Public Health 201 (2021) 89–97

Contents lists available at ScienceDirect

Public Health

journal homepage: www.elsevier.com/locate/puhe

Original Research

The detection of the epidemic phase of COVID-19 and the timing of social distancing policies in Korea

Woohyeon Kim*

Korea Institute of Public Finance, 336, Sicheong-daero, Sejong 30147, Republic of Korea

ARTICLE INFO

Article history:
Received 25 January 2021
Received in revised form
22 September 2021
Accepted 8 October 2021
Available online 18 October 2021

Keywords:
Hidden Markov model
Epidemic phase
COVID-19
Government response stringency index
Effective reproduction number

ABSTRACT

Objectives: Observing cumulative and new daily confirmed cases of COVID-19, disease control authorities respond to a surge in cases with social distancing measures or economic lockdown. The question in this article is whether we can gather more useful information from a readily available time series data set of day-to-day changes in confirmed cases of COVID-19.
Study design: Time-series data analysis was done using a hidden Markov model.
Methods: Day-to-day differences in confirmed cases of COVID-19 in Korea from February 19, 2020, to July 13, 2021, were modeled via a hidden Markov model. The results from the model were compared with the effective reproduction number and the Korean government's response.
Results: The model reports that Korea was in an epidemic phase from August 2020 and from mid-November 2020, the second and third epidemic waves. The government's response, represented by the Government Response Stringency Index, was not timely during the epidemic phases. The results from the model may also be more helpful to detect the onset of the epidemic phase of an infectious disease than the effective reproduction number.
Conclusions: The model can reveal a hidden epidemic phase and help disease control authorities to respond more promptly and effectively.

Introduction

The unprecedented COVID-19 pandemic has thrown the world into crisis. The numbers of cumulative or newly confirmed cases or deaths are released by the media, with the information provided by research centers or websites, such as Johns Hopkins University or Worldometer. Governments respond to surges in confirmed cases with social distancing measures or economic lockdowns. However, the up-to-date case numbers may not be sufficient for health authorities to judge whether or not a serious epidemic phase is underway, requiring tougher action. People may not understand the implicit meaning of the daily fluctuation of the time series data of confirmed cases. A surveillance system with scientific support should process an up-to-date data set and share its understanding of pandemic risk with the public.

South Korea has repeatedly imposed different levels of social distancing measures and partial economic lockdowns, and the government has produced guidelines on easing or tightening these measures. For example, as of December 2020, the Korea Disease Control and Prevention Agency (KDCA) tightened social distancing from level 2 (rapid local transmission, initial phase of national transmission) to level 3 (national epidemic) when the weekly average number of confirmed cases exceeded 800–1000.[a] However, the threshold of 800–1000 cases seems to be unsubstantiated. Even when the actual number of cases did exceed the threshold, KDCA was often reluctant to implement a tougher lockdown policy.[b] Understandably, KDCA assesses a variety of different economic and social factors in addition to the pandemic risk. However, hesitation also comes about because KDCA is not able to detect the true risk of an epidemic phase from the daily number of cases and

* Tel.: +82-44-414-2338; fax: +82-44-414-2309.
E-mail addresses: 7bumblebees@gmail.com, whkim@kipf.re.kr.

[a] The KDCA website, http://ncov.mohw.go.kr/en, Accessed: January 26, 2021. The guidelines of KDCA have changed throughout the pandemic. However, the number of confirmed cases remains as the important determinant in adjusting social distancing measures.

[b] For example, Korea suffered 900 confirmed daily cases during the third week of December 2020, arguably meeting the requirement for social distancing level 3. However, KDCA decided to remain at level 2+ at that time.

https://doi.org/10.1016/j.puhe.2021.10.002
0033-3506/

deaths. This hesitation may confuse the public who have to live their daily lives under social distancing rules.

To help with this situation, established statistical methods are available to detect the early onset of an epidemic. Regression models and other statistical treatments based on historical data sets, from simple summary statistics to cumulative sum statistics, have been actively used for early detection.[1–4] The seminal research introduced a periodic regression to model a fluctuation of weekly pneumonia-influenza deaths in the United States.[5] The model fitted the regular pattern of death cases with the historical data to identify an irregular surge of cases over a predetermined threshold. Several drawbacks have been pointed out, including the need for non-epidemic data to model a normal trend[6] or the independent observations assumption.[7] The need for a long-term non-epidemic data set to model the baseline is particularly vulnerable to a newly discovered infectious disease. ARIMA(Autoregressive Integrated Moving Average)-type time series modeling can also be used to model a fluctuation and detect an irregular perturbation.[8] However, time series modeling may depend on stationarity and single distribution assumptions, which are hard to satisfy in many cases, including epidemics.[9]

Some researchers have paid attention to a hidden Markov model to relax the strong assumptions above.[6,9] Epidemiological data can be readily separated and modeled with different states, an epidemic and a non-epidemic phase underlying the Markov chain. Martínez-Beneito et al.[7] further developed the idea by modeling the week-to-week differences in influenza incidence rate in Spain. They identified an epidemic phase, the period when strong containment measures would be needed to curb the spread of the virus.

This article models the daily confirmed cases of COVID-19 in Korea following the model suggested by Martínez-Beneito et al.;[7] this information is readily available to the public as well as to disease control authorities. The model becomes particularly helpful in understanding when and where an epidemic breaks out in each country or region with an estimated probability of being in an epidemic phase. The epidemic phases of COVID-19 are identified from the daily confirmed cases in Korea from February 2020 to mid-2021. It is then considered how well the epidemic phases correlate with the timing of social distancing and lockdown policies.

The COVID-19 situation and social distancing policies in Korea

The first case of COVID-19 in Korea was reported on January 20, 2020. Since then, it is believed that Korea has been relatively successful in curbing the spread of the virus compared with many other countries. Korean people have conformed to COVID-19 prevention measures, wearing face masks, supporting the aggressive “trace, test, and treat” strategy, and following social distancing rules.[10]

However, efforts to contain the spread of coronavirus have not always been successful. Panel (a) in Fig. 1 illustrates the daily change in confirmed cases in Korea from mid-February 2020 to mid-July 2021. It can be seen that there were at least three distinct outbreaks of COVID-19. Because COVID-19 is a highly contagious disease, the momentary carelessness of a small group of people can lead to widespread exposure to the virus. It is believed that the first two surges of the virus originated exclusively from activities in some local churches.[c] On the other hand, the cause of the third nationwide outbreak starting in November 2020 is unclear.

Whenever the virus has surged, KDCA has taken infectious disease prevention and control measures. In addition, different levels of social distancing have been applied for people in local outbreak areas or nationwide as needed. Because social distancing measures and the related lockdown of small businesses hurt the economy, KDCA has a difficult task in maintaining a balance between preventing outbreaks and sustaining the economy. Therefore, KDCA carefully defined some rules, ranging from mild distancing in daily life to enhanced social distancing. As of December 2020, Korea had five different levels of social distancing, depending on the severity and scale of virus transmission and the pandemic (Table 1).

From KDCA's standpoint, determining when to intervene and adjust social distancing measures is very important. According to the rules in Table 1, changes in daily confirmed cases are the determining factor for imposing social distancing measures. According to the rules, KDCA should tighten restrictions from level 2 (regional) to level 2.5 or 3 (national) when the 7-day average of daily cases peaks at or exceeds 400–500 or when there is a sudden surge in confirmed cases (e.g. doubling or a sudden increase in daily confirmed cases). However, it is not obvious what an average of over 400–500 daily cases means in terms of virus control or how to determine whether a doubling in cases is sudden enough to provoke a shift to the next level of rules.

The effective reproduction number (R_t) provides important information for health authorities.[d] By definition, the number of infected people increases when $R_t > 1$. Much of the literature on epidemiology and economics considers the R_t rate when constructing modeling for the COVID-19 pandemic.[13,14] KDCA reports that it refers to the effective reproduction number as one of the subindicators used to adjust levels of social distancing.[15] However, it is not clear how KDCA incorporates information about R_t into the criteria shown in Table 1. Therefore, some experts in Korea recommend that KDCA should actively use the reproduction number rather than just tracing changes in confirmed cases.[16] The reproduction number has an intuitive meaning. The condition when $R_t > 1$ indicates that a virus is spreading and action is required to contain it. However, R_t is time-lagged information because the information represents a delayed dynamics of transmission.[17] Furthermore, crucial information, including the serial interval and time of symptom onset, may not be readily available to correctly estimate R_t, especially for a newly emerging infectious disease.[17] Authorities may run the risk of releasing biased estimation results without credible prior information.

This study exploits the advantages of the hidden Markov modeling in the context of contagious diseases as suggested by Martínez-Beneito et al.[7] The hidden Markov model has several advantages over other information, including the effective reproduction number. First, the model only requires information that is readily available publicly, that is, daily changes in confirmed cases. This simplicity enables us to generate relevant information in a timely manner, even for a newly infectious disease. Second, the model contemporaneously sheds light on the hidden status of a current epidemic. This information would help authorities to base their decisions to implement painful social distancing and economic lockdown on more complete evidence. The estimation results effectively complement the frequently referenced metrics of the COVID-19 era, including R_t.

[c] On February 18, 2020, a super-spreader was identified in the Shincheonji Daegu branch of the Church, leading to 5212 cases nationwide, particularly in the Daegu-Gyeongbuk area of Korea.[11] On August 3, 2020, Sarang Jeil Church in Seoul became another outbreak epicenter, resulting in 1163 cases nationwide according to KDCA.

[d] The reproduction number represents the average number of subsequent cases from a primary case.[12]

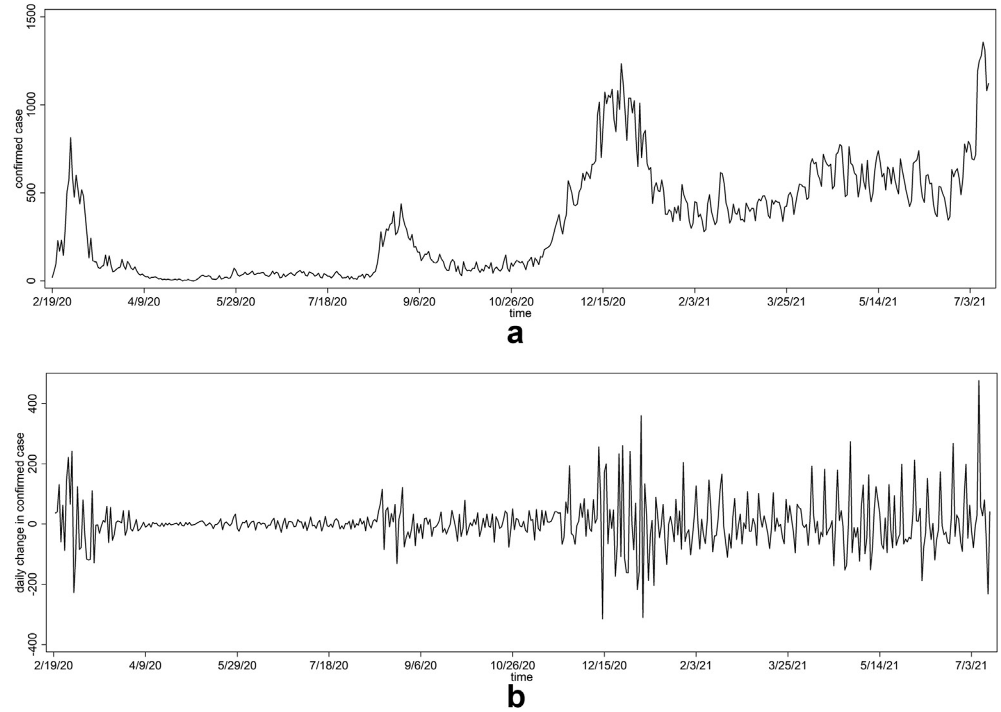

Fig. 1. Daily COVID-19 confirmed case, South Korea, February 19, 2020, to July 13, 2021.

Methods

The question in this article is whether we can gather more useful information from a readily available time series data set: day-to-day changes in confirmed cases of COVID-19. Specifically, the question is how we can determine whether we are in an epidemic phase (the onset of an epidemic) from changes in daily confirmed cases of COVID-19. The hidden Markov model can systematically analyze information on an infectious disease. The model distinguishes the epidemic phase, in which an infectious virus spreads rapidly and the variance in the number of cases increases from a non-epidemic phase with a narrow range of changes in daily case numbers. Indeed, we observe large variations in day-to-day differences in cases at a time when episodes of COVID-19 in Korea were waxing and waning, as shown in panel (b) in Fig. 1. Accordingly, it is reasonable to identify epidemic and non-epidemic phases by observing variations in day-to-day differences in confirmed cases.

Day-to-day differences in cases—$Y_{i,j}$—are modeled on the observations above, where i represents 17 first-tier administrative divisions (metropolitan areas and provinces) in Korea, and j stands for days from February 19, 2020, to July 13, 2021, which is the period of the data set analyzed.

$$Y_{i,j} \mid (Z_{i,j} = 0) \sim \ \mathrm{N}(0,\ \sigma^2_{0,i})$$

$$Y_{i,j} \mid (Z_{i,j} = 1) \sim \ \mathrm{N}(\rho Y_{i,j-1},\ \sigma^2_{1,i})$$

The model shows that if we are in a non-epidemic phase—that is, $Z_{i,j} = 0$—$Y_{i,j}$ follows a normal distribution with mean 0 and variance $\sigma^2_{0,i}$. Once we are in an epidemic phase—that is, $Z_{i,j} = 1$—the variance in the distribution increases to $\sigma^2_{1,i} > \sigma^2_{0,i}$, which indicates that the variance in the day-to-day differences in cases is larger in an epidemic phase than in a non-epidemic phase. In addition, in an epidemic phase, it is reasonable to model the difference in cases today as correlated with the difference in cases yesterday through the parameter ρ, given the characteristics of infectious diseases that spread from an infected person to a healthy person. The hidden daily epidemic status $Z_{i,j}$ is assumed to follow a Markov process through $P_{k,m} = P(Z_{i,j+1} = m \mid Z_{i,j} = k)$, where $k = 0, 1,\ m = 0, 1$. Therefore, the daily epidemic phase transition is governed by the four parameters, $P_{0,0}$, $P_{0,1}$, $P_{1,0}$, and $P_{1,1}$.

A Bayesian framework is used to obtain posterior distributions with appropriate prior distributions for the parameters in the model, $P_{0,0},\ P_{0,1},\ P_{1,0},\ P_{1,1},\ \rho,\ \sigma^2_{0,i},\ \sigma^2_{1,i}$. Following the previous study,[7] hyper-prior distributions are used to represent the condition $\sigma^2_{0,i} < \sigma^2_{1,i}$. More specifically, four ordered statistics $(\theta_{(1)},\ \theta_{(2)},\ \theta_{(3)},\ \theta_{(4)})$ are drawn from a uniform distribution $U(a,b)$, and let $\sigma^2_{0,i}$ and $\sigma^2_{1,i}$ come from $U(\theta_{(1)}, \theta_{(2)})$ and $U(\theta_{(3)}, \theta_{(4)})$, respectively.[e] The condition $\sigma^2_{0,i} < \sigma^2_{1,i}$ in the model is satisfied in this way. The parameters $P_{0,0}$ and $P_{1,1}$ depend on the $beta(0.5,\ 0.5)$ priors; ρ starts from a prior $U(-1, 1)$. After fitting the model, all posterior distributions for parameters are obtained along with samples of daily epidemic status, $Z_{i,j}$, through Gibbs sampling. The posterior average of samples of $Z_{i,j}$ represents the posterior daily probability of being in an epidemic phase for region i at time j.

[e] The parameters of the precedent uniform distribution a, b are assigned according to the variance of the day-to-day differences, $Y_{i,j}$.

Table 1
Social distancing in Korea, basic rules.

Level	Level 1 Distancing in daily life	Level 1.5 Regional level	Level 2	Level 2.5 National level	Level 3
Concept	Distancing in daily life	Local transmission	Rapid transmission, starting phase of national transmission	National transmission	National epidemic
Situation	Daily disinfection and social distancing/control of disease under the medical capacity	Transmission lasts equal to or over 7 days in a specific region threatening the medical system's capacity	Shows increases in transmission despite of Level 1.5 actions/observations of national transmission	National transmission lasts equal to or over 7 days exceeding the capacity of current medical system/surge in number of confirmed cases nationwide and threat of collapse of current medical system	
Criteria	- Average of daily confirmed cases per week • Seoul metro region: below 100 • Chungcheong, Honam, Gyeongbuk, Gyeongnam: below 30 • Gangwon, Jeju: below 10	- Average of daily confirmed cases per week • Seoul metro region: equal to or over 100 • Chungcheong, Honam, Gyeongbuk, Gyeongnam: equal to or over 30 • Gangwon, Jeju: equal to or over 10 - Average of daily confirmed cases per week of ages equal to or above 60 • Seoul metro region: equal to or over 40 • Chungcheong, Honam, Gyeongbuk, Gyeongnam: equal to or over 10 • Gangwon, Jeju: equal to or over 4	When applied to one of the following criteria ① Increase of confirmed cases by 200% lasts after the Level 1.5 actions in epidemic regions ② Level 1.5 actions last for 7 days or longer in two or more regions ③ Number of national daily confirmed cases surpasses 300 for 7 days or longer	- Average of daily confirmed cases per week peaks to or over 400–500 **OR** doubling or sudden increase in confirmed cases during Level 2 ※ The ratio of new confirmed cases of 60 or older, accommodation capability of severe patients, etc. will be considered when increasing the level to 2.5	-Average of daily confirmed cases hits 800–1000 or over **OR** doubling or sudden increase in confirmed cases during Level 2.5 ※ The ratio of new confirmed cases of 60 or older, accommodation capability of severe patients, etc. will be considered when increasing the level to 3
Core Message	Comply with COVID-19 precautionary acts in normal daily/social/economic lives	Regional transmission, thorough social distancing in high-risk regions	Rapid regional transmission, refrain from outings and gathering in high-risk regions and using public facilities	National transmission, stay at home if possible, and refrain from outings and using public facilities	- National epidemic - Stay at home - Minimize contact with others

Ministry of Health and Welfare in Korea, Social Distancing Basic Rules, translated in George Mason University, Mason Korea https://masonkorea.gmu.edu/corona/national-regulations-in-korea/social-distancing, updated: December 11, 2020.

The daily time series of confirmed cases for the 17 administrative divisions[f] of Korea compiled by Statistics Korea were collected for the period February 19, 2020, to July 13, 2021. It is interesting to see how the estimated daily probability of being in an epidemic phase is correlated with other relevant information, namely, day-to-day changes in confirmed cases and the effective reproduction number R_t. The advantages of the new information from the hidden Markov model are presented by comparing the regional population-weighted average of the probabilities with other measures. The corresponding daily effective reproduction number for Korea was extracted from Our World in Data.[12] In addition, a critical policy question in terms of disease control is how the government of Korea actually responded to the COVID-19 situation by adjusting social distancing levels. It is difficult to clearly determine how local governments and KDCA reacted over time. Notwithstanding the current national rules from KDCA, shown in Table 1, rules were constantly revised in line with the ever-changing nature of the pandemic. Furthermore, the local government in each region can tighten or loosen the social distancing level at their discretion. The Government Response Stringency Index[g] is a standardized measure showing how a government's policies and responses evolve.[18] Albeit an imperfect measure, the index can be used to understand how the Korean government reacted on the whole and whether its responses can be considered appropriate in the light of the estimated probabilities of being in an epidemic phase according to the model. The Government Response Stringency Index is also available in Our World in Data.

Results

The posterior distribution of parameters in the model is shown in Table 2, where $\widehat{\rho}$ is estimated to be negative, probably reflecting the serrate-shaped time series data of day-to-day differences in confirmed cases in panel (b) in Fig. 1. This itself may not represent the dominant characteristics of the data flows shown in Fig. 1. On the contrary, the daily transition probabilities $\widehat{P_{0,0}}$ and $\widehat{P_{1,1}}$ are estimated to be extremely high, at 98.7% and 95.7%, respectively, exhibiting the path-dependent tendency of an infectious disease. Therefore, the estimates have the potential to fit the data flows well along with the differences in variances, $\widehat{\sigma^2_{0,i}} < \widehat{\sigma^2_{1,i}}$ coming from $\widehat{\theta_{(1)}} \sim \widehat{\theta_{(4)}}$.

After estimating the probability of being in an epidemic phase for each region i at time j, it is informative to see how the flows of the probabilities and the actual numbers of cases are correlated. The daily number of cases was plotted, and circles were overlaid for the days when the estimated probability of being in an epidemic phase was greater than 50%. Although there may be other ways of interpreting and using the results, it seems reasonable to regard a probability of greater than 50% as a warning sign, following previous studies.[7]

[f] The administrative divisions comprise eight special or metropolitan cities (Seoul, Busan, Daegu, Incheon, Gwangju, Daejeon, Ulsan, and Sejong) and nine provinces (Gyeonggi, Gangwon, Chungbuk, Chungnam, Jeonbuk, Jeonnam, Gyeongbuk, Gyeongnam, and Jeju).

[g] The Government Response Stringency Index, part of the Oxford COVID-19 Government Response Tracker, is a composite measure, which uses nine metrics to measure a government's strictness of policy response. The metrics are school closures, workplace closures, cancellation of public events, restrictions on public gatherings, closures of public transport, stay-at-home requirements, public information campaigns, restrictions on internal movement, and international travel controls. The index ranges from 0 (the least strict response) to 100 (the most strict response).[18]

Table 2
The posterior distribution of parameters.

Parameters	Mean	Standard deviation	25%	median	75%
$\widehat{P_{0,0}}$	0.987	0.002	0.985	0.987	0.988
$\widehat{P_{1,1}}$	0.957	0.006	0.953	0.957	0.961
$\widehat{\theta_{(1)}}$	5.002	0.001	5.000	5.001	5.002
$\widehat{\theta_{(2)}}$	5.006	0.004	5.002	5.004	5.008
$\widehat{\theta_{(3)}}$	6.770	1.339	5.688	6.582	7.602
$\widehat{\theta_{(4)}}$	64.813	6.211	60.433	63.873	68.557
$\widehat{\rho}$	−0.316	0.022	−0.331	−0.316	−0.303

The plots for two regions, the city of Daegu and the Gyeongbuk province, are shown in Fig. 2. As explained in Section 2, the first outbreak in Korea occurred in these two areas during February and March 2020. The model performs well in the sense that the probabilities of being in an epidemic phase capture the onset and decline of the pandemic in February and March 2020. More helpfully, the epidemic probabilities beneath the actual confirmed cases distinguish an epidemic from a non-epidemic in a more scientific manner.

Fig. 2 also illustrates the relatively strong performance of the model for other periods of the pandemic in Korea. The second wave of the pandemic, in August 2020, occurred mostly in Seoul, the capital city of Korea, and nearby metropolitan areas, Incheon city and Gyeonggi province. From mid-August 2020, the model warns of the onset of the pandemic in these areas. The model provides alerts again for Seoul and Incheon from early- or mid-November 2020 during the nationwide third wave of the pandemic. For the third wave, it is interesting that the model flags warnings for Gangwon and Gyeongnam provinces, which show upward trends of confirmed cases from mid-November 2020. These two areas did not previously suffer from the pandemic during the first and second waves. On the other hand, Sejong city and Jeonnam province, for example, do not show upward trends in the number of confirmed cases during the third epidemic wave; these two areas are known to be successful in containing the outbreak because of their population size and density, showing relatively stable case numbers over the period. The model hardly gives any warning for Sejong city and Jeonnam province.

As each regional epidemic probability is effective in analyzing and detecting the early onset of the epidemic locally, a local population-weighted average of the probability of being in an epidemic phase illustrates another way of viewing the national pandemic. Panel (a) in Fig. 3 shows daily confirmed cases and the hidden local population-weighted average of epidemic probabilities. The average probabilities stand out during the second and third waves of the pandemic in Korea and beyond. Although the numbers of confirmed cases in the first pandemic were greater than those in the second, the model is silent for the first period. This result indicates that a locally severe outbreak in the first period may not have been serious at the national level, meaning that locally intensive disease controls were appropriate at that time. On the other hand, there was a need for KDCA to focus on social distancing and other control measures nationwide during the second and third waves. The model helps to understand the real-time epidemic situation locally and nationally and to ensure that appropriate measures are taken.

The effective reproduction number and the hidden Markov model exhibit quite different patterns in some periods. Again, a reproduction number greater than 1 is a warning sign of being in an epidemic phase. Panel (b) in Fig. 3 uses circles to identify the days where $R_t > 1$. Although the time series of the effective reproduction number corresponds fairly well with the first, second, and third waves of the pandemic in Korea, the numbers are also greater than 1 for most of May, June, and July 2020. The changes in confirmed cases

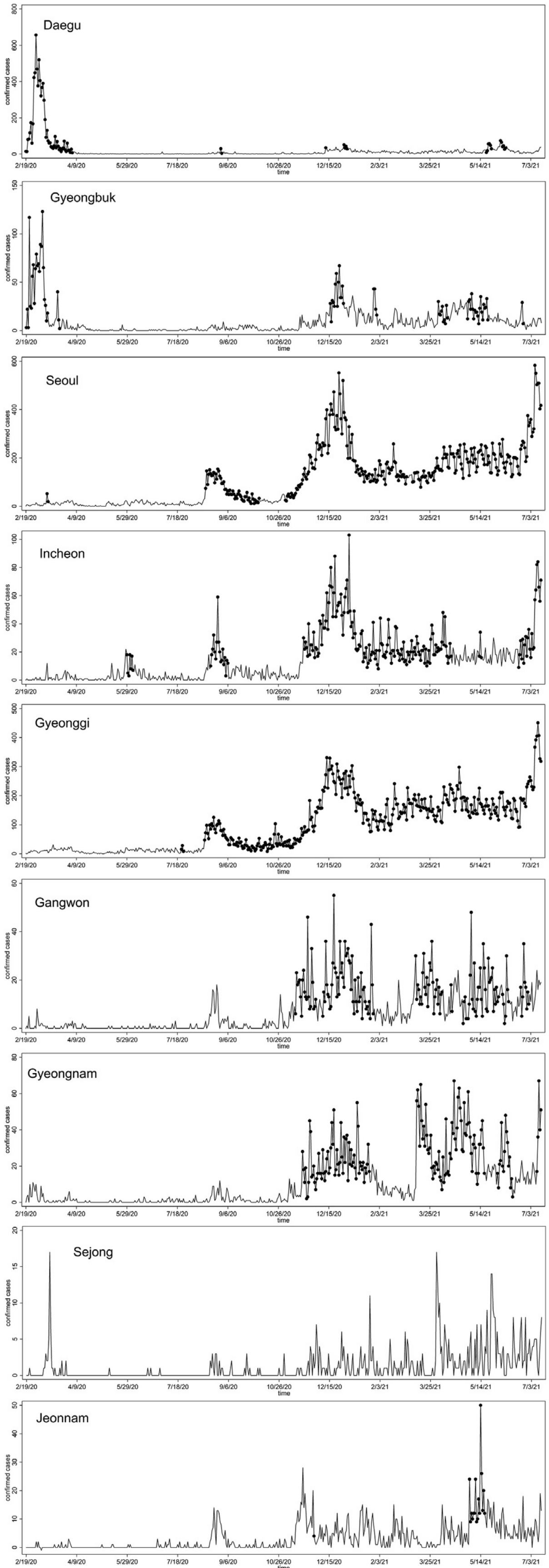

Fig. 2. Numbers of confirmed cases and the probabilities of being in an epidemic phase in different regions.

remained stable during this period when KDCA lowered the level from social distancing (level 2) to distancing in daily life (level 1). The figure shows that the effective reproduction number may be excessively sensitive for correctly detecting the onset of infectious disease. From a public policy point of view, the hidden Markov model more clearly distinguishes epidemic and non-epidemic phases.

An additional distinction is that the hidden Markov model presents a more conservative identification of onset compared with the effective reproduction number. For example, the reproduction number produces a warning sign until August 30, 2020, three days after the number of cases reached the peak during the second wave. However, the hidden Markov model remains cautious until September 9, 2020, when the time series of confirmed cases appears to be completely back to normal. For the third wave and beyond, the distinction is more pronounced, as the hidden Markov model consistently flags warnings, whereas the effective reproduction number does not.

Finally, a central policy question is whether the actual government responses in Korea correspond to the hidden status of the epidemic. Panel (c) in Fig. 3 shows the Government Response Stringency Index of Korea.[18] The index does not seem to be highly correlated with the probability of being in an epidemic phase according to the model. Therefore, from a policy point of view, this implies that the Korean government could have been more aggressive in its response in the periods with warning signs, that is, the second and third waves of the pandemic.

Discussion

Since 2020, the world has faced the highly contagious disease COVID-19. In Korea, adopting an aggressive “trace, test, and treat” strategy with tough social distancing and economic lockdown rules has been considered relatively successful in containing the spread of the epidemic.[19,20] However, local lockdowns and social distancing policies have taken a heavy toll on the economy, particularly on vulnerable economic groups, such as small business owners. According to Korea Credit Data,[h] retail sales in 2020 were lower than 2019 almost every week. Therefore, the assessment of the risk of pandemic locally and nationally in an accurate and timely manner is more important than ever before.

This study has shown how a hidden Markov model can be used to understand real-time COVID-19 situations. The model reports that Korea was in an epidemic phase during August 2020 and in the period from mid-November onward, the second and third waves. The results can help both the authorities and the public understand the current spread of the virus and take appropriate action. According to the results of the model, the policy responses in Korea may not have been as timely as they could have been. Finally, the effective reproduction numbers appear to represent different information compared with the results of the model. The hidden Markov model clearly separates epidemic and non-epidemic phases, which, from a policy point of view, is more useful for detecting the onset of an infectious disease and adjusting relevant disease control measures.

To evaluate whether the model performs well in other settings, COVID-19 cases in five other countries (the United States, the United Kingdom, India, New Zealand, and Brazil) were analyzed with the same model.[i] The model continuously raises a warning

[h] Korea Credit Data (KCD) is a for-profit financial technology company that collects and provides business transaction information. It compares changes in sales in the year 2020 with the same weeks in the previous year. See the online supplemental material.

[i] The results are available in the online supplemental material.

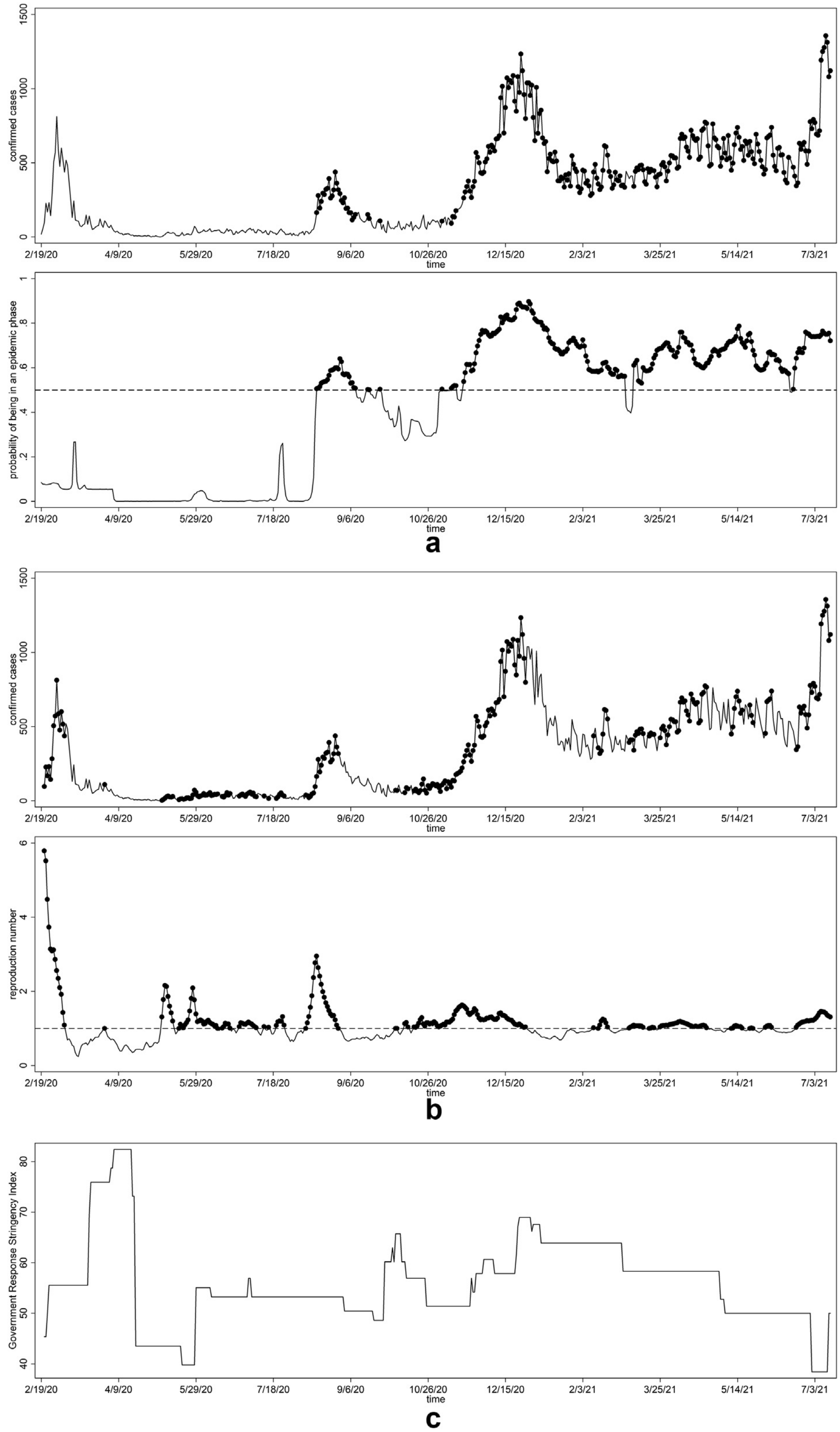

Fig. 3. Population-weighted averages of the probability of being in an epidemic phase, the effective reproduction numbers, and Government Response Stringency Index for Korea.

flag for the United States, India, and Brazil, who have suffered from a high number of cases of COVID-19 throughout the pandemic, whereas it stays relatively silent for New Zealand and during some calm periods for the United Kingdom.

Compared with other established models, the model has some attractive features for identifying the outbreaks of infectious diseases. The hidden Markov model itself fits quite naturally with the mixture of distributions explaining different states of epidemic and non-epidemic periods. Furthermore, conventional Serfling-type classical regression models usually require long series of historical epidemic data to perform well.[3] When it comes to a newly emerging infectious virus such as COVID-19, this means its performance for surveillance, monitoring, and evaluation may be weaker. The model introduced in this study was estimated via Bayesian framework, an intuitive way to understand the current status in the absence of sufficient data.[21] A priori knowledge of infectious disease is combined with gradually updating new daily information, which resembles the way we process newly available information.

In addition, the effective reproduction number, a well-known measure for understanding the intensity of an infectious disease outbreak in epidemiology, may not be sufficient to capture the dynamics of disease spread, especially from the public health policy point of view. As shown in Fig. 3, during the period under study, the reproduction number turns out to be sensitive with respect to the threshold $R_t > 1$. Specific and timely warning and social distancing implementation may be difficult if depending solely on the observation of changes in the reproduction numbers. The model in this study successfully differentiates epidemic and non-epidemic phases amid extreme fluctuations in confirmed cases.

From a quarantine perspective, the model can provide informative answers on how to prepare to treat COVID-19 patients with respect to medical resources, such as hospital beds, staff, and so on. This is because the model in this study models daily differences in confirmed cases, not the number of cases itself. Therefore, the probability of being in an epidemic phase itself is related to the differences in confirmed cases locally and nationally by model construction. The disease control and prevention agencies who have responsibility to distribute the resources to hospitalize and treat patients may benefit from the scientific results by modeling the daily differences in confirmed cases locally and nationally.

In addition, the epidemic probability from the model may be used to perform a cost-benefit analysis of social distancing policies. As mentioned, social distancing policies and economic lockdowns have been painful, especially for small business owners. The government should measure the total benefits and costs of strengthening or weakening lockdown policies when needed. The probability of the severity of virus spread can be a readily available component for measuring the benefits and costs of those policies in cost-benefit quantitative analysis.

There are limitations to the model which should be explored in future research. First, the model gives a warning during a period of a rapid decline of confirmed cases by construction because it is designed to recognize a large variation of differences as an epidemic phase. In Fig. 3, we can observe a clear difference of warning signs between the hidden Markov model and the effective reproduction numbers in the winter of 2020. The reproduction numbers explain the decline of number of cases in a timely manner, whereas the hidden Markov model displays a more conservative attitude and continues to give a warning until a stationary time series of confirmed cases is observed. This study did not analyze how to evaluate and determine how conservative we should be in terms of quarantine policy. Both methods have their own pros and cons, but these may need to be explored.

In relation to the limitation mentioned previously, some may point out that the model can become silent during a plateau in the time series of confirmed cases. Theoretically, it is possible for the model to stay calm when a high number of cases continues with little fluctuation. This is a possible limitation of the model and should be further examined, although, considering the nature of infectious disease, the situation of a high constant plateau in a series of confirmed cases may be unlikely.

The model estimates a daily probability of being in an epidemic phase but does not directly show when to adjust the level of social distancing. This research follows previous studies regarding the period when an epidemic probability becomes greater than 50% as an epidemic phase.[7] In a real setting, the threshold for detecting an epidemic phase may not apply for all related authorities or the public. Future research should scrutinize the relationship between social distancing measures and the probability of being in an epidemic phase during the COVID-19 outbreak. Understanding this relationship is essential in an ex-ante social distancing and lockdown policy simulation.

Although disease control authorities set social distancing and lockdown measures based on the information observed, tracing daily changes in confirmed cases may not tell them directly what to do. The main contribution of the model in this article is that it can reveal a hidden epidemic phase and guide disease control authorities to respond in a more scientific manner. Although authorities have their own disease control guidelines (see, for example, Table 1), it may be difficult for the authorities to take persuasive action against vocal complaints from the public who are suffering from prolonged lockdown and social distancing measures. Therefore, evaluating the real-time level of pandemic risk becomes more important for communicating with the public and taking appropriate action.

Author statements

Ethical approval

Not required. This study analyzed data in the public domain.

Funding

This research did not receive any funding from any sources.

Competing interests

None declared.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.puhe.2021.10.002.

References

1. Hutwagner LC, Maloney EK, Bean NH, Slutsker L, Martin SM. Using laboratorybased surveillance data for prevention: an algorithm for detecting salmonella outbreaks. *Emerg Infect Dis* 1997;**3**(3):395.
2. Tay Ee Laine, Grant Kristina, Kirk Martyn, Mounts Anthony, Kelly Heath. Exploring a proposed who method to determine thresholds for seasonal influenza surveillance. *PLoS One* 2013;**8**(10):e77244.
3. Cowling Benjamin J, Wong Irene OL, Ho Lai-Ming, Riley Steven, Leung Gabriel M. Methods for monitoring influenza surveillance data. *Int J Epidemiol* 2006;**35**(5):1314–21.
4. Wang Ruiping, Jiang Yonggen, Michael Engelgau, Zhao Genming. How to select a proper early warning threshold to detect infectious disease outbreaks based on the China infectious disease automated alert and response system (cidars). *BMC Publ Health* 2017;**17**(1):1–10.
5. Serfling Robert E. Methods for current statistical analysis of excess pneumonia-influenza deaths. *Publ Health Rep* 1963;**78**(6):494.
6. Rath Toni M, Carreras Maximo, Sebastiani Paola. Automated detection of influenza epidemics with hidden markov models. In: *International symposium on intelligent data analysis*. Springer; 2003. p. 521–32.
7. Martínez-Beneito Miguel A, Conesa David, López-Quílez Antonio, López-Maside Aurora. Bayesian Markov switching models for the early detection of influenza epidemics. *Stat Med* 2008;**27**(22):4455–68.

8. Watier Laurence, Richardson Sylvia, Hubert Bruno. A time-series construction of an alert threshold with application to s. bovismorbificans in France. *Stat Med* 1991;**10**(10):1493–509.
9. Le Strat Yann, Carrat Fabrice. Monitoring epidemiologic surveillance data using hidden markov models. *Stat Med* 1999;**18**(24):3463–78.
10. Lim Soo, Yoon Holl, Song Kyoung-Ho, Kim Eu Suk, Kim Hong Bin. Face masks and containment of COVID-19: experience from South Korea. *J Hosp Infect* 2020;**106**(1):206–7.
11. Kim Mi-Young, Kweon Sang-hui, Baek Soo-Jin, Jeon Byoung-Hak, Yoo Hyo-Soon, Park Young-Joon, et al. Weekly report on the covid-19 situation in the Republic of Korea (as of April 11, 2020). *Public Health Weekly Report* 2020;**13**(16):972–85. Korea Disease Control and Prevention Agency.
12. Arroyo-Marioli Francisco, Bullano Francisco, Kucinskas Simas, Rondón-Moreno Carlos. Tracking R of COVID-19: a new real-time estimation using the Kalman filter. *PLoS One* 2021;**16**(1):e0244474.
13. Budish Eric B. *Maximize utility subject to $R \leq 1$: a simple price-theory approach to covid-19 lockdown and reopening policy*. NBER Working Paper, (w28093). 2020.
14. Choi Sunhwa, Moran Ki. Estimating the reproductive number and the outbreak size of covid-19 in korea. *Epidemiology and health* 2020;**42**.
15. KDCA. *Media briefing on 1 Nov. 2020 from kdca*. 2020. *in* Korean.
16. Kim Nam-Jong. Economic discussions on how to react the covid-19 pandemic considering the effective reproductive number. *Finance Brief* 2020;**29**(24): 3–11. *in* Korean.
17. Lim Jun-Sik, Cho Sung-Il, Ryu Sukhyun, Pak Son-Il. Interpretation of the basic and effective reproduction number. *J Preventive Med Public Health* 2020;**53**(6):405.
18. Hale Thomas, Petherick Anna, Phillips Toby, Webster Samuel. *Variation in government responses to covid-19*. Blavatnik school of government working paper, 31. 2020.
19. Peck Kyong Ran. Early diagnosis and rapid isolation: response to COVID-19 outbreak in korea. *Clin Microbiol Infect* 2020;**26**(7):805–7.
20. Park Sang Woo, Sun Kaiyuan, Viboud Cécile, Grenfell Bryan T, Dushoff Jonathan. Potential role of social distancing in mitigating spread of coronavirus disease, South Korea. *Emerg Infect Dis* 2020;**26**(11):2697.
21. Long Bo, Chang Yi. *Relevance ranking for vertical search engines*. Newnes; 2014.