

Estimating and forecasting the spread of COVID-19 in South Korea: A Bayesian SIHR-based dynamic model with non-pharmaceutical interventions

Hyunho Choi^{1*}, Dayun Kang^{1*}, Unkyoung Lee², Jong-Hun Kim³, Sunwha Choi⁴, Jungsoon
Choi^{5,6**}

¹ Department of Applied Statistics, Hanyang University, Seoul, Republic of Korea;
gusgh4950@hanyang.ac.kr (H.C.); dayun4927@hanyang.ac.kr (D.K.)

² Department of Preventive Medicine, Hanyang University, Seoul, Republic of Korea;
ukyoungee22@hanyang.ac.kr (U.L.)

³ Department of Social and Preventive Medicine, Sungkyunkwan University School of
Medicine, Suwon, Republic of Korea; kimjh32@skku.edu (J.H.K.)

⁴ Department of Cancer Control and Population Health, Graduate School of Cancer Science
and Policy, National Cancer Center, Goyang, Republic of Korea; ssunhwa.choi@gmail.com (S.C.)

⁵ Department of Mathematics, Hanyang University, Republic of Korea;
jungsoonchoi@hanyang.ac.kr (J.C.)

⁶ Research Institute for Natural Sciences, Hanyang University, Republic of Korea

* The first two authors contributed equally to the manuscript.

** Corresponding author's e-mail address: jungsoonchoi@hanyang.ac.kr; Tel.: +82-2-2220-

2621

Abstract

Background: As of May 22, 2020, the total number of confirmed COVID-19 cases is over 5 million worldwide and more than 300 thousand people have lost their lives to the virus. South Korea also experienced a sharp increase

in late February, but owing to non-pharmaceutical interventions, the number of confirmed cases has been decreasing since March. In this study, we aimed to investigate the transmission dynamics with these effects as well as forecast the spread of COVID-19 in South Korea, using a flexible statistical model.

Methods: We analyzed the COVID-19 data obtained at the Korea Centers for Disease Control and Prevention from Feb 18 to April 30. Using a Bayesian susceptible-infectious-hospitalized-removed (SIHR) dynamic model, we estimated the dynamic transmission rate considering the non-pharmaceutical intervention effects and forecast the confirmed cases.

Results: The estimated transmission rate without any control effects was 0.4605 with 95% credible interval (0.4468, 0.4745). During the days with effects between February 26 and March 6, the daily transmission rate decreased by about 89.48% of that of the previous day. With consistent control effects, it remained at 0.1549 with 95% credible interval (0.1497, 0.1602). Based on the estimated transmission rate, the forecast number of COVID-19 infections in South Korea showed an overall decreasing pattern.

Conclusions: We considered and estimated the dynamic effects of the non-pharmaceutical interventions on COVID-19 using a Bayesian SIHR-based model. This study shows that non-pharmaceutical interventions including active testing, quarantine and isolation, personal preventive measures, and social distancing are crucial to curb the transmission. Our findings contribute to a better understanding of non-pharmaceutical interventions in COVID-19.

Keywords: COVID-19, SIHR model, Statistical model, Bayesian inference, South Korea, Non-pharmaceutical intervention

Background

An outbreak of novel coronavirus (SARS-CoV-2) was first reported by the Chinese government on December 31, 2019. It occurred in Wuhan, the provincial capital of Hubei Province. Since then, the virus has so far affected over 210 countries [1]. As of May 22, 2020, the total number of confirmed cases around the world is 5,161,289 and 334,677 people have lost their lives to the virus [1]. To control the spread of COVID-19, many countries have implemented various non-pharmaceutical interventions such as active testing, quarantine and isolation, personal

protective equipment (PPE), social distancing, the closure of schools, and national lockdowns. Thus, it is important to investigate the transmission dynamics considering the non-pharmaceutical interventions and forecast the number of COVID-19 infections. Such study can allow to provide real-time information to healthcare workers and help policymakers to create more evidence-based policies.

South Korea has also been affected by the virus since the index case occurred on January 20. Until February 17, 17 out of 30 cases in the country were imported from abroad, accounting for more than 50%. However, the country's 31st patient, a follower of Shincheonji, which is a minor South Korean Christian sect, occurred in Daegu on February 18. The patient was found to have had attended a few services in Daegu where a mass infection in the religious group occurred. Thus, the number of new confirmed cases dramatically increased to about 400 cases until February 22 (4 days). As a result, the Korean government raised the coronavirus alert to the highest level on February 23 [2], and strongly encouraged citizens to practice social distancing, wear face masks, and wash hands frequently, to try and control the spread of COVID-19. In addition, COVID-19 diagnostic testing capabilities have been dramatically expanded, and a large-scale and rapid virus testing was conducted to identify the infected patients at the early stage. Finally, the daily average number of new confirmed cases in mid-April was less than 10, and the government shifted to mitigation. Thus, it is important to estimate the dynamic transmission rate considering the effects of the non-pharmaceutical interventions in South Korea. Forecasting the spread of COVID-19 is also important because the pandemic is on-going. However, some studies on COVID-19 in South Korea assumed a constant transmission rate using mathematical models to forecast the number of COVID-19 infections [3-5]. Recently, a deterministic SIR-based econometric model was used to estimate the effects of non-pharmaceutical intervention policies in South Korea, but not to forecast the number of coronavirus infections [6].

In this work, we estimated the impact of the non-pharmaceutical interventions on COVID-19 in South Korea from February 18 to March 31 using a Bayesian susceptible-infectious-hospitalized-removed (SIHR) statistical model. In addition, we forecast the number of new COVID-19 infections in South Korea, using the proposed model. To the best of our knowledge, no studies estimated intervention effects and forecast the spread of COVID-19 together.

Methods

Data sources

The COVID-19 dataset was obtained from the Korea Centers for Disease Control and Prevention (KCDC) website (<http://www.cdc.go.kr/>) and the Gyeonggi Infectious Disease Control Center (GIDCC) website (<http://gidcc.or.kr>). The KCDC website reports daily new confirmed cases through press releases. The GIDCC additionally provides information on the symptom onset date of the confirmed patients in Gyeonggi.

Since we mainly wanted to investigate the transmission dynamics within the local community, we did not use the cases imported from abroad. As the number of cases increased from February 18 and most of the earlier cases were travel-related, we used the number of confirmed cases from February 18 to April 30. There were 10,125 confirmed cases during this period. In the proposed SIHR model, we require the symptom onset date information rather than the confirmation date information. However, we had the symptom onset date information only for Gyeonggi, which accounts for 5.2% of the total data. Based on the Gyeonggi dataset, we assumed a Gamma distribution with mean 4.35 and variance 11.76 for the period from symptom onset to the confirmation. We randomly sampled the periods from the distribution and estimated the symptom onset date of the individuals in other provinces from February 18 to April 30. Based on the symptom onset date information, we used the number of cases from February 18 to March 31 for the modeling and the number of cases from April 1 to April 30 for the calibration.

Figure 1 (a) shows the time series plots of the daily number of patients with symptom onset and confirmed cases and Figure 1 (b) shows corresponding cumulative cases in South Korea from February 17 to March 31. The number of cases escalated dramatically from mid- to end-February and seem to have decreased since then. The number increased sharply in late February due to the mass infection in Shincheonji. Figure 2 represents the cumulative number of confirmed cases for each province. The two dark red areas on the map are Daegu and North Gyeongsang province, where the Shincheonji mass infection occurred.

Statistical model

Many researchers have conducted mathematical model-based studies to investigate the transmission dynamics and forecast the number of COVID-19 infections. A mathematical susceptible-infectious-recovered (SIR) and susceptible-exposed-infectious-recovered (SEIR) were considered to forecast the number of coronavirus infections in China [7, 8]. A flowchart of the SIHR model is presented in Figure 3. The classical mathematical

SIHR model has four compartments: susceptible (S), infectious (I), hospitalized (H), and removed (R). The basic model framework in the mathematical SIHR model can be described as follows:

$$\frac{dS}{dt} = -\beta \frac{I}{N} S,$$

$$\frac{dI}{dt} = \beta \frac{I}{N} S - \alpha I,$$

$$\frac{dH}{dt} = \alpha I - \gamma H,$$

$$\frac{dR}{dt} = \gamma H,$$

where the parameters α, β, γ are isolation, transmission, and removal rates, respectively. The inverse of the isolation rate, $1/\alpha$ indicates the average interval from the symptom onset to hospitalization, that is, the infection transmission period. The inverse of the removal rate, $1/\gamma$ is the average quarantine period from hospitalization to removal. The mathematical model assumes that the sum of the individuals in the four categories is the total population (N) and does not change over time. A mathematical framework is convenient for a compartment model using differential equations. However, a mathematical model, especially a deterministic model, relies heavily on parameter assumptions and cannot contain uncertainty in the real world.

We proposed a Bayesian statistical SIHR model to investigate the transmission dynamics of COVID-19 and forecast the number of COVID-19 infections in South Korea. A Bayesian approach considers the parameter distributions to explain the uncertainty of the parameters, and it can allow a natural and flexible framework to capture complicated dynamics. Thus, we are able to figure out how the intervention effects affect the transmission rate. This would aid the authorities to implement more appropriate and effective intervention policies to curb the transmission.

We assume that the number of new infectious individuals with symptoms at time t , x_t , that is, the number of

individuals moving from S to I at time t , follows a Poisson distribution with mean λ_t ,

$$x_t \sim Poi(\lambda_t), \quad t = 1, \dots, 75.$$

Based on the mathematical model described above, the time-varying transmission rate is considered as follows:

$$\log(\lambda_t) = \log(\delta^{h(t)}\beta) + \log(I_{t-1}) + \log(S_{t-1}) - \log(N),$$

$$\text{where } h(t) = \begin{cases} 0, & t \leq T_1 \\ t - T_1, & T_1 < t \leq T_2, \\ T_2 - T_1, & t > T_2 \end{cases}$$

where β , I_{t-1} , and S_{t-1} indicate the transmission rate, the number of infectious individuals, and susceptible individuals at time $t - 1$, respectively. The parameter $\delta^{h(t)}$ indicates the daily reduction of the transmission rate due to the preventive measures and T_1 and T_2 represent the time points when the intervention effects start and become consistent, respectively. We took account of the effects on the transmission rate, since the more people comply with the preventive measures, the fewer will be infected with the virus and pass it on to others. To better understand the model, the model formula is re-expressed as follows:

$$\log(\lambda_t) = \log(\beta) + \log(I_{t-1}) + \log(S_{t-1}) - \log(N), \quad t \leq T_1,$$

$$\log(\lambda_t) = \log(\delta^{t-T_1}\beta) + \log(I_{t-1}) + \log(S_{t-1}) - \log(N), \quad T_1 < t \leq T_2,$$

$$\log(\lambda_t) = \log(\delta^{T_2-T_1}\beta) + \log(I_{t-1}) + \log(S_{t-1}) - \log(N), \quad t > T_2,$$

and the transmission rate before the control effects begin is β . During the period from T_1 to T_2 , it becomes δ times that of the previous day and at time point T_2 , it becomes $\delta^{T_2-T_1}\beta$. After time point T_2 , it remains the same as $\delta^{T_2-T_1}\beta$. Therefore, from T_1 to T_2 , individuals and the government made vigorous efforts to curb the transmission and their effects remained constant from time T_2 as the number of new infections stabilized.

In the same manner as the mathematical model, the number of individuals in compartments H and R at time t are $H_t = \alpha I_{t-1} + (1 - \gamma)H_{t-1}$ and $R_t = \gamma H_{t-1} + R_{t-1}$, respectively. Based on our dataset, we assumed that α is $1/4$ and γ is $1/14$ [9].

For the transmission rate parameters, we used a Bayesian framework with prior distributions. For the parameters β and δ , we used a Gamma distribution with mean 1 and variance 10 and Beta distribution with mean $1/2$ and

variance 1/12, respectively.

To compare the model performance of the SIHR model with the non-pharmaceutical intervention effects, we additionally considered an SIHR model without the effects as a competing model, as follows:

$$\log(\lambda_t) = \log(\beta) + \log(I_{t-1}) + \log(S_{t-1}) - \log(N),$$

where the transmission rate β remains consistent during the study period.

We determined the time points T_1 and T_2 based on the model performance using mean squared prediction error (MSPE) and deviance information criterion (DIC, [10]). As shown in Table 1, February 26 and March 6 for T_1 and T_2 showed the best model performance, which is our final model. In addition, we compared the MSPE and DIC values of the models with or without intervention effects. The model with non-pharmaceutical intervention effects showed a dramatic improvement in both prediction power and model fitness.

Table 1. Performance of the proposed models and competing model.

Model	T_1	T_2	MSPE ^a	pD ^b	DIC ^c
	Feb 25	Mar 4	688.45	1.983	687.483
	Feb 25	Mar 6	504.49	1.988	665.527
	Feb 25	Mar 8	567.08	1.988	704.562
Model with	Feb 26	Mar 4	602.53	2.006	683.867
intervention	Feb 26	Mar 6	490.94	2.018	677.084
effects	Feb 26	Mar 8	602.29	1.968	728.054
	Feb 27	Mar 4	612.12	1.994	696.903
	Feb 27	Mar 6	578.58	2.002	706.242
	Feb 27	Mar 8	745.67	1.987	769.500
Model without					
intervention			6512.68	0.987	2532.44
effects					

^a MSPE: Mean squared prediction error, ^b pD: the effective number of parameters, ^c DIC: Deviance information

criterion

For parameter estimation, we used Markov Chain Monte Carlo (MCMC) sampling utilizing the R-package R2WinBUGS [11]. We used two parallel chains with different initial values. To check sample convergence, we utilized the Gelman-Rubin statistic, trace plots, and auto-correlation. For each chain, after the burn-in of 10,000 samples, 10,000 samples were used for parameter estimation with thin 10.

Results

Table 2 shows the parameter estimates of the proposed model. The transmission rate before the non-pharmaceutical intervention effects, β , is 0.4605 with 0.4468 and 0.4745 for lower and upper bounds of the 95% credible interval, respectively. During the days with the effects, the transmission rate decreases by about 89.48% every day as the estimate of δ is 0.8948. Since the non-pharmaceutical intervention effects stopped increasing, the transmission rate, $\beta\delta^{10}$, remains at 0.1549, which is nearly 33.6% of the initial transmission rate.

Table 2. Parameter estimates in the proposed model.

Parameter	Mean	95% CI ^d (Lower, Upper)
β	0.4605	(0.4468, 0.4745)
δ	0.8948	(0.8924, 0.9012)

^d CI: Credible Interval

Figures 4 (a) and (b) show the number of patients with the forecast number of new infections for the next thirty days using the model without intervention effects and the proposed model, respectively. The points represent the number of patients based on the symptom onset date and the line shows the posterior means of the estimated number of new infections. Grey shades indicate the 95% credible interval of the estimated number of new infections. The two graphs represent a dramatic improvement in prediction performance in the proposed model.

Figure 5 shows the cumulative number of patients and also represents the dramatic improvement in model performance in the proposed model. As shown in Figures 4 (b) and Figure 5 (b), the proposed model fits the dataset quite well. The estimated number of infected individuals is the highest on February 27 and has decreased since then.

Discussion

We examined the dynamic effects of non-pharmaceutical interventions for COVID-19 using a Bayesian SIHR-based dynamic model and forecasted the number of individuals infected with COVID-19 in South Korea. The transmission rate after the period with vigorous non-pharmaceutical interventions was nearly one third of the initial transmission rate. As the number of cases escalated sharply from February 18, it seems that the authorities carried out active testing to isolate patients and people maintained social distancing with PPE. In addition, the media emphasized personal hygiene regulations such as wearing masks and washing hands frequently. Due to these efforts, the transmission rate started to decrease from February 26 and, according to our results, reached its lowest level on March 6.

We also determined that the forecast number of infections was the highest at the end of February. However, there are many other factors that may lead to a resurgence in COVID-19 infections. For example, since the Korean government shifted from strict social distancing to routine distancing on April 20 and schools have started to reopen, more people will likely come into contact with each other in the community. Thus, unexpected mass infections may occur.

To the best of our knowledge, this is the first study to estimate the impact of non-pharmaceutical interventions on COVID-19 as well as forecast the spread of COVID-19 together. To do such study, we developed a statistical compartment model with non-pharmaceutical interventions within a Bayesian framework. We modeled the uncertainty of the transmission dynamics and could obtain reasonable results. Motivated by the intervention analysis in a time series modeling, we also modeled the impact of non-pharmaceutical interventions in the transmission rate. The proposed model is very flexible because it can be easily extended by adding or removing some compartments. The period and structure of intervention effects can vary based on the dataset or type of interventions.

Despite the strengths of our model, there are some limitations. First, we were able to obtain the limited information for the symptom onset date. Due to the lack of such data, the symptom onset dates were estimated using the confirmation date information, and the estimates were used for modeling. However, we can expect more accurate results if we have complete symptom onset data. Second, for computing efficiency, we manually determined the period with intervention effects, T_1 and T_2 , but the period itself can be estimated through the statistical model. Lastly, imported cases from foreign countries were not considered in the model since we were focused on the local transmission within the community. However, except for the patients who were screened and quarantined immediately at the airport, the model can also contain a compartment of the individuals who spread out to the local community as they can be one main source of transmission.

For future research, we are planning to conduct spatial and spatio-temporal studies on the COVID-19 infections in South Korea. Since understanding and estimating spatial and spatio-temporal dynamics is essential in the prevention of the transmission of infectious diseases, the relevant studies would provide valuable information to control COVID-19.

Conclusions

To the best of our knowledge, this study is the first to model the COVID-19 infection trend and forecast future trends using a Bayesian statistical SIHR model with the preventive measures effects. With the assumption that the effects increase daily during a certain period, we found out that the transmission rate dropped to nearly one third of the initial transmission rate. Our findings show that the number infections in South Korea has been decreasing and non-pharmaceutical interventions are very important to curb transmission.

Abbreviations

COVID-19: Coronavirus disease 2019; SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; SIHR: Susceptible-infectious-hospitalized-removed; PPE: Personal protective equipment; MCMC: Markov Chain Monte Carlo; MSPE: Mean squared prediction error; DIC: Deviance information criterion; CI: Credible interval

Declarations

Ethics approval and consent to participate

No human or animal samples were included in the research presented in this article; therefore, ethical approval was not necessary.

Consent for publication

Not applicable

Availability of data and materials

The datasets supporting the conclusions of this article are available in the websites <http://www.cdc.go.kr/> and <http://gidcc.or.kr>.

Competing interests

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This work was supported by the research fund of the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education [grant number NRF-2018R1D1A1B07047712] and by the Government-wide R&D Fund project for infectious disease research (GFID), Republic of Korea [grant number HG18C0088]. The funders had no role in study design, data collection and analysis, interpretation of results, decision to publish, or preparation of the manuscript.

Authors' contributions

J.C. designed the study; U.L contributed to data acquisition; H.C., D.K., and J.C. carried out the statistical analysis; H.C., D.K., J.H.K., S.C., and J.C. drafted the manuscript. All authors contributed to the interpretation of data and revision of the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We thank all of those in the healthcare fields who were struggling to overcome the COVID-19 outbreak. This

study was performed under the research project titled “Research and Development on Integrated Surveillance System for Early Warning of Infectious Diseases (RISEWIDS).”

References

1. World Health Organization. Coronavirus disease 2019 (COVID-19): Situation Report - 73. 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200402-sitrep-73-covid-19.pdf?sfvrsn=5ae25bc7_6.
2. Korean Society of Infectious Diseases, Korean Society of Pediatric Infectious Diseases, Korean Society of Epidemiology, Korean Society for Antimicrobial Therapy, Korean Society for Healthcare associated Infection Control and Prevention, Korea Centers for Disease Control and Prevention. Report on the epidemiological features of coronavirus disease 2019 (COVID-19) outbreak in the Republic of Korea from January 19 to March 2, 2020. *J Korean Med Sci.* 2020;35:e112. <https://doi.org/10.3346/jkms.2020.35.e112>.
3. Choi S, Ki M. Estimating the reproductive number and the outbreak size of Novel Coronavirus disease (COVID-19) using mathematical model in Republic of Korea. *Epidemiol Health.* 2020;42.
4. Godio A, Pace F, Vergnano A. SEIR modeling of the Italian epidemic of SARS-CoV-2 using computational swarm intelligence. *Int J Environ Res Public Health.* 2020;17(10):3535. <https://doi.org/10.3390/ijerph17103535>.
5. Zhou X, Ma X, Hong N, Su L, Ma Y, He J, et al. Forecasting the worldwide spread of COVID-19 based on logistic model and SEIR model. *medRxiv.* 2020.
6. Hsiang S, Allen D, Annan-Phan S, Bell K, Bolliger I, Chong T, et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature.* 2020;1-9. <https://doi.org/10.1038/s41586-020-2404-8>.
7. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet.* 2020;395(10225):689-97. [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9).

8. Sun H, Qiu Y, Yan H, Huang Y, Zh Y, Chen SX. Tracking and predicting COVID-19 epidemic in ChinaMainland. medRxiv. 2020.
9. Ki M. Epidemiologic characteristics of early cases with 2019 novel coronavirus (2019-nCoV) disease in Korea. Epidemiol Health. 2020;42. <https://doi.org/10.4178/epih.e2020007>.
10. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. J R Stat Soc Series B Stat Methodol. 2002;64(4):583-639. <https://doi.org/10.1111/1467-9868.00353>.
11. Sturtz S, Ligges U, Gelman AE. R2WinBUGS: a package for running WinBUGS from R. 2005.

Figure Legends

Figure 1. Time-series plots of the number of confirmed and cumulative cases in South Korea.

Figure 2. Province-level choropleth map of the number of confirmed cases in South Korea. (This picture was produced using R version 3.6.3 <https://www.r-project.org/>.)

Figure 3. Flowchart of SIHR model.

Figure 4. The number of patients with symptoms using the competing model and proposed model.

Figure 5. The cumulative number of patients with symptoms the competing model and proposed model.