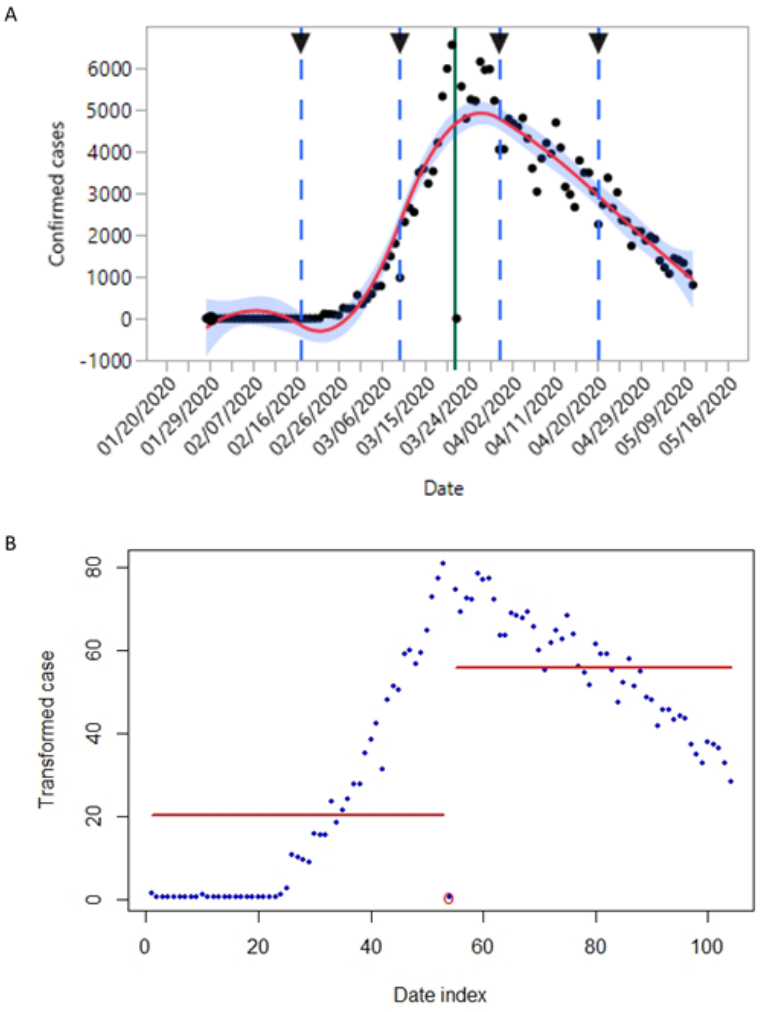
In the **Eastern Mediterranean** region, the cases in Iran peaked between March 2 and April 10, then started a downtrend thereafter but was moving towards a predicted upward direction after April 17 (the third change point, see Supplemental Figure 11). The last CI for Iran from April 18 to June 5 ranges from 1,431 to 1,865 cases, showing an upward trend for Iran (see Table 2). Saudi Arabia instituted travel restrictions and other measures (e.g., public education about hygienic measures such as hand washing) to control the spread of the epidemic.[8,9] The cases in Saudi Arabia had a slow increasing trend until April 16, and quickly picked up an upward trend thereafter (see Supplemental Figure 12). The CI for the average case number from April 17 to June 5 ranges from 1,479 to 1,862 cases (see Table 2) indicating a continuing upward trend.

In the **Western Pacific** (WPRO) region, China saw the earliest breakout in Wuhan. The Chinese government restricted travel from and to Hubei province and implemented a number of measures to contain the outbreak, including quarantine measures affecting 60 million people and strict limitations on travel.[12] The public health interventions included intensive case and contact tracking, isolation of moderately ill patients in containment centers, social distancing, and shutting down public life in the province.[12] The virus was contained after the whole country was on lock-down for roughly two months (see Supplemental Figure 13). The last CI for China showed the estimated average case number from March 25 to June 5 ranged from 18 to 35, indicting a containment. In

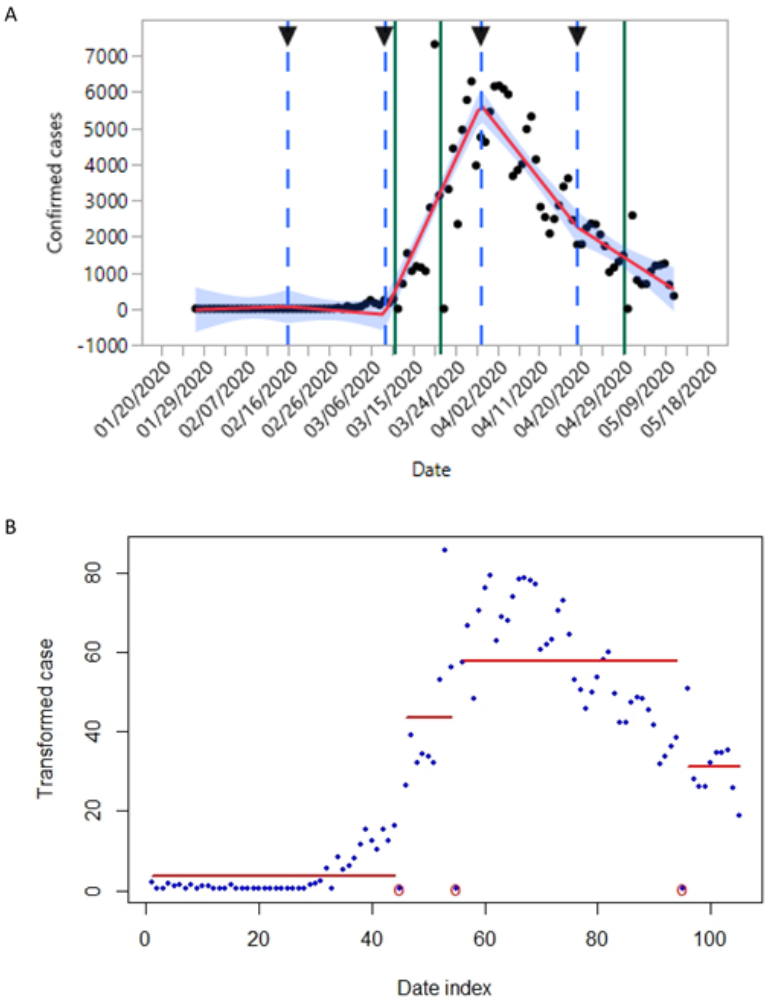
Korea, there were 12 cases of the novel coronavirus disease on February 1, 2020. Korea was able to better control the outbreak through a strategy of mass screening, quarantines in local areas, contact tracing, physical distancing, and a national mobile phone alert system to warn residents of the movements of affected individuals to detect possible outbreaks.[13,14] Our analysis results showed that Japan, Korea, Australia and New Zealand have seen spikes in earlier days of the breakout in their respective countries and are now in containment (see Supplemental Figures 14-17) and the last CI for each of these four countries in Table 2. The cases in Singapore were contained very well in early months from January to mid-March, a gradual increase trend was seen after March 16 (the first green vertical line, see Supplemental Figure 18), the trend progressed until April 13 and beyond April 16 it took a steeper upward trend. The trend is predicted to continue to progress with 3-digit numbers beyond May 11th and the last CI for the period from April 17 to June 5, (591, 707) (see Table 2) confirms this observation.

Supplematal Figures:

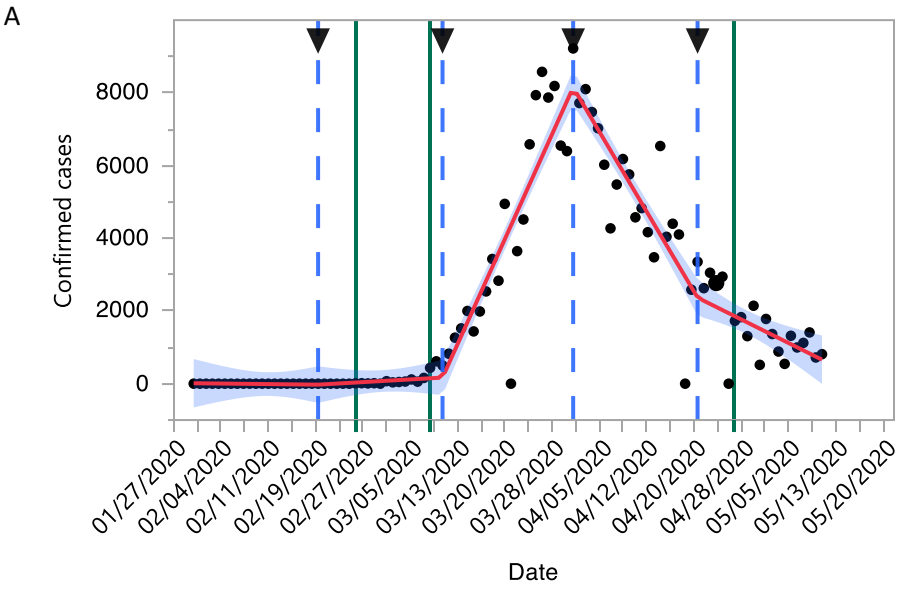
Supplemental Figure 1. **B-spline fitting of the trend of new daily cases for Italy (A) with change points identified according to a change point model (B).** The middle vertical green line is marked at the change point (red cricle in panel B) identified by the online change point detection algorithm with the posterior probability of 0.8145.

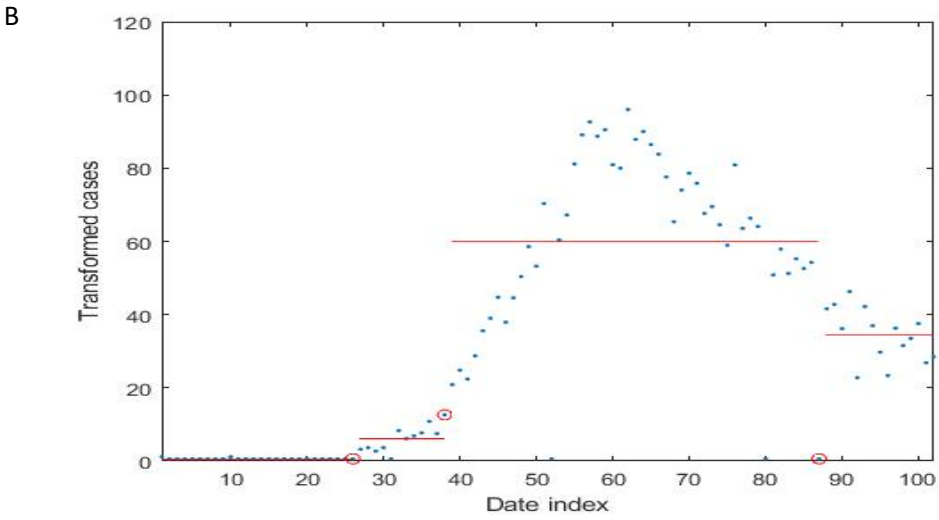


Supplemental Figure 2. **B-spline fitting of the trend of new daily cases for Germany (A) with change points identified according to a change point model (B).** The three vertical green lines are marked at the change points identified by the online change point detection algorithm with posterior probability respectively as 0.6445, 0.7786, and 0.7455.

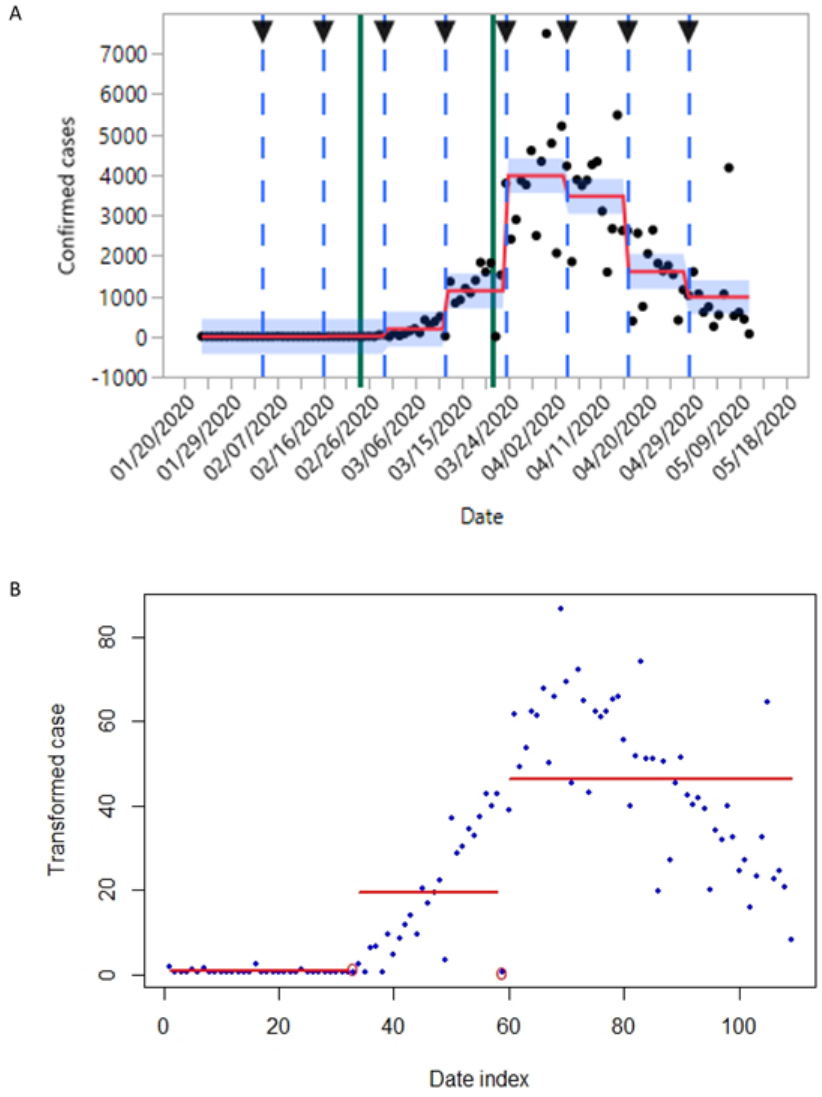


Supplemental Figure 3. **B-spline fitting of the trend of new daily cases for Spain (A) with change points identified according to a change point model (B).** The three vertical green lines are marked at the change points (red cricles in panel B) identified by the mean and variance change point model with p-values respectively as 0.0000, 0.0000, and 0.0021.

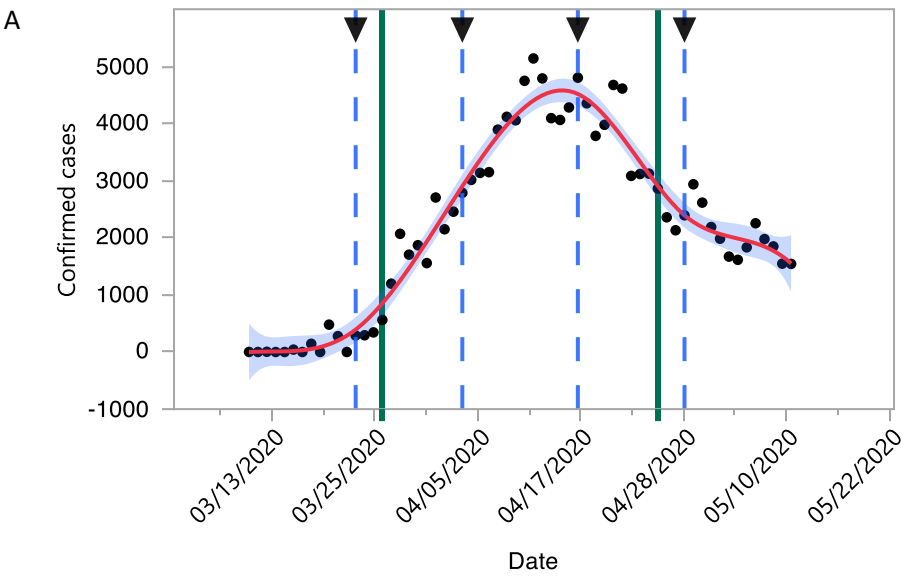


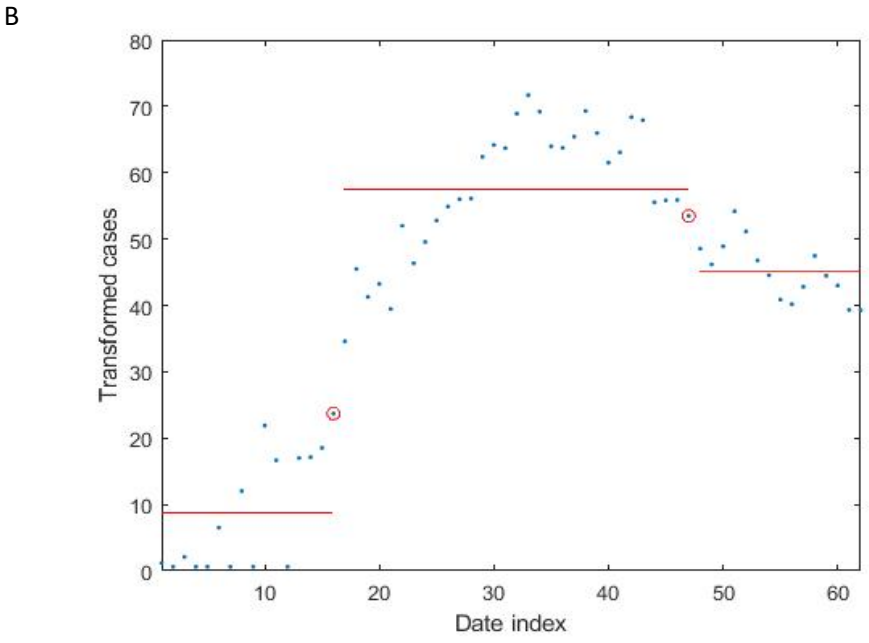


Supplemental Figure 4. **B-spline fitting of the trend of new daily cases for France (A) with change points identified according to a change point model (B).** The two vertical green lines in panel A are marked at the change points (red cricles in panel B) identified by the online change point detection algorithm with posterior probability respectively as 0.5383 and 0.7029.

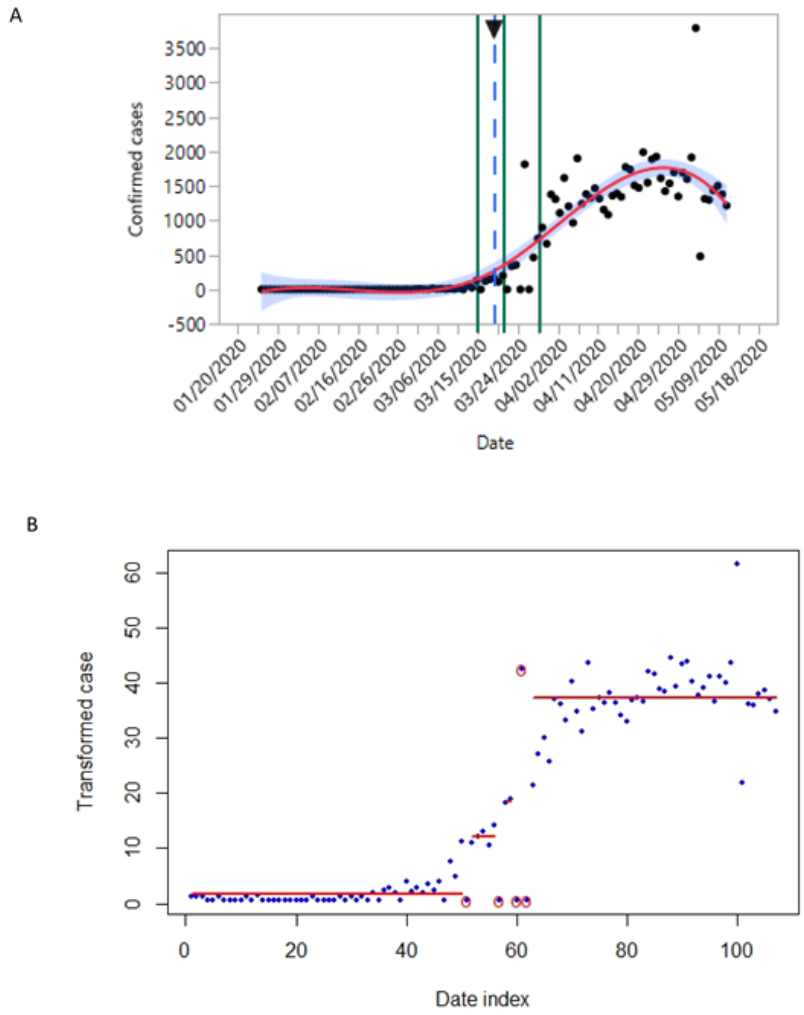


Supplemental Figure 5. **B-spline fitting of the trend of new daily cases for Turkey (A) with change points identified according to a change point model (B).** The two vertical green lines are marked at the change points (red cricles in panel B) identified by the mean and variance change point model with p-values respectively as 0.0000 and 0.0344.

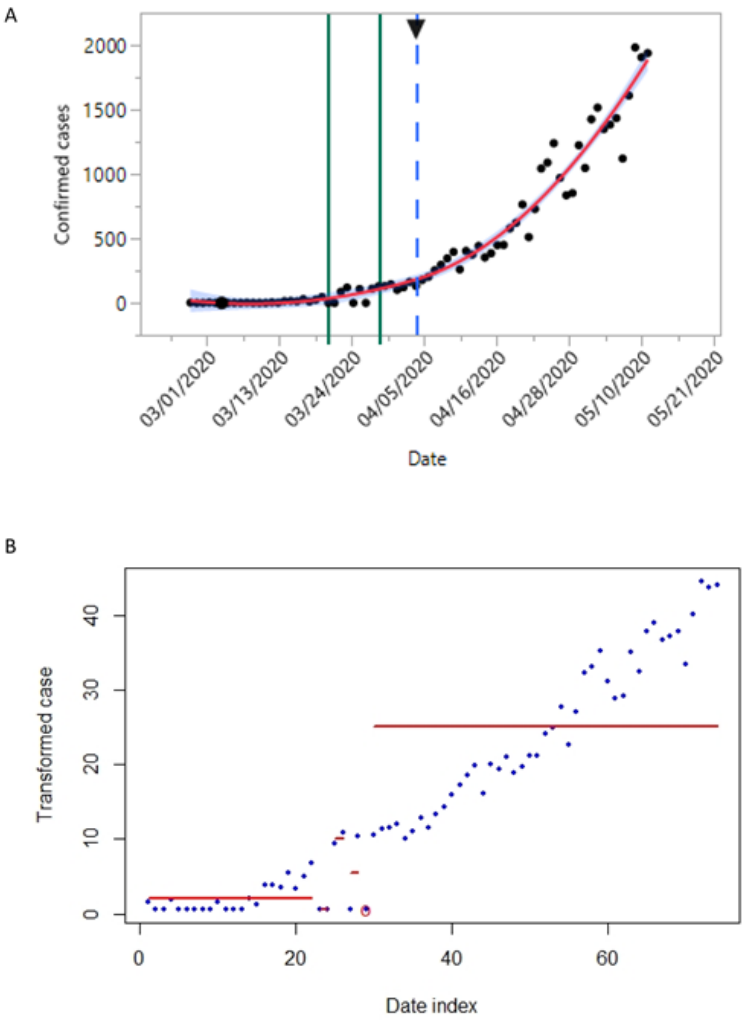




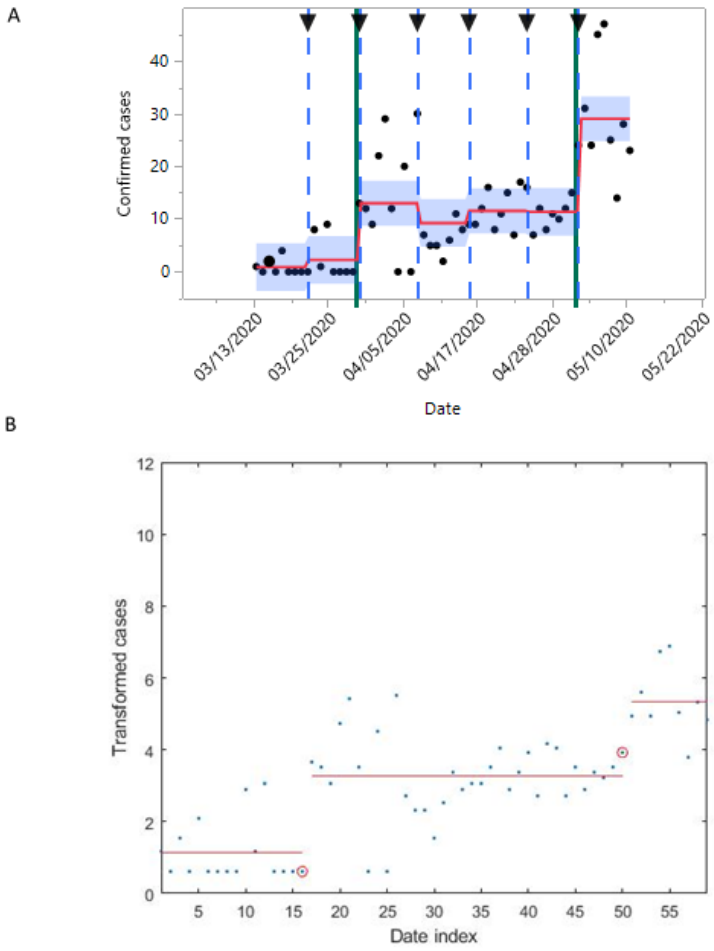
Supplemental Figure 6. **B-spline fitting of the trend of new daily cases for Canada (A) with change points identified according to a change point model (B).** The three vertical green lines are marked at the change points (red cricles in panel B) identified by the online change point detection algorithm with posterior probability respectively as 0.5367, 0.5535, and 0.5751.



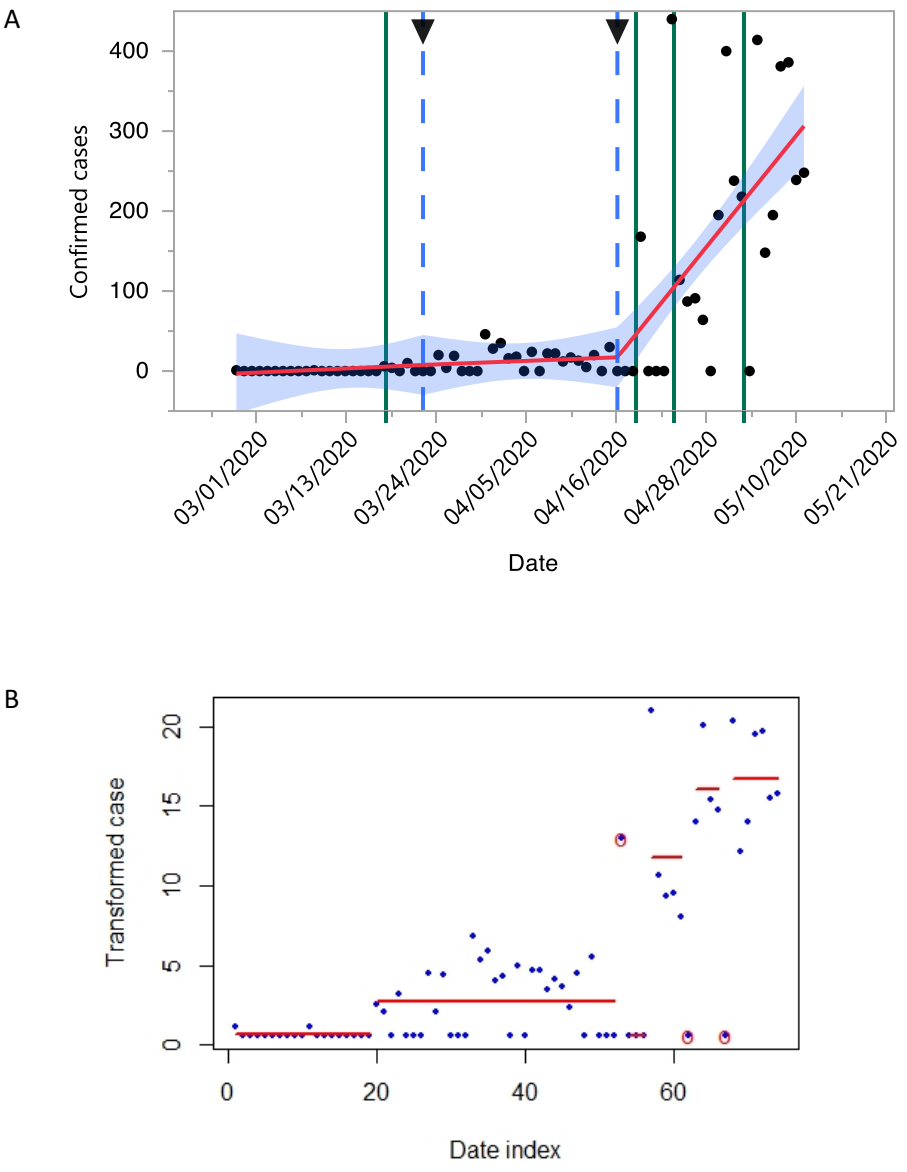
Supplemental Figure 7. **B-spline fitting of the trend of new daily cases for Mexico (A) with change points identified according to a change point model (B).** The two vertical green lines are marked at the change points identified (red cricles in panel B) by the online change point detection algorithm with posterior probability respectively as 0.5698 and 0.5172.



Supplemental Figure 8. **B-spline fitting of the trend of new daily cases for Kenya (A) with change points identified according to a change point model (B).** The first two vertical green lines are marked at the change points (red circles in panel B) identified by the mean and variance change point model with p-values respectively as 0.0009 and 0.0230.

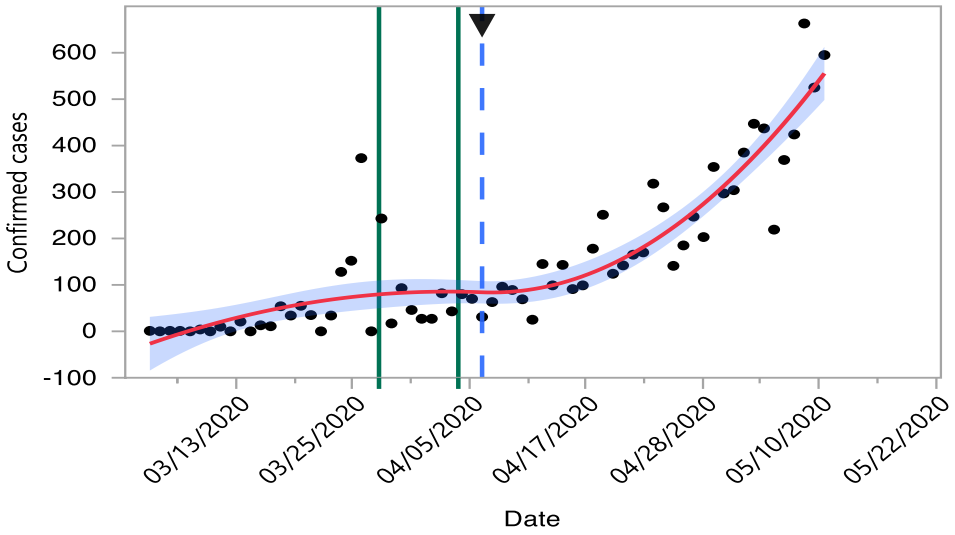


Supplemental Figure 9. **B-spline fitting of the trend of new daily cases for Nigeria (A) with change points identified according to a change point model (B).** The four vertical green lines are the marked at change points (red cricles in panel B) identified by the online change point detection algorithm with posterior probability respectively as 0.6382, 0.6917, 0.8862, and 0.5777.

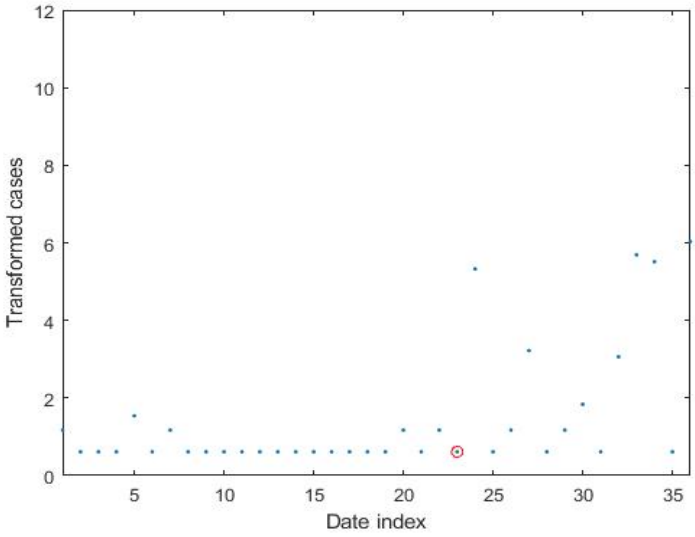


Supplemental Figure 10. **B-spline fitting of the trend of new daily cases for South Africa (A) with change points identified according to a change point model (B).** The two vertical green lines are marked at the change points (red cricles in panel B) identified by the mean and variance change point model with p-values respectively as 0.0000 and 0.0672.

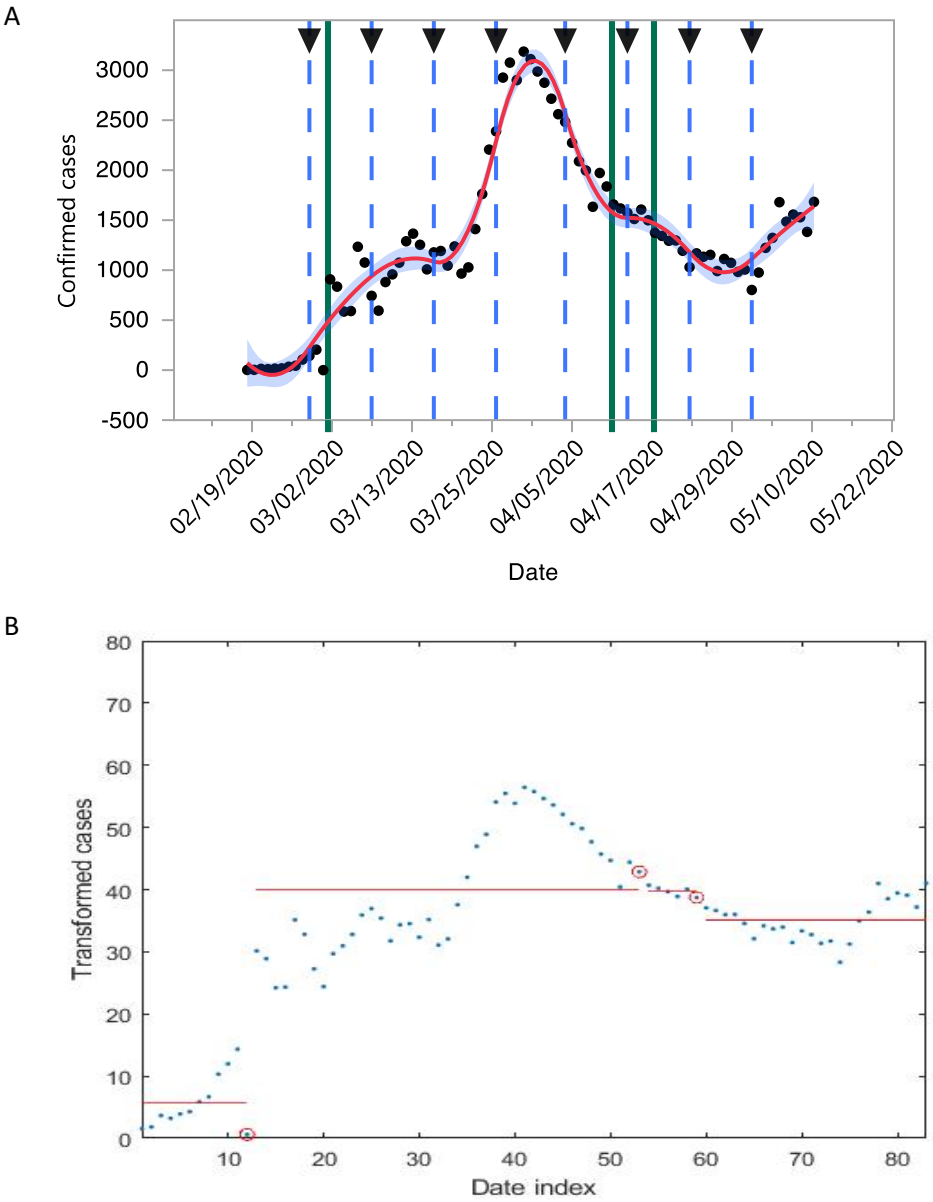
A



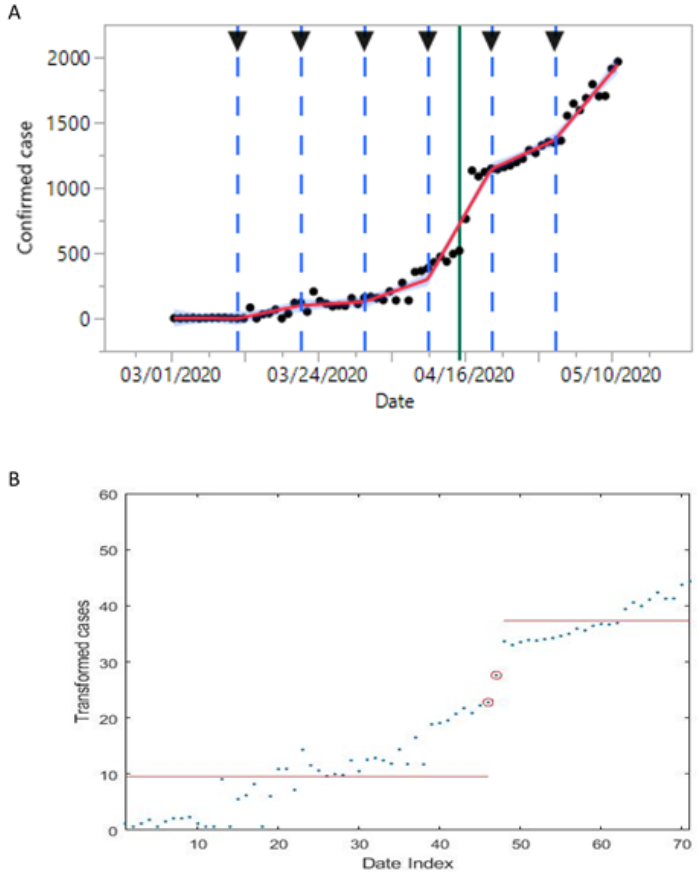
B



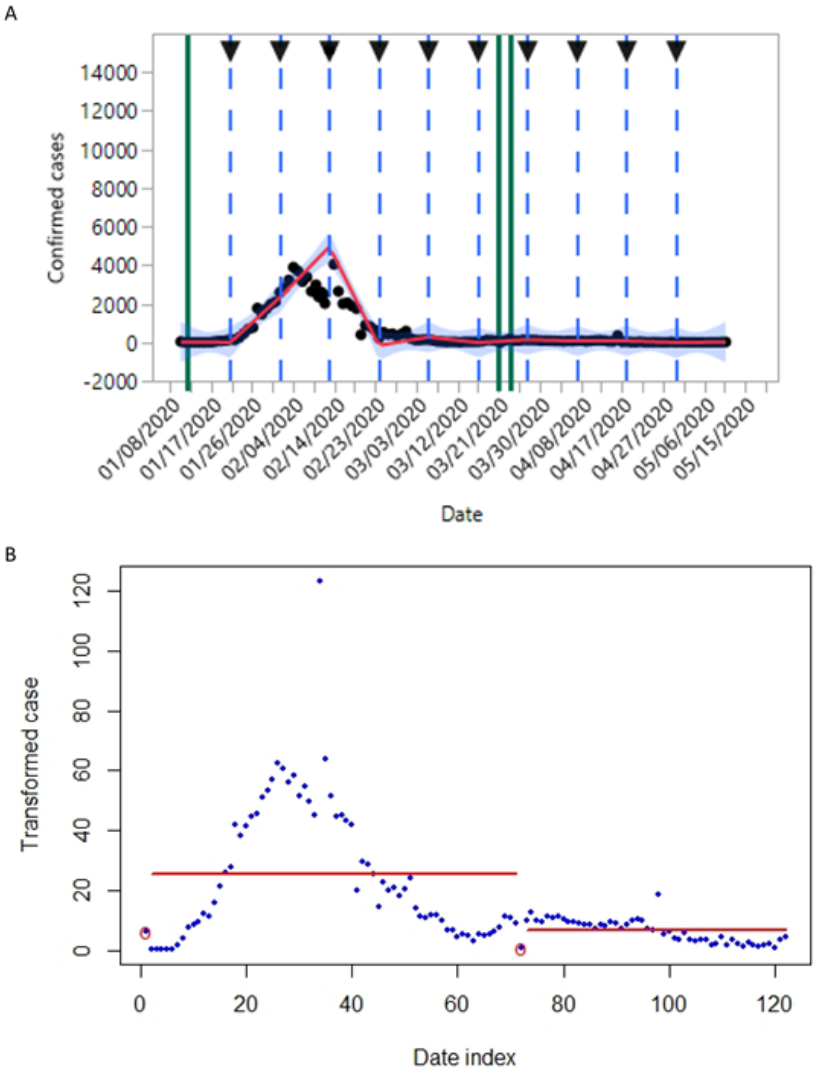
Supplemental Figure 11. **B-spline fitting of the trend of new daily cases for Iran (A) with change points identified according to a change point model (B).** The three vertical green lines in panle A are the change points (red cricles in panel B) identified by the mean and variance change point modell with p-values respectively as 0.0000, 0.0022 and 0.0740.



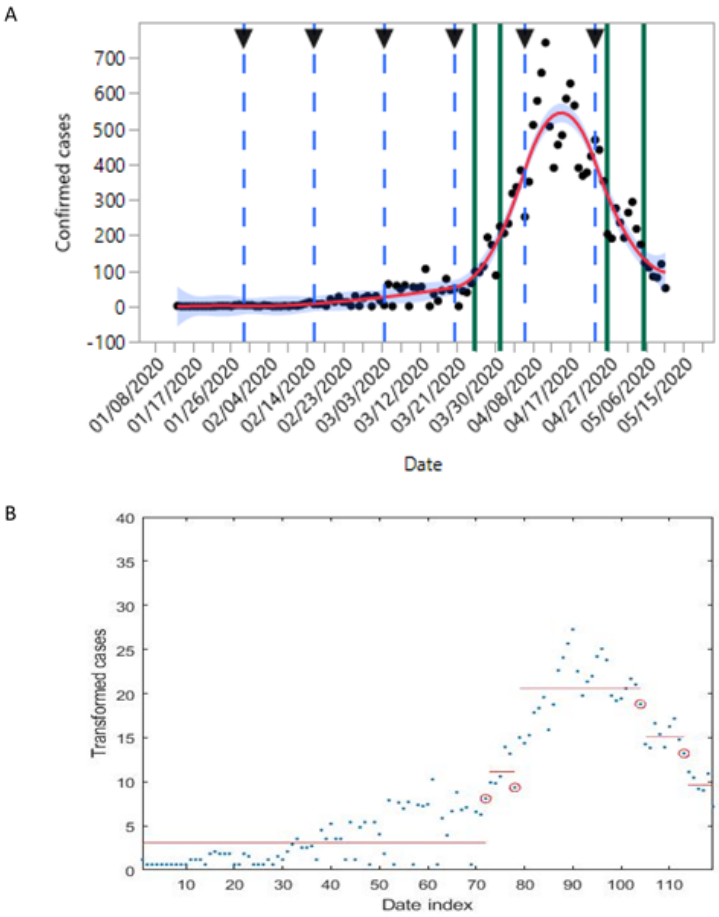
Supplemental Figure 12. **B-spline fitting of the trend of new daily cases for Saudi Arabia (A) with change points identified according to a change point model (B).** The vertical green line is marked at the two adjcent change points (red cicles in panel B) identified by the mean and variance change point model with p-values respectively as 0.0000 and 0.0672.



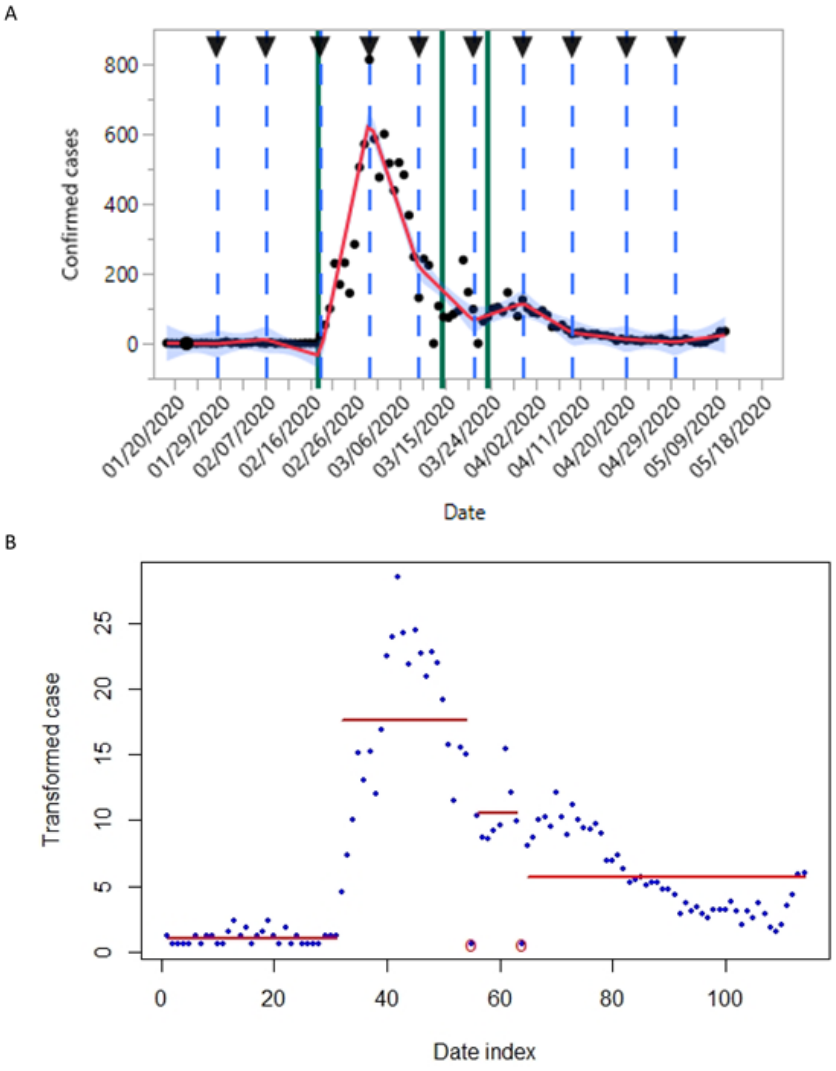
Supplemental Figure 13. **B-spline fitting of the trend of new daily cases for China (A) with change points identified according to a change point model (B).** The three vertical green lines are marked at the change points (red cicles in panel B) identified by the online change point detection algorithm with posterior probability respectively as 0.5555, 0.5535, 0.5031.



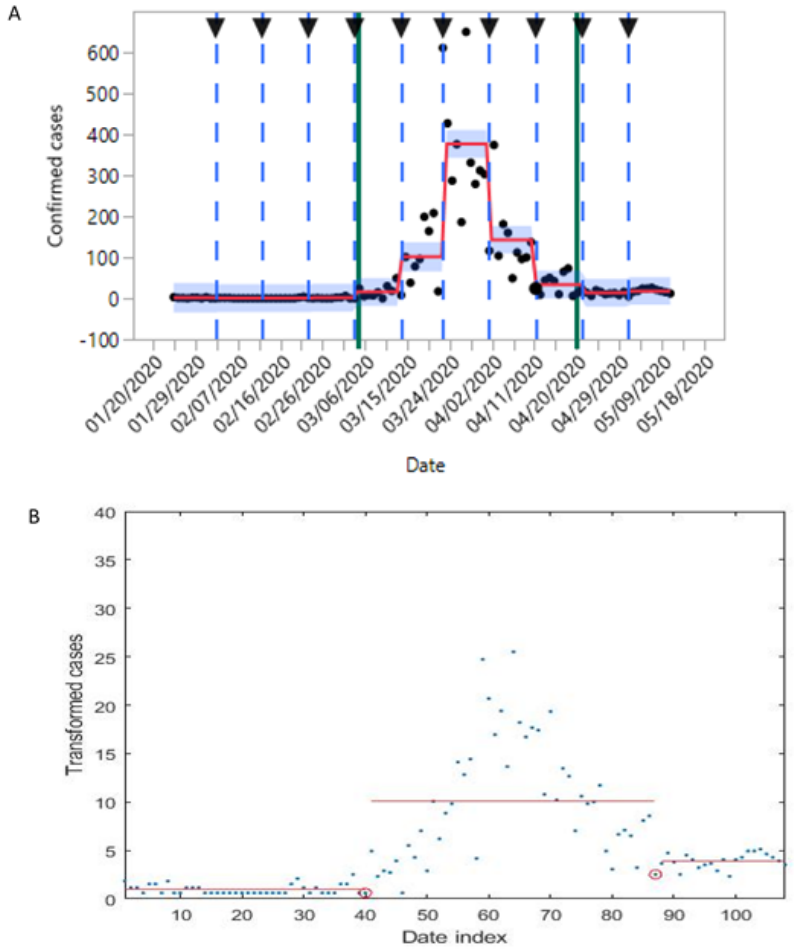
Supplemental Figure 14. **B-spline fitting of the trend of new daily cases for Japan (A) with change points identified according to a change point model (B).** The four vertical green lines are marked at the change points (red cicles in panel B) identified by the mean and variance change point model with p-values respectively as 0.0000, 0.0312, 0.0083, and 0.0548.



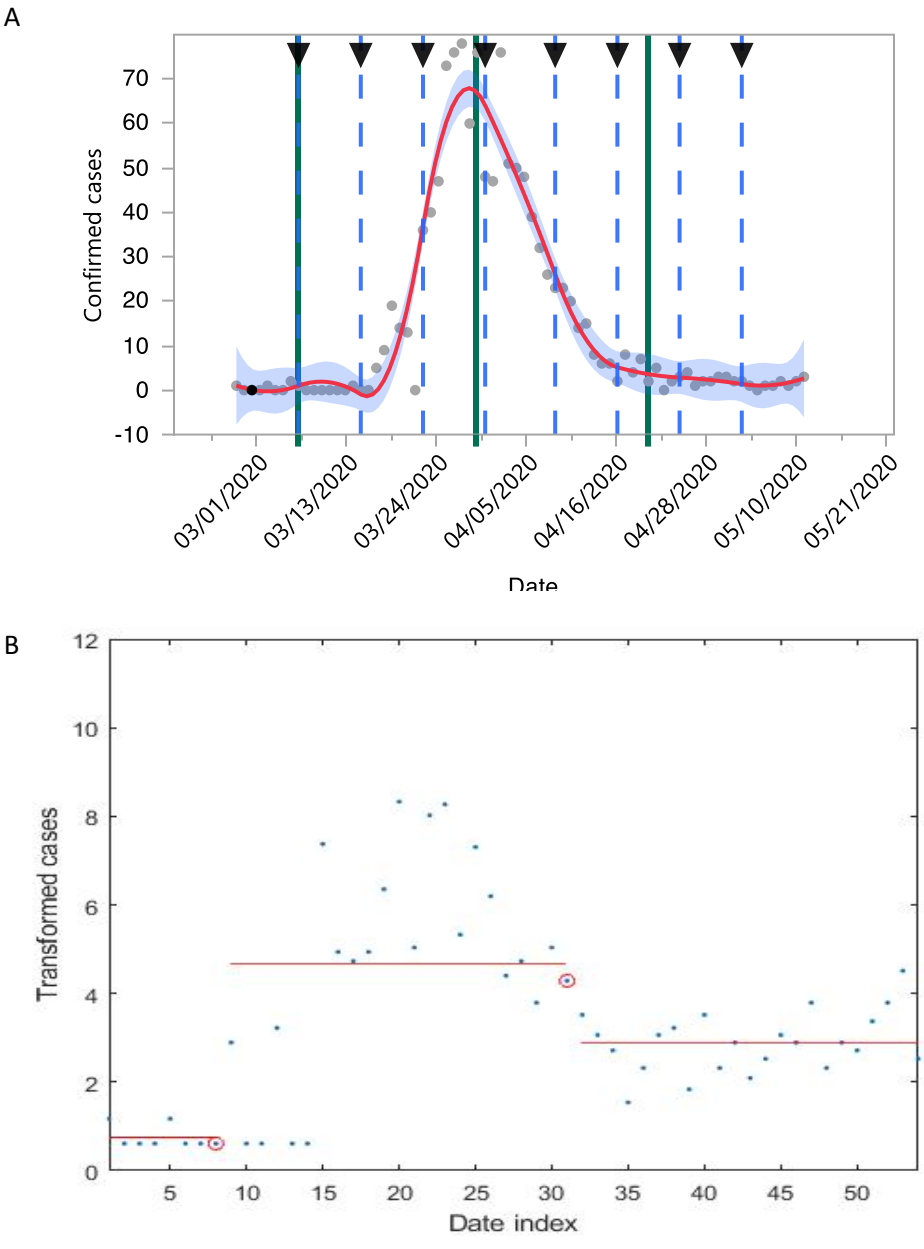
Supplemental Figure 15. **B-spline fitting of the trend of new daily cases for Korea (A) with change points identified according to a change point model (B).** The three vertical green lines are marked at the change points (red cicles in panel B) identified by the online change point detection algorithm with posterior probability respectively as 0.6314, 0.9517, 0.8482.



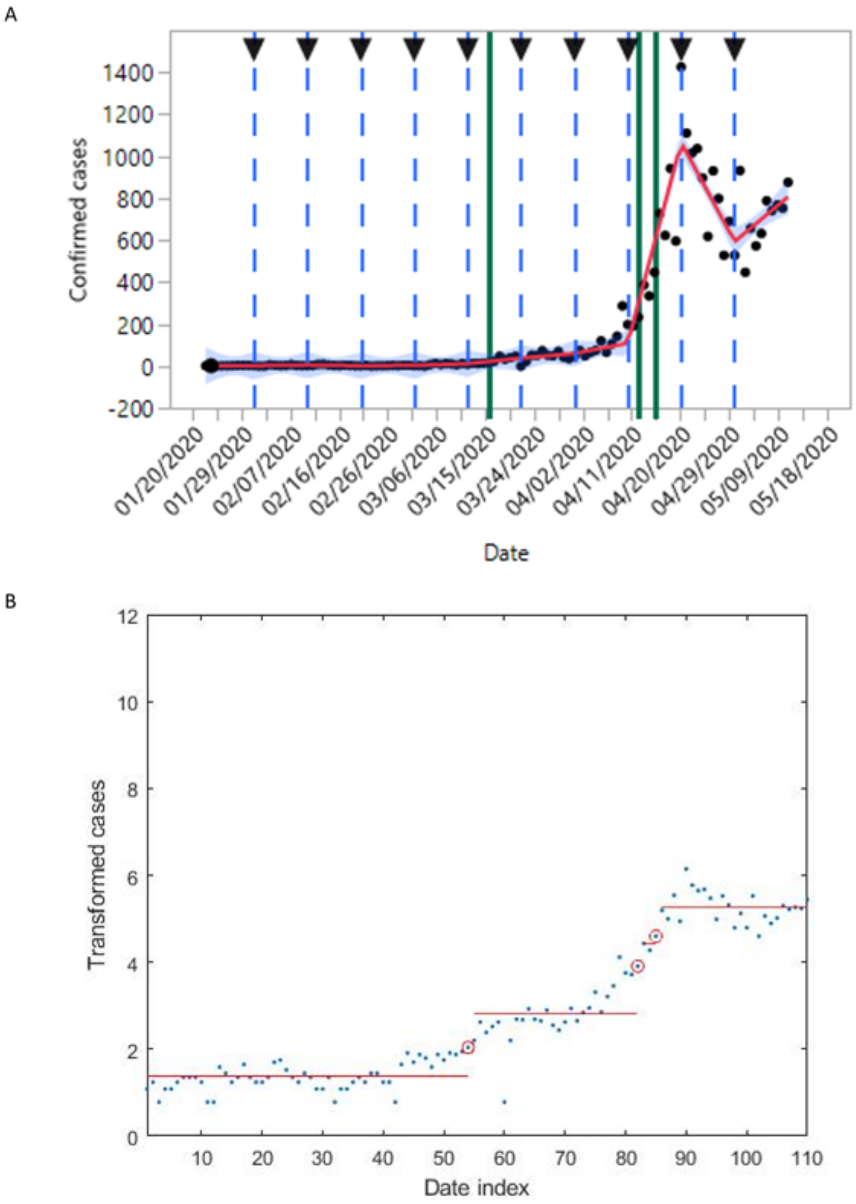
Supplemental Figure 16. **B-spline fitting of the trend of new daily cases for Australia (A) with change points identified according to a change point model (B).** The two vertical green lines are marked at the change points (red cricles in panel B) indetified by the mean and variance change point model with p-values respectively as 0.0000 and 0.0000.



Supplemental Figure 17. **B-spline fitting of the trend of new daily cases for New Zealand (A) with change points identified according to a change point model (B).** The first two vertical green lines in panel A are marked at the change points (red cicles in panel B) identified by the mean and variance change point model with p-values respectively as 0.0000 and 0.0016. The last vertical green line in panel A is marked at a point whose p-value of 0.3966 is not small enough for being considered as a change point according to the algorithm. The possible interpretation of that point is that it may indicates the start of the next downturn phase on April 20 for New Zealand.



Supplemental Figure 18. **B-spline fitting of the trend of new daily cases for Singapore (A) with change points identified according to a change point model (B).** The three vertical green lines in panle A are marked at the change points (red circles in panel B) identified by the mean and variance change point model with p-values respectively as 0.000, 0.0000, and 0.0802.



Reference for Supplemental Materials:

1. Miller JC. Mathematical models of SIR disease spread with combined non-sexual and sexual transmission routes. Infect Dis Model. 2017;2(1):35-55.

2. Eilers PHC, Marx B. Flexible smoothing with B-splines and penalties. Statistical Science. 1996;11 (2): 89–121.

3. Chen J, Gupta AK (2012). “Parametric Statistical Change Point Analysis - With Applications to Genetics, Medicine, and Finance,” second edition, Birkhauser, and copyrighted by Springer Science+Business Media LLC, New York.

4. Pan A, Liu L, Wang C, et al. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China [published online ahead of print, 2020 Apr 10]. JAMA. 2020;323(19):1 ‐ 9. doi:10.1001/jama.2020.6130

5. Ji T, Chen J. Modeling the Next Generation Sequencing Read Count Data for DNA Copy Number Variant Study. Statistical Applications in Genetics and Molecular Biology. 2015;14: 361–374.

6. Chen J, Wang YP. A Statistical Change Point Model Approach for the Detection of DNA Copy Number Variations in Array CGH Data. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2009; 6:529-541.

7. Yiğiter A, Chen J, An L, Danacioğlu N. An online copy number variant detection method for short sequencing reads, Journal of Applied Statistics. 2015; 42(7):1556-1571.

8. Ebrahim SH, Memish ZA. Saudi Arabia’s drastic measure to curb the COVID-19 outbreak: temporary suspension of the Umrah pilgrimage. J of Travel Medicine 2020;1-2.
9. Demirbilek Y, Pehlivanturk G, Ozge Z, et al. COVID-19 outbreak control, example of ministry of health of Turkey. Turkish Journal of Medical Sciences. Turk J Med Sci 2020;50:489-94.

10. Salcedo A, Yar S, Cherelus G. Coronavirus Travel Restrictions, Across the Globe. The New York Times. June 8, 2020. https://www.nytimes.com/article/coronavirus-travel-restrictions.html

11. Kapata N, Ihekweazu C, Ntoumi F, et al. Is Africa prepared for tackling the COVID-19 (SARS-CoV-2) epidemic. Lessons from past outbreaks, ongoing pan-African public health efforts, and implications for the future. Int J Infectious Diseases 2020;93:233-6.
12. Salzberger B, Gluck T, Ehrenstein B. Successful containment of COVID-19: the WHO-report on the COVID-19 outbreak in China. Infection 2020;48:151-3.

13. Moatti J-P. The French response to COVID-19: intrinsic difficulties at the interface of science, public health, and policy. The Lancet 2020;5:e255.

14. Saez M, Tobias A, Varga D, et al. Effectiveness of the measures to flatten the epidemic cure of COVID-19. The case of Spain. Science of the Total Environment 2020;727:138761.