Network Modeling Analysis in Health Informatics and Bioinformatics (2023) 12:33
https://doi.org/10.1007/s13721-023-00430-5

SHORT COMMUNICATION

COVID-19 lag time and case fatality rate calculation tool, as well as a tool to identify when policymakers made mistakes

Yoshiyasu Takefuji[1]

Received: 2 June 2023 / Revised: 20 July 2023 / Accepted: 25 August 2023

Abstract

There is no tool for calculating a case fatality rate (CFR) and a lag time of COVID-19. This paper proposes a new policymaker tool, covidlag for policymakers to calculate the CFR and the lag time with associated relationship from infection to death. The more the infections, the more the COVID-19 deaths. The less the infections, the less the deaths. We took advantage of this correlation between infection and death. In other words, the number of daily cases and that of daily deaths are used for calculating CFRs and lag times in the US. Scoring policies are based on dividing the number of daily cumulative deaths by the population in millions. The proposed covidlag algorithm with a strong correlation from infection to death with maxima and minima can generate lag times and CFRs in ongoing time series. covidlag allows users to provide the appropriate degree of curve-fitting polynomials of infection and death to generate maxima and minima for calculating time lag and CFR. The introduced hiscovid discovered that New Zealand made a single mistake in March 2022 while the US made multiple major mistakes during the pandemic. The hiscovid tool found that the COVID-19 surge in New Zealand is due to lifting of travel border restrictions. The CFR and lag time calculations play a key role for policymakers in analyzing the COVID-19 pandemic.

1 Introduction

This paper proposes new policymaker applications for enhancing decision support system with big data. The proposed applications are intended to monitor the COVID-19 pandemic in cities and to improve and mitigate the pandemic by updating COVID-19 policies. Therefore, the proposed paper will fulfill the aim and scope of the journal.

Pandemic analysis is essential for cities and society. To not only observe the progression of the COVID-19 pandemic, but also to mitigate and end it, two indicators are important for policymakers to determine policy toward COVID-19: a lag time from infection to death and a case fatality rate (CFR) or deaths per infection (Hsiang et al. 2020; Russell 2020). Misjudgment by policymakers can cause unnecessary COVID-19 deaths in cities and society. This paper will focus on CFRs and lag times for COVID-19 pandemic analysis for improving human wellbeing.

Therefore, policymakers must correctly determine the state of the COVID-19 pandemic whether they should strengthen their policies or not (Davies et al. 2021). Both lag time and CFR are important indicators that are constantly changing due to COVID-19 variants. The true CFR is unknown because the exact number of infected people is unknown. The calculated CFR in this paper is the estimated value.

If the lag time from infection to death is short, the infected patients must be treated urgently. If the lag time is long, policymakers must provide adequate hospital facilities for the infected patients.

CFR is computed by dividing the number of death due to COVID-19 by the number of cases or infected individuals for retrospective observational study. In other words, CFR in a retrospective study is the ratio of the number of deaths divided by the number of confirmed cases. CFR plays a key role for policymakers in determining policy toward COVID-19. Note that the infection fatality rate is an only estimate, the proportion of people who die after having the infection overall, but we have no data. This paper focuses on the CFR using ongoing time-series data.

In the conventional methods, the number of sampled days and its range (start and end dates) have to be manually determined to calculate a value of the CFR. Remember that CFR is primarily used for retrospective observational study and not for ongoing time-series data. Besides, the number

✉ Yoshiyasu Takefuji
takefuji@keio.jp

1 Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan

of sampled days and its range can significantly influence the CFR.

Unnikrishnan et al. proposed the CFR algorithm using a delay-adjusted case fatality ratio (Unnikrishnan et al. 2021). However, their algorithm must supply the correct baseline CFR and must assume that the delay is based on the fixed lognormal distribution. However, we do not know what is the correct baseline CFR. The algorithm is used only for retrospective observational studies.

Newall et al. (2020) also used the similar delay-adjusted case fatality ratio based on Nishiura's method (Nishiura et al. 2009). However, their method is also used only for retrospective observational study and not for ongoing time-series data.

Rothman et al. proposed the CFR algorithm based on the first-order linear regression (Rothman et al. 2021). However, the algorithm is not used for ongoing time-series computation but for retrospective observational studies.

As far as we know as of today, the conventional delay-adjusted case fatality algorithms cannot easily handle ongoing time-series data. In other words, the conventional algorithms are all used for retrospective observational studies. This is because determining the number of sampled days and its range (start and end dates) is difficult to manually calculate the CFR in ongoing time-series data. There is no satisfactory algorithm in the world that can automatically or semi-automatically calculate ongoing time-series CFRs and time lags.

The proposed algorithm of covidlag, based on strong correlated transitions from infection to death, can assist researchers in calculating lag times and CFRs using ongoing time-series data. In other words, the more the infections, the more the COVID-19 deaths. The less the infections, the less the deaths. In the proposed algorithm, the correlation between infections and deaths is utilized. In other words, the algorithm uses the maximum and minimum using the transition relationship between the peak of infection and the peak of death. Instead of three determinants such as the number of days and its range with start and end dates, researchers must determine and provide an appropriate degree of the curve-fitting function to generate maxima and minima by observing the generated graph with the proposed tool. In other words, with covidlag, a single determinant such as the curve-fitting degree is only needed while the conventional calculations need three determinants such as the number of days, the range start, and the range end dates. However, if there are no maxima or minima in the infection and death graph, we cannot theoretically compute the CFR and the lag time.

First, the proposed algorithm of covidlag computes two curve-fitting functions using the number of daily cases and the number of daily deaths in the time series, respectively. Two curve-fitting functions from infection to death are usually strongly correlated. No infection means no death. The more people infected, the more deaths.

Second, the two curve-fitting functions are used to generate the extreme values including maxima and minima. In other words, overlearning or overfitting using data means no extreme value. The proposed algorithm allows the user to provide the degree of curve-fitting polynomial where r-squared plays a key role in finding an appropriate degree of the curve-fitting polynomial.

Finally, the lag time can be calculated by the difference between the peak day of infection and the peak day of death, and the CFR can be calculated by dividing the number of deaths at peak by the number of cases at peak, respectively. In this paper, maxima in the extreme values are only discussed. Similarly, minima can be used for computing lag time and CFR. The algorithm was tested by observing many countries included in the scraped csv files. The proposed covidlag tool can run on Windows, MacOS, and Linux operating systems, respectively, as long as Python is installed on the system. The covidlag tool provides a semi-automated and accurate picture of the COVID-19 epidemic that policymakers can use to determine whether current policies should be strengthened or relaxed. The covidlag tool has been downloaded by 19,209 times worldwide as of July 21, 2023. The number of users indicates that the applicability, the usability, and the usefulness were justified.

This single metric corresponds to individual policy outcomes and is validated by four peer-reviewed articles (Takefuji 2021a, b, 2022a, b). The single metric was proposed and introduced in the herd immunity debate in Sweden (Takefuji 2021a). The single metric indicates that the less the number of COVID-19 deaths, the better the policy. Therefore, the single metric was used for revealing the best COVID-19 policy in the world. Scores based on the single metric can be used for poorly scored countries to learn good strategies from countries with excellent scores (Takefuji 2021a, 2021b, 2022a, b).

The tool, hiscovid, can be used for analyzing individual travel policies and identifying mistakes by policymakers (Takefuji 2022d, 2023). The hiscovid tool is based on a single metric with a scoring policy of dividing the number of COVID-19 deaths by the population in millions in time series. In other words, the hiscovid tool can discover and show when policymakers made mistakes. This paper will show the result of New Zealand and the US for comparison of their policies.

2 Methods/calculations

The proposed method for calculating lag time and CFR is based on a strong correlation between infection and death or associated relationship from infection to death. If a strong correlation is not found for a country, the proposed algorithm cannot calculate lag time and CFR. In the proposed algorithm, it is assumed that the higher the number of infected people, the higher the number of deaths. In other words, the proposed method calculates the maxima or minima of the first infection wave and the second mortality wave with a time lag between them. To calculate the maxima and minima of two waves, two curve-fitting functions will be estimated with the linear polynomial regressions. The accuracy of curve-fitting functions can be measured by the r-squared or r2 value, which compares real and estimated values. The proposed covidlag tool allows users to change the order of linear polynomial regressions and observe the r-squared values to find the best estimate. Therefore, the lag time can be calculated by the difference between the peak day of infection and the peak day of mortality, and the CFR can be calculated by dividing the number of deaths at peak by the number of cases at peak, respectively.

First, two curve-fitting functions are generated by ongoing time-series daily infection and daily death data: the number of daily cases and that of daily deaths, respectively. The extreme values including maxima and minima can be computed by two curve-fitting functions. The order of the curve-fitting function plays an important role in calculating lag time and CFR. In other words, the proposed algorithm allows the user to provide and change the order of the curve-fitting function. To select the appropriate order for the curve-fitting function, r2 or r-squared provides a statistical measure of how close the data are to the fitted regression line. A maximum of infection and a maximum of death can calculate a lag time and a CFR.

The algorithm was tested by observing many countries included in the scraped csv file: https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/owid-covid-data.csv.

The owid-covid-data.csv file contains the number of daily cases and daily deaths of 238 locations in countries and continents due to COVID-19.

The proposed algorithm is implemented by Python program. The open-source program is called covidlag, a python package index assisting tool for calculating lag time and case fatality rate in ongoing time-series data: https://pypi.org/project/covidlag/.

To use covidlag, it can be installed by pip command in the terminal:

$ pip install covidlag

The hiscovid tool can be also installed on your system as long as Python is installed. In hiscovid, calculating time-series scores can identify when policymakers made mistakes. The lower the score, the better the policy. The steeper the slope of the graph, the more rapidly COVID-19 is spreading.

$ pip install hiscovid

3 Results

The covidlag takes three determinants: country name, the number of days, and the degree of polynomial. As of 2021–10-07, the following command generated Fig. 1 with three determinants for the United States using 600 days with the 13th degree polynomial:

$ covidlag 'United States' 600 13

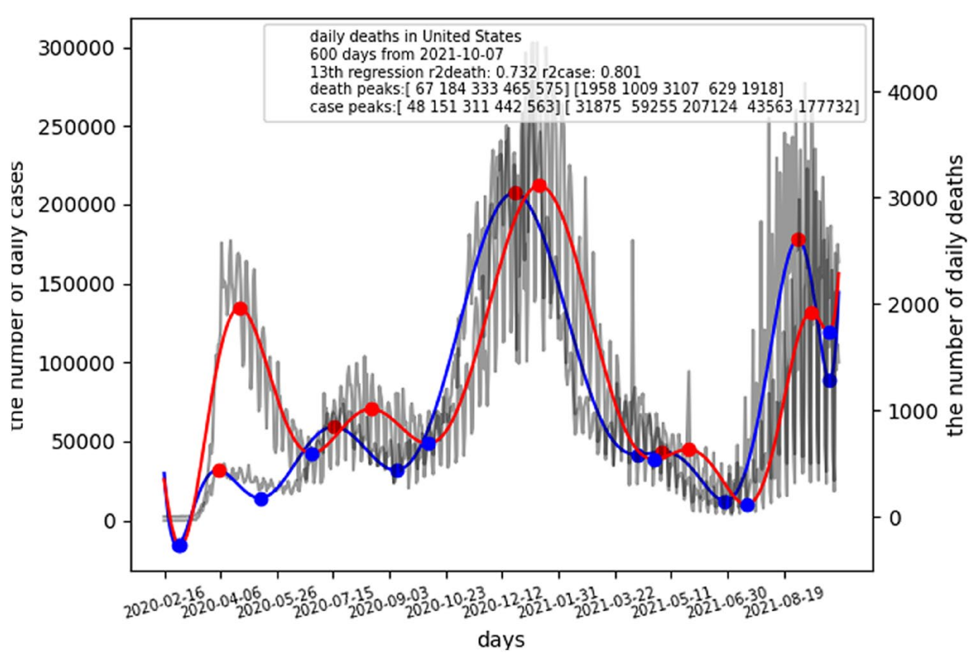

Fig. 1 Lag times and CFRs for the United States as of Oct. 7, 2021, 600 days from 2021-10-07 with 13th degree polynomial regression. *X*-axis indicates days while *Y*-axis on the left side is the number of daily cases and *Y*-axis on the right side is the number of daily deaths

Based on a strong correlation between infection and death, Fig. 1 shows the two curve-fitting functions for the United States. The blue line indicates the curve-fitting function for infection and the red line shows that for death. In Fig. 1, five red points represent maxima and five blue points represent minima. In Fig. 1, r2death: 0.732 and r2case:0.802 are r-squared where its r-squared indicates a goodness-of-fit measure for linear regression models. Figure 1 shows that X-axis indicates days while Y-axis on the left side is the number of daily cases and Y-axis on the right side is the number of daily deaths.

Extreme values of maxima (red points) and minima (blue points) are calculated. Figure 1 shows only maxima points and their values of five death peaks and five case peaks, respectively, in the legend. The number indicates day-index. For example, 68 in death peaks corresponds to 2020-04-23 and 49 in case peaks to 2020-04-04. The lag time of case peak on 2020-04-04 would be $23-4=19$ days. That on 2020-07-16 is $184-151=33$ days, that on 2020-12-23 is $333-311=22$ days, that on 2021-05-03 is $465-442=23$ days, and that on 2021-09-01 is $575-563=12$ days, respectively.

As shown in Fig. 1, the latest infection peak is 563 corresponding to 2021-09-01 and the latest death peak is 575 corresponding to 2021-09-13. The latest CFR at infection peak on 2021-09-01 is $1918/177732=0.011$. CFR at infection peak on 2020-04-04 is $1958/31875=0.061$, that on 2020-07-16 is $1009/59255=0.017$, that on 2020-12-23 is $3107/207124=0.015$, that on 2021-05-03 is $629/43563=0.014$, respectively.

The hiscovid tool can be used for analyzing individual COVID-19 policies by country. To obtain the policy outcomes of New Zealand and the US, the following command should be executed. Figure 2 shows the result of the policy outcomes of New Zealand and the US.

$ hiscovid 'New Zealand' 'United States'

For the detection of maxima and minima in the number of infections and deaths, the 7-day variation can be regarded as noise. However, the covidlag tool allows users to manually specify the number of sampled days. The longer the number of sampled days, the more effectively the 7-day variation noise can be eliminated.

The coefficient of determination R-squared is a measure of how well the polynomial equation fits the data, but it does not need to be a high degree to fit the 7-day variation. Two user-specified parameters, such as the number of sampled days and the degree of polynomial curve fitting, can affect the coefficient of determination R-squared. Therefore, users are allowed to specify these parameters. The covidlag tool is an assistive tool for researchers and policymakers to calculate the lag time and case fatality ratio. However, if there is no strong correlation between cases and deaths, covidlag is not able to provide accurate predictions.

4 Discussion

Using two tools, covidlag and hiscovid, policymakers analyze the dataset, covidlag calculates lag times and CFRs, and hiscovid identifies when and to what extent policymakers made mistakes. The proposed tools are indispensable for policymakers to mitigate the COVID-19 pandemic.

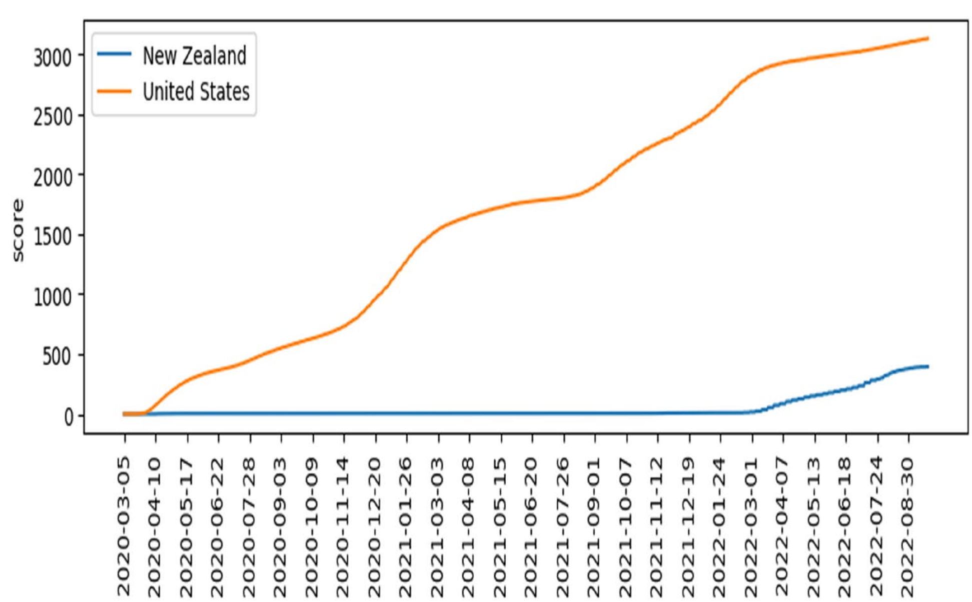

Fig. 2 Time-series policy analysis of New Zealand and the US as of Sept. 21, 2022. Score in time series is calculated by dividing the number of COVID-19 deaths by the population in millions

4.1 covidlag tool

Figure 1 clearly shows that the first CFR was very high compared with other four CFRs. The CFRs are gradually decreasing. The lag time is changing from 19, 33, 22, 23, and 12 days. The latest lag time was suddenly shortened.

As long as extreme values can be calculated, the accurate lag time and CFR will be calculated based on the correlation between infection and death. The proposed method can be used for ongoing time-series data for analysis of the pandemic.

However, the disadvantage of the proposed method lies in that if no extreme value, CFR cannot be calculated. In the midst of a resurge, extreme values cannot be calculated.

Although the proposed method is recently released, 18,237 times have downloaded the covidlag program according to https://pepy.tech/. The large number of users shows that the applicability, the practicality, and the usefulness were justified.

The proposed algorithm allows the user to choose an appropriate order of the curve-fitting lines between infection and death by r-squared value and the algorithm can generate the necessary values for calculating lag times and CFRs. If the appropriate order of the curve-fitting lines is not selected, the proposed algorithm cannot calculate lag times and CFRs. The covidlag tool allows users to select proper determinants such as the number of days and the degree of a curve-fitting polynomial.

The covidlag is detailed at the following site:

https://pypi.org/project/covidlag/

The covidlag tool is validated via Code Ocean for software reproducibility (Takefuji 2022a).

4.2 hiscovid tool

The hiscovid tool can reveal when and to what extent policymakers made mistakes. If the calculated graph is a flat graph, the policy is strongly suppressing the COVID-19 pandemic. The steeper the slope of the graph, the stronger the resurgence of "COVID-19".

Figure 2 shows New Zealand made a single mistake in March 2022. New Zealand's lifting of border restrictions has led to the first surge of COVID-19 in New Zealand (Post 2022; NPR 2022). In Fig. 2, a flat graph of New Zealand until March 2022 indicates that the prevalence of COVID-19 is suppressed with mandatory test-isolation policy by law. The test-isolation policy is to test and identify infected individuals at an early stage and to isolate them from uninfected people during the quarantine period.

The steeply sloping line in the graph indicates a rapid increase in COVID-19. The COVID-19 policy requires real time, evidence-based research and policy updates, keeping track of the number of COVID-19 deaths as well as the economy.

Figure 2 shows that there is no flat graph observed in the US. In other words, the US has never suppressed COVID-19. This is because the US has no mandatory test-isolation policy by law. The steeper the slope of the graph, the stronger the resurgence.

The graph of New Zealand shows that the quarantine period and border regulations can play a key role in mitigating the COVID-19 pandemic. The shorter the quarantine period, the more likely COVID-19 is to spread. The longer the quarantine period, the fewer the number of COVID-19 cases.

Alexander et al. in 2021 reported that there were 33,630 cases of the delta variant of COVID-19 in England in 2021. The delta variant had a number of mutations that make it more transmissible and more likely to cause severe illness. Some experts believed that the delta variant could be even more deadly than the original strain of COVID-19, as the latest delta variant is more likely to lead to severe illness and death, with the latest lag time shortening gradually.

Keep in mind that covidlag is an assistive tool that can be used to calculate the lag time and case fatality ratio in the next 7 days. It is only effective when there is a strong association between cases and deaths. In other words, the tool does not minimize prediction error or the magnitude of error. However, users can specify the number of sampled days and the degree of polynomial curve fitting to calculate the coefficient of determination R-squared which is a measure of how well the polynomial equation fits the data.

5 Conclusion

This paper demonstrated that the proposed covidlag tool allows policymakers to assist in the calculation of CFR and lag time for evaluating the effectiveness of their policies. The covidlag tool has been downloaded 18,237 times worldwide which indicates that the proposed tool is one of the most popular COVID-19 analysis tools in the world. The hiscovid can identify when policymakers made mistakes in time-series scores. Both tools are used to help policymakers make the right decisions for mitigating the COVID-19 pandemic.

The covidlag tool is only effective when there is a strong correlation between cases and deaths. However, the tool does not quantify the degree of correlation, but R-squared is provided as a reference. Quantifying the degree of correlation may be a future work.

The covidlag tool is only effective when there is a strong correlation between cases and deaths. The tool does not quantify the degree of correlation, but R-squared is provided as a reference. In other words, the degree of trust with time-series data is missing. There are the Pearson correlation coefficient and the dynamic time warping which is a technique for comparing two time series that may have different lengths or different sampling rates.

As the future scope, one way to compute the degree of trust with time-series data for predicting the short period in the future is to use an Anticipated Learning Machine (ALM) (Chen et al. 2020). This method is based on non-linear dynamics and can transform recent correlation/spatial information of high-dimensional variables into future dynamical/temporal information of any target variable, thereby overcoming the small-sample problem and achieving multistep-ahead predictions. Extensive experiments on real-world data demonstrate significantly superior performances of ALM over all of the existing 12 methods. However, their method was not used to calculate the lag time and case fatality ratio.

Author contributions Y.T. completed this research and wrote this article.

Funding This research has no fund.

Declarations

Conflict of interest The author has no conflict of interest.

References

Alexandar S, Ravisankar M, Kumar RS, Jakkan K (2021) A comprehensive review on Covid-19 Delta variant. Int J Pharmacol Clin Res 5(83–85):7

Chen C et al (2020) Predicting future dynamics from short-term time series using an anticipated learning machine. Natl Sci Rev 7(6):1079–1091. https://doi.org/10.1093/nsr/nwaa025

Davies NG, Jarvis CI, CMMID COVID-19 Working Group et al (2021) Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. Nature 593:270–274. https://doi.org/10.1038/s41586-021-03426-1

Hsiang S, Allen D, Annan-Phan S et al (2020) The effect of large-scale anti-contagion policies on the COVID-19 pandemic. Nature 584:262–267. https://doi.org/10.1038/s41586-020-2404-8

Newall AT, Leong RNF, Nazareno A, Muscatello DJ, Wood JG, Kim WJ (2020) Delay-adjusted age- and sex-specific case fatality rates for COVID-19 in South Korea: evolution in the estimated risk of mortality throughout the epidemic. Int J Infect Dis 101:306–311. https://doi.org/10.1016/j.ijid.2020.09.1478. (**Epub 2020 Oct 2**)

Nishiura H, Klinkenberg D, Roberts M, Heesterbeek JA (2009) Early epidemiological assessment of the virulence of emerging infectious diseases: a case study of an influenza pandemic. PLoS One 4(8):e6852. https://doi.org/10.1371/journal.pone.0006852

NPR (2022) New Zealand changes its tack on surging COVID-19 cases. https://www.npr.org/2022/03/08/1085095406/new-zealand-changes-tack-on-surging-covid-19-cases

Rothman JE, Eidelberg D, Rothman SL, Holford TR, Rothman DL (2021) Analysis of the time course of COVID-19 cases and deaths from countries with extensive testing allows accurate early estimates of the age specific symptomatic CFR values. PLoS One 16(8):e0253843. https://doi.org/10.1371/journal.pone.0253843

Russell TW et al (2020) Using a delay-adjusted case fatality ratio to estimate under-reporting. Technical Report. Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine.

Takefuji Y (2021a) Correspondence: open schools, Covid-19, and child and teacher morbidity in Sweden. N Engl J Med 384:e66. https://doi.org/10.1056/NEJMc2101280

Takefuji Y (2021b) Analysis of digital fences against COVID-19. Health Technol. https://doi.org/10.1007/s12553-021-00597-9

Takefuji Y (2022ba) Deathdaily: a Python Package Index for predicting the number of daily COVID-19 deaths. Netw Model Anal Health Inform Bioinform 11:14. https://doi.org/10.1007/s13721-022-00359-1

Takefuji Y (2022b) Discovering COVID-19 state sustainable policies for mitigating and ending the pandemic. Cities (lond, Engl) 130:103865. https://doi.org/10.1016/j.cities.2022.103865

Takefuji Y (2022d) Sustainable policy: don't get infected and don't infect others. J Hazard Mater Adv. https://doi.org/10.1016/j.hazadv.2022.100165

Takefuji Y (2023) hiscovid for visualizing and identifying when policymakers made mistakes against COVID-19. Softw Impacts 15:100466

Takefuji Y (2022a) covidlag for calculating lagtime and CFR [Source Code]. https://doi.org/10.24433/CO.8676218.v1

Unnikrishnan J, Mangalathu S, Kutty RV (2021) Estimating under-reporting of COVID-19 cases in Indian states: an approach using a delay-adjusted case fatality ratio. BMJ Open 11:e042584. https://doi.org/10.1136/bmjopen-2020-042584

Washington Post (2022) New Zealand announces plan to reopen to the world after nearly two years of coronavirus travel restrictions. https://www.washingtonpost.com/world/2022/02/03/new-zealand-border-reopening-covid/