Data-driven analysis on the simulations of the spread of COVID-19 under different interventions of China

Ting Tian[1#], Jingwen Zhang[1#], Shiyun Lin[1#], Yukang Jiang[1#], Jianbin Tan[1#], Zhongfei Li[2], Xueqin Wang[3,1*]

[1]School of Mathematics, Sun Yat-sen University, Guangzhou, 510275, China

[2]School of Management, Sun Yat-sen University, Guangzhou, 510275, China

[3]School of Management, University of Science and Technology of China, Hefei, 230026, China

* Correspondence author: wangxq20@ustc.edu.cn

Dr. T. Tian, J. Zhang, S. Lin, Y. Jiang, and J. Tan contributed equally to this article.

Abstract

Objectives

Since February 2020, COVID-19 has spread rapidly to more than 200 countries in the world. During the pandemic, local governments in China implemented different interventions to efficiently control the spread of the epidemic. Characterizing transmission of COVID-19 under some typical interventions is to help countries develop appropriate interventions.

Methods

We established compartmental model that allowed the number of infected and infectious to be unknown and the effective reproduction number to change over time, thus the effects of policies could be reasonably reflected and estimated. By using the epidemic data of three representative cities of China (Wuhan, Wenzhou and Shenzhen), we migrated the estimated policy modes to other countries.

Results

The smallest expected cumulative confirmed cases under different interventions would be 5936 with 95% CI (5012,6966) for South Korea, 146012 with 95% CI (140504, 154264) for Italy, and 400642 with 95% CI (390331,409431) for the United States until May 31, 2020, respectively.

Conclusions

Based on the simulation of epidemic in South Korea, Italy and the United States, it is reasonable that South Korea and Italy continue to maintain their current policies, while the implementation of interventions of Wenzhou may significantly decrease the magnitude of the outbreak of COVID-19 for the United States.

Keywords: Compartmental model; Time-varying reproduction number; Infected individuals without isolation; Mode migration; Policy intervention

Introduction

Since December 2019, the first report of COVID-19 which was detected in Wuhan was sent to the World Health Organization (WHO) by Chinese government (World Health Organization, 2020a). The epidemic has rapidly spread to all provinces in China and more than 200 countries around the world. WHO declared that COVID-19 became a pandemic on March 11, 2020 (World Health Organization, 2020b). China issued a series of intervention policies to reduce the transmission of COVID-19, such as suspending public transportations in areas of severe outbreak (e.g., Wuhan), cancelling public gathering and delaying reopening of enterprises. These endeavors have led to the significant suppression of the epidemic of COVID-19. There was no domestically newly reported confirmed case for the first time in China on March 18, 2020 (Xinhuannet, 2020). The local and national responses of China have made

remarkable contributions to the prevention and control of COVID-19 at the expense of the economic loss. Taking individual consumption industry as an example, the amount of total retail sales of consumer goods decreased by about 20.5% with a sharp drop of catering services by 43.1% from January to February 2020 compared to the same duration of 2019 according to the National Bureau of Statistics in China (2020). Consequently, policymakers need to anticipate the likely outcome of interventions in terms of their own epidemic situations.

Predicting the scale of the epidemic is essential to inform policymakers the likely outcomes of potential intervention options (Lipsitch et al., 2011), i.e. the expected scale of cumulative confirmed cases for COVID-19 could provide insights into the likely outcomes of different intervention policies. The estimation of the time-varying reproduction numbers could indicate the effects of intervention policies (Cauchemez et al., 2006, Wallinga and Teunis, 2004). Yang et al. (2020) showed that if the implementation of intervention policies delayed 5 days, the expected scale of cumulative confirmed cases would be over 3 times larger than actual ones in China. In this study, researchers used epidemic data from Hubei province, Zhejiang province, and Guangdong province. We selected the most severe cities in these three provinces as our important examples, i.e. Wuhan, Wenzhou, and Shenzhen. In these cases, their time-varying reproduction numbers were used to unfold the effects of interventions, and thus to predict the expected scale of COVID-19 by using the real-time epidemic

data in various countries. In doing so, the possible courses of COVID-19 under these interventions in South Korea, Italy, and the United States can be made.

Wuhan was the first city of the COVID-19 outbreak with the most severe epidemic in China. It suspended public transportation, cancelled all outbound trains and flights (The prevention and control headquarters of Wuhan, 2020) and started to construct a specialist emergency hospital (i.e. Huoshenshan Hospital) for patients on January 23. All residential areas were closed on February 11, 2020. Wenzhou was the city with the largest number of cumulative confirmed cases in Zhejiang province. In Wenzhou, the inter-provincial and municipal bus services were suspended on January 27(Wenzhou Municipal People's Government, 2020a), 14 expressway entrances were temporarily closed from 10 p.m. onwards on January 29 (Wenzhou Municipal People's Government, 2020b), and on February 1, it announced that every household might send one person every two days outside for necessity purchases (Wenzhou Municipal People's Government, 2020c). Shenzhen was the city of the first confirmed case in Guangdong province, and its prevention and control strategy was adopted as a classical case in the report of WHO (2020). On January 24, all the entertainment places were closed. It announced that all the individuals returning from Hubei needed to be isolated at home or other suitable facilities for 14 days on February 2 (Shenzhen Municipal People's Government, 2020a), and the residential units with confirmed cases were forced to implement “hard quarantine” for 14 days and checkpoints were set up in all entrances of residential areas on February 7 (Shenzhen Municipal

People's Government, 2020b). All details about the interventions implemented in these cities were given in Figure 1.

In this paper, we used the time-varying reproduction number to reflect the effects of interventions implemented in Wuhan, Wenzhou, and Shenzhen, China. Our goal was to predict the likely outcomes of these interventions in South Korea, Italy and the United States so that the lessons can be learned from the key decisions made in the representative cities.

Methods

Data sources

The epidemic data were extracted for Wuhan, Wenzhou, and Shenzhen from the Chinese Center for Disease Control and Prevention (China CDC) (National Health Commission of the People's Republic of China, 2020), and the data for South Korea, Italy and the United States were downloaded through April 23, 2020 from COVID-19 Knowledge & Data Hub (Sciences, 2020).

Model assumptions and solutions

We introduced Susceptible-Infected and infectious without isolation-Hospitalized in isolation-Removed (SIHR) model with four compartments: Susceptible (S), infected and infectious without isolation (I), hospitalized in isolation (H), removed (R)(Hsieh et al., 2004, Li, 2018). Based on the definitions, we made the following assumptions:

(a) The susceptible (S) individuals have no immunity to the disease;

(b) The infected and infectious without isolation (I) were assumed to be infectious since they caught the virus, i.e., ignoring the latent period, and would transform into the hospitalized in isolation eventually after symptoms onset;

(c) The hospitalized (H) individuals were in isolation and treated carefully and would not transmit COVID-19 to the susceptible;

(d) The removed (R) individuals would not be infected again.

It is crucial to evaluate the number of individuals who are infected and infectious without isolation for the control of epidemic, which remain unknown and are needed to be estimated in SIHR model. Based on the above assumptions, the transfer relationships between S, I, H, R were defined (see Figure 1 in the supplements) and the corresponding dynamic system of these four compartments along with parameters α, β, γ were given in the supplements.

The basic reproduction number R_0, which is the expected number of secondary cases produced by a single (typical) infection in a completely susceptible population (Jones, 2007), is defined by SIHR model:

$$R_0 = \alpha/\beta,$$

where $1/\beta$ is the mean of the incubation period of COVID-19, and is assumed to be 5 days in terms of studies about COVID-19 (Backer et al., 2020, Guan et al., 2020, Li

et al., 2020) by ignoring the latent period. To obtain the time-varying reproduction number R_t, we assumed that the parameter α was time variant and would decrease by the influence of interventions. We used logistic function (Tan et al., 2020) to simulate the decreasing trend of the α:

$$\alpha_{\alpha_0,d,m}(t) = \frac{\alpha_0}{1+\exp\left(\lambda_m\left(t-d-\frac{m}{2}\right)\right)},$$

where α_0 denotes the maximum contact numbers per person of COVID-19 during the early outbreak, d is the time when the control measures start to be effective and α starts to decline, m represents the duration of a process where the epidemic is to nearly vanish, λ_m is chosen as $\frac{2\log\left((1-\varepsilon)/\varepsilon\right)}{m}$ and ε is fixed to be 0.01. The above parameters α_0, d, m are estimated from real epidemic data.

The smaller the values of d and m, the earlier effectiveness and the stronger intensity of interventions were implemented, respectively. The graph of time-varying α is given in Figure 2 in supplements.

By the setting above, R_t was calculated as:

$$R_t = \frac{\alpha_{\alpha_0,d,m}(t)}{\beta}.$$

We introduce randomness to SIHR model and applied the Bayesian method combined with Markov Chain Monte Carlo (MCMC) for the estimation of parameters and missing data (Bishop, 2006, Soubeyrand, 2016). Once the prior distribution of the parameters was given, we used Block Metropolis–Hastings algorithm (Andrieu et al., 2000, Andrieu et al., 2003) to simulate the posterior distribution of the parameters and the number of individuals in each class. The point estimates were presented as the

median of the posterior distribution while 95% credible intervals were constructed with 2.5% and 97.5% quantiles.

Results

The estimation of policy pattern

To describe the effects of interventions in different cities, we used the data from the starting dates of the COVID-19 outbreak to the dates of well-controlled in each representative cities individually, indicating that the time range for each representative city would be different (Green et al., 2002).

Among them, the starting date of the outbreak was defined as the day when the number of hospitalized individuals (H) started to increase and continued to increase for the next three consecutive days, while the date of well-controlled was at which the number of hospitalized individuals (H) started to decrease and continued to decrease for the next three consecutive days(Green et al., 2002). Based on these definitions, we used the data from January 21 to February 9 in Wenzhou and from January 19 to February 6 in Shenzhen. In Wuhan, the starting date was January 15, but the clinical standard was changed on February 12, we used February 11 as the data cutoff for Wuhan. The trends of R_t for representative cities were shown in Figure 2.

The initial value of R_t for Wuhan was greater than that for Shenzhen, followed by Wenzhou. The effectiveness of intervention policies for Shenzhen started earlier than that for Wuhan, followed by Wenzhou. The intensity of policies for Wenzhou

was stronger than that for Wuhan, followed by Shenzhen. Comparing with details about the interventions implemented in these cities in Figure 1, the effects of these policies could be reasonably reflected and estimated.

The mode migration

We applied the SIHR model whose R_t is time-invariant in South Korea, Italy and the United States. We used the real-time epidemic data from February 26 to February 28 in South Korea, from March 28 to March 30 in Italy and from March 29 to March 31 in the United States and estimated the posterior distributions country by country. And then we migrate the intensity of policy intervention of Wuhan, Wenzhou, and Shenzhen to South Korea, Italy, and the United States, and predict the posterior distributions of numbers of infected and infectious without isolation and cumulative confirmed cases. The simulation of COVID-19 trend for constant R_t and time-varying R_t of three policies of China in South Korea, Italy, and the United States are given as follows (Figures 3-5).

From our prediction, if interventions were not implemented since 23 February, the expected cumulative confirmed cases would be 7705971with corresponding 95% credible intervals (CI) (2419211,17969284) on March 7 when the expected cumulative confirmed cases were over 10% population for the first time in South Korea (Figure 3A). Similarly, the expected cumulative confirmed cases would be 6680363 with 95% CI (442494,35236491) on March 29 when the expected

cumulative confirmed cases were over 10% population for the first time in Italy (Figure 4A). And the expected cumulative confirmed cases will be 34179685 with 95% CI (5274919,136469414) on April 17 when the expected cumulative confirmed cases will over 10% population for the first time in the United States, if interventions were not implemented since March 13 (Figure 5A). Overall, if interventions were not implemented, it can be expected that how tremendously large populations would be infected in these countries. Thus, the interventions are critical to control the fast spread of the disease. In SIHR model, the estimated number of infected and infectious without isolation (I) could be used to estimate the stage of the epidemic, while the expected cumulative confirmed cases under different policy patterns reflected the magnitude of the outbreak of infectious diseases.

Based upon our prediction, the actual cumulative confirmed cases of COVID-19 in South Korea from February 26 to March 9 were within the ranges of expected cumulative confirmed cases for Shenzhen pattern, whose intensity was relatively mild compared with both Wuhan and Wenzhou patterns. According to the report (Normile, 2020), South Korea has implemented similar interventions to Shenzhen pattern, where it implemented policies closely following and isolating the close contact individuals of confirmed cases to reduce the transmission of COVID-19. Also, as its estimated number of I was 2639 with corresponding 95% CI (2306,3055) on February 26, there is not much difference for the expected cumulative confirmed cases in the simulation of three urban interventions in South Korea (Figure 3B). This indicated

that mild policy like Shenzhen pattern implemented at an early stage of epidemic could effectively curb the outbreak of infectious diseases. It is reasonable that South Korea would maintain its current intervention.

For Italy, there is a substantial difference in the simulation of three policies implemented, the expected cumulative confirmed cases will be 179571 with 95% CI (165868, 205749) under Wuhan pattern, 146012 with 95% CI (140504, 154264) under Wenzhou pattern and 214881 with 95% CI (187566, 265198) under Shenzhen pattern on May 31, respectively (Figure 4B). The likely outcomes of different interventions were distinct since the estimated number of I was large (25768) on March 28, which implies that it is cautious to carry out a policy at this stage of epidemic. Note that the expected number of cumulative confirmed cases under the interventions of Wenzhou for Italy would be the smallest. There were the most intense efforts in the interventions of Wenzhou for both suspending inner and outer public transportations (middle-distance prevention) and restricting residents out (short-distance prevention). It was required that such a large-scale response mounted to block the broad extent of transmission not only inside of city but also inside of communities.

Comparing the actual situations of Italy, the cumulative confirmed cases were 187,327 on April 23, 2020, which were 28.7% more than the simulation of COVID-19 under interventions of Wenzhou. If taking such highly stringent interventions, the epidemic situation could be mitigated where the expected number

of cumulative confirmed cases would be 146,012 on May 31, decreasing by 19% and 32% compared to the implementation of policy of Wuhan and Shenzhen, respectively.

In the United States, there is a considerable difference from the policy of Shenzhen compared with other two patterns. The expected cumulative confirmed cases will be 694055 with 95% CI (665937,725910) under Wuhan pattern, 400642 with 95% CI (390331,409431) under Wenzhou pattern and 1096227 with 95% CI (1037815,1172518) under Shenzhen pattern on May 31, respectively. One reason for this distinction is that the number of I was 83232 (Figure 5B) on March 29, which implied certain risks for long-term outbreak of epidemic. The cumulative confirmed cases were 849094 on April 23, 2020, and were 113.3% more than the simulation of COVID-19 under interventions of Wenzhou, which implied that the implementations of interventions of Wenzhou may significantly decrease the magnitude of the outbreak of COVID-19 for the United States. If taking the relatively mild interventions of Shenzhen, the situation would get worse and the expected cumulative confirmed cases would be over 1000000 in the United States on May 31.

Discussion

Our simulation results indicated that the mild interventions were difficult to control the magnitude of outbreak of infectious diseases when the number of infected and infectious without isolation (I) was quite large, but it was effective to implement at the early stage of epidemic. Once the epidemic became severe, the highly stringent

interventions were needed to be implemented for the control of epidemic. Although it still takes time to make a reasonable assessment of the patterns and effects of these policies, the transmission patterns of COVID-19 under some typical intervention measures could be learned from them, and the transmission trend of COVID-19 could be predicted.

In general, there are no one-size-fits-all interventions applicable for all regions (Wu et al., 2020). Taking any potential interventions depends on the specific setting of the region. We did not quantitatively measure the details of interventions and did not consider public health and economic costs of interventions in our analysis. However, the likely outcomes of interventions were predicted to inform the policymakers to consider the appropriate sets of proactive interventions.

Author contributions:

All authors attest that they meet the ICMJE criteria for authorship. TT and XW had roles in the study design, data collection, data analysis, data interpretation, literature search, and developed the idea and research. TT, JZ, SL, YJ and JT wrote the first draft of the manuscript and all other authors discussed results and edited the manuscript. JZ, SL, YJ and JT had roles in the data collection, data analysis, and ZL had roles in the data interpretation. TT, JZ, SL, YJ, and JT contributed equally to this article. All authors reviewed and approved the final version of the manuscript.

Conflict of interest

All authors declare no competing interests.

Acknowledgements

We would like to thank all individuals who are collecting epidemiological data of the COVID-19 outbreak around the world.

References

1. Andrieu C, de Freitas N, Doucet A. Sequential Bayesian estimation and model selection for dynamic kernel machines. 2000.
2. Andrieu C, De Freitas N, Doucet A, Jordan MI. An introduction to MCMC for machine learning. Machine learning 2003;50(1-2):5-43.
3. Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020. Euro Surveill 2020;25(5).
4. Bishop CM. Pattern recognition and machine learning: springer, 2006.
5. Cauchemez S, Boelle PY, Thomas G, Valleron AJ. Estimating in real time the

efficacy of measures to control emerging communicable diseases. Am J Epidemiol 2006;164(6):591-7.

6. Green MS, Swartz T, Mayshar E, Lev B, Leventhal A, Slater PE, et al. When is an epidemic an epidemic? The Israel Medical Association journal: IMAJ 2002;4(1):3-6.

7. Guan W-j, Ni Z-y, Hu Y, Liang W-h, Ou C-q, He J-x, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. New England Journal of Medicine 2020.

8. Hsieh Y-H, Lee J-Y, Chang H-L. SARS epidemiology modeling. Emerging infectious diseases 2004;10(6):1165.

9. Jones JH. Notes on R0. Califonia: Department of Anthropological Sciences 2007.

10. Li MY. An introduction to mathematical modeling of infectious diseases: Springer, 2018.

11. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. New England Journal of Medicine 2020.

12. Lipsitch M, Finelli L, Heffernan RT, Leung GM, Redd SC, Group HnS. Improving the evidence base for decision making during a pandemic: the example of 2009 influenza A/H1N1. Biosecur Bioterror 2011;9(2):89-115.

13. National Bureau of Statistics in China. National Economy Withstood the Impact of COVID-19 in the First Two Months; 2020. Available from: http://www.stats.gov.cn/english/PressRelease/202003/t20200316_1732244.html.

14. National Health Commission of the People's Republic of China. Distribution of COVID-19 cases in China; 2020. Available from: http://2019ncov.chinacdc.cn/2019-nCoV/index.html. [Accessed April 26, 2020].

15. Normile D. Coronavirus cases have dropped sharply in South Korea. What's the secret to its success. Science 2020.

16. World Health Organization. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). 2020.

17. Sciences CAo. COVID-19 Knowledge & Data Hub; 2020. Available from: http://geodoi.ac.cn/covid-19/en/index.aspx.

18. Shenzhen Municipal People's Government. Circular of the prevention and control headquarters of pneumonia epidemic caused by new coronavirus infection in Shenzhen; 2020a. Available from: http://www.sz.gov.cn/szzt2010/yqfk2020/szzxd/zczy/zcwj/fkzc/content/post_6725195.html.

19. Shenzhen Municipal People's Government. Shenzhen issued three epidemic notice! To the citizens living in shenzhen, to the personnel in shenzhen, the enterprises in shenzhen; 2020b. Available from: http://www.sz.gov.cn/cn/xxgk/zfxxgj/zwdt/content/post_6697629.html.

20. Soubeyrand S. Construction of semi-Markov genetic-space-time SEIR models and inference. 2016.

21. Tan J, Jinag Y, Tian T, Wang X. P-SIHR probabilistic graphical model: An

applicable model of COVID-19 in estimating the number of infectious individuals without isolation and time-varying reproduction number. Acta Mathematicae Applicatae Sinica 2020;42(3).

22. The prevention and control headquarters of Wuhan. Circular of the prevention and control headquarters of pneumonia epidemic caused by new cornoavirus infection in Wuhan (No.1); 2020. Available from: http://www.gov.cn/xinwen/2020-01/23/content_5471751.htm.

23. Wallinga J, Teunis P. Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures. American Journal of Epidemiology 2004;160(6):509-16.

24. Wenzhou Municipal People's Government. Notice of the Leading Group for the Prevention and Control of COVID-19 in Wenzhou (No. 3); 2020a. Available from: http://www.wenzhou.gov.cn/art/2020/1/28/art_1219304_41858870.html. [Accessed March 31, 2020].

25. Wenzhou Municipal People's Government. Notice of the Leading Group for the Prevention and Control of COVID-19 in Wenzhou (No. 4); 2020b. Available from: http://www.wenzhou.gov.cn/art/2020/1/30/art_1219304_41862456.html.

26. Wenzhou Municipal People's Government. Notice of the Leading Group for the Prevention and Control of COVID-19 in Wenzhou (No. 6); 2020c. Available from: http://www.wenzhou.gov.cn/art/2020/2/2/art_1219304_41867466.html. [Accessed March 31, 2020].

27. World Health Organization. Rolling updates on coronavirus disease (COVID-19); 2020a. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen. [Accessed April 20, 2020].

28. World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19-11 March 2020; 2020b. Available from: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

29. 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. The Lancet 2020.

30. Xinhuannet. No local newly confrimed cases for the first time in China; 2020. Available from: http://www.xinhuanet.com/politics/2020-03/18/c_1125730531.htm.

31. Yang Z, Zeng Z, Wang K, Wong S-S, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. Journal of Thoracic Disease 2020.

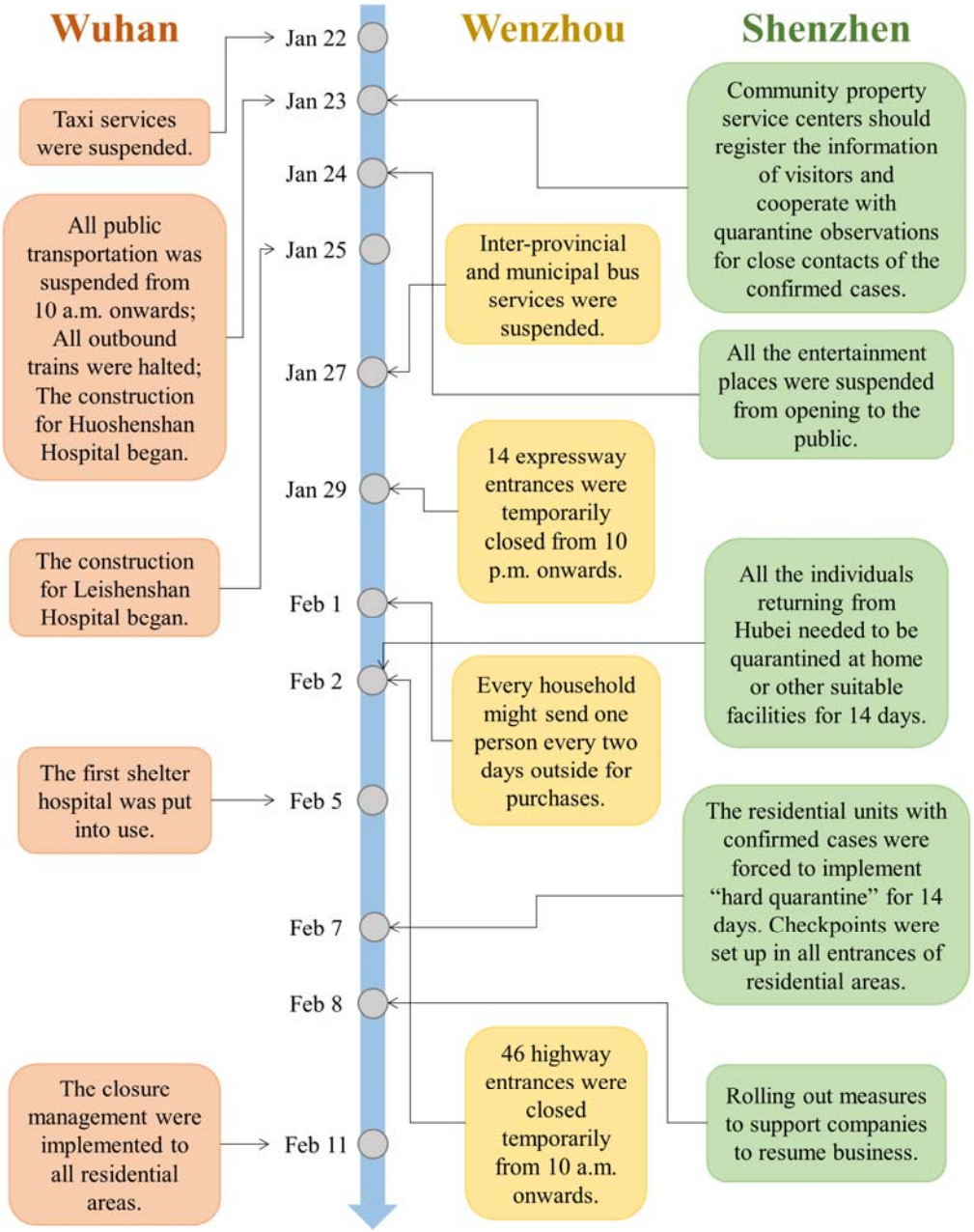

Figure 1. The timeline of interventions implemented in Wuhan, Wenzhou, and Shenzhen.

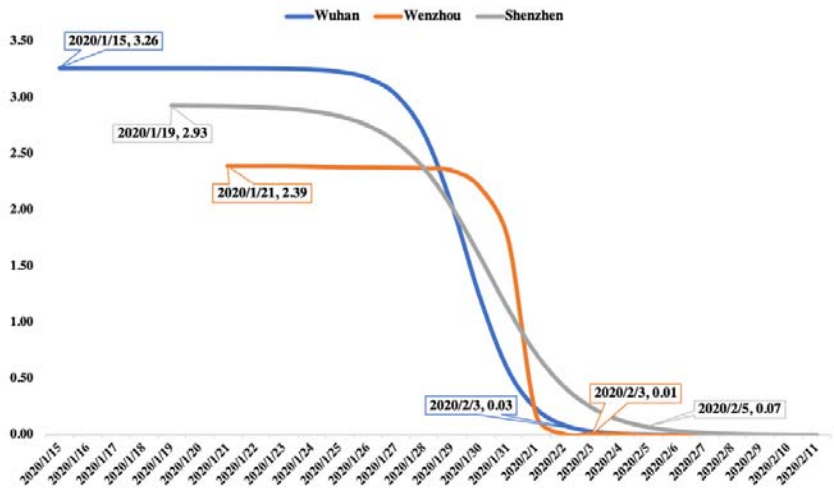

Figure 2. The trends of R_t for Wuhan, Wenzhou and Shenzhen.

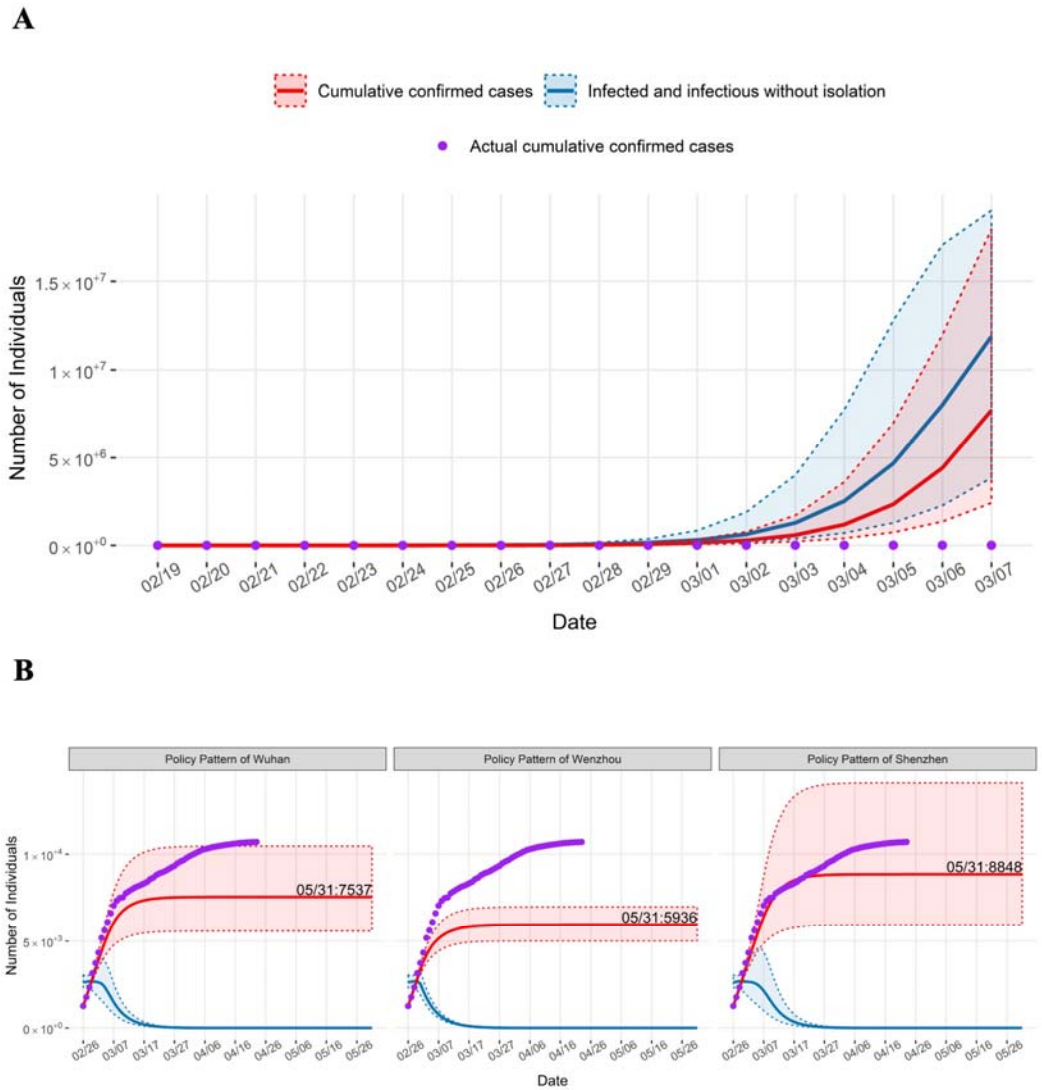

Figure 3. The simulation of COVID-19 trend for constant R_t (A) and time-varying R_t (B) of three policies in South Korea. The red curve with corresponding dashed lines represents the expected cumulative confirmed cases with 95% CI and the blue curve with corresponding dashed lines represent the expected number of infected and infectious without isolation (I) with 95% CI.

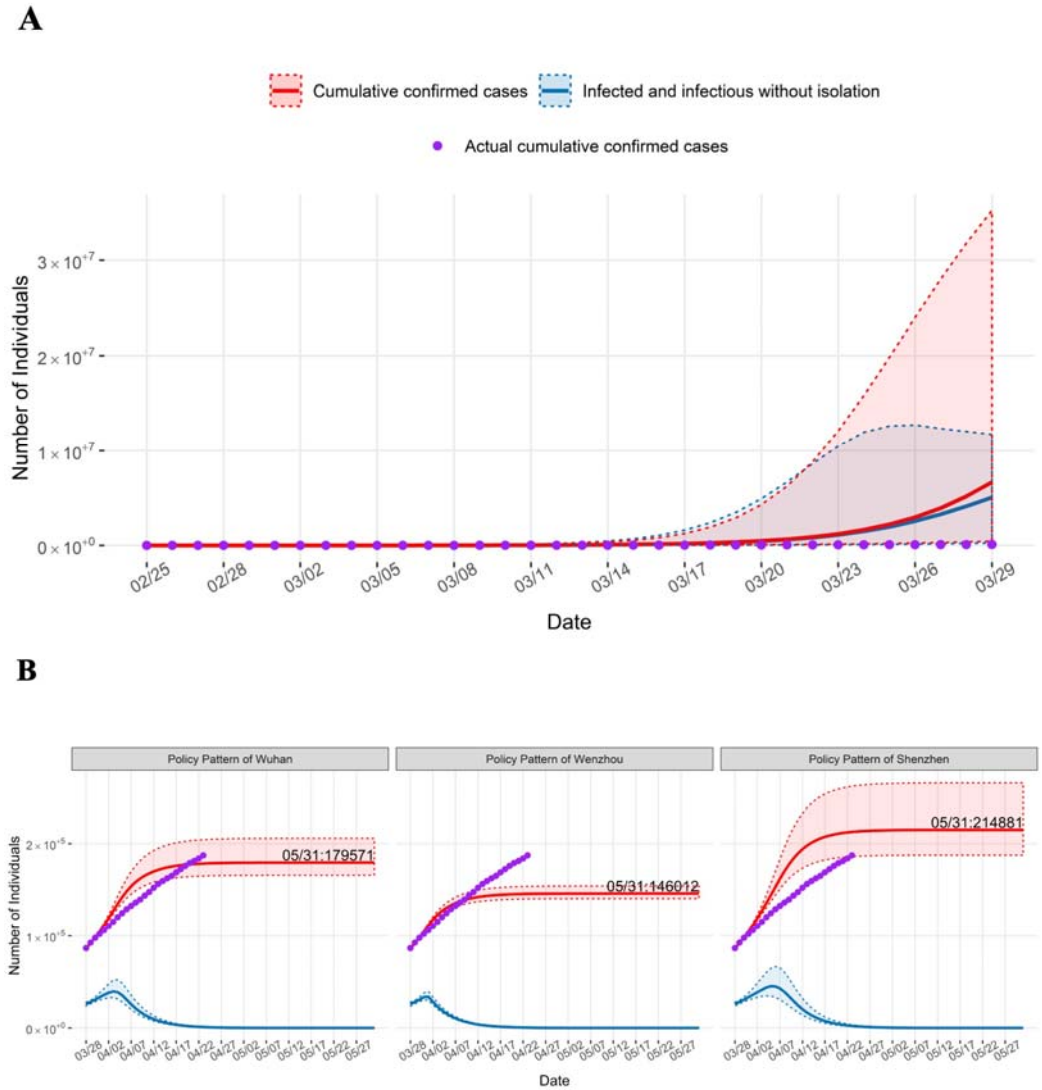

Figure 4. The simulation of COVID-19 trend for constant R_t (A) and time-varying R_t (B) of three policies in Italy. The red curve with corresponding dashed lines represents the expected cumulative confirmed cases with 95% CI and the blue curve with corresponding dashed lines represent the expected number of infected and infectious without isolation (I) with 95% CI.

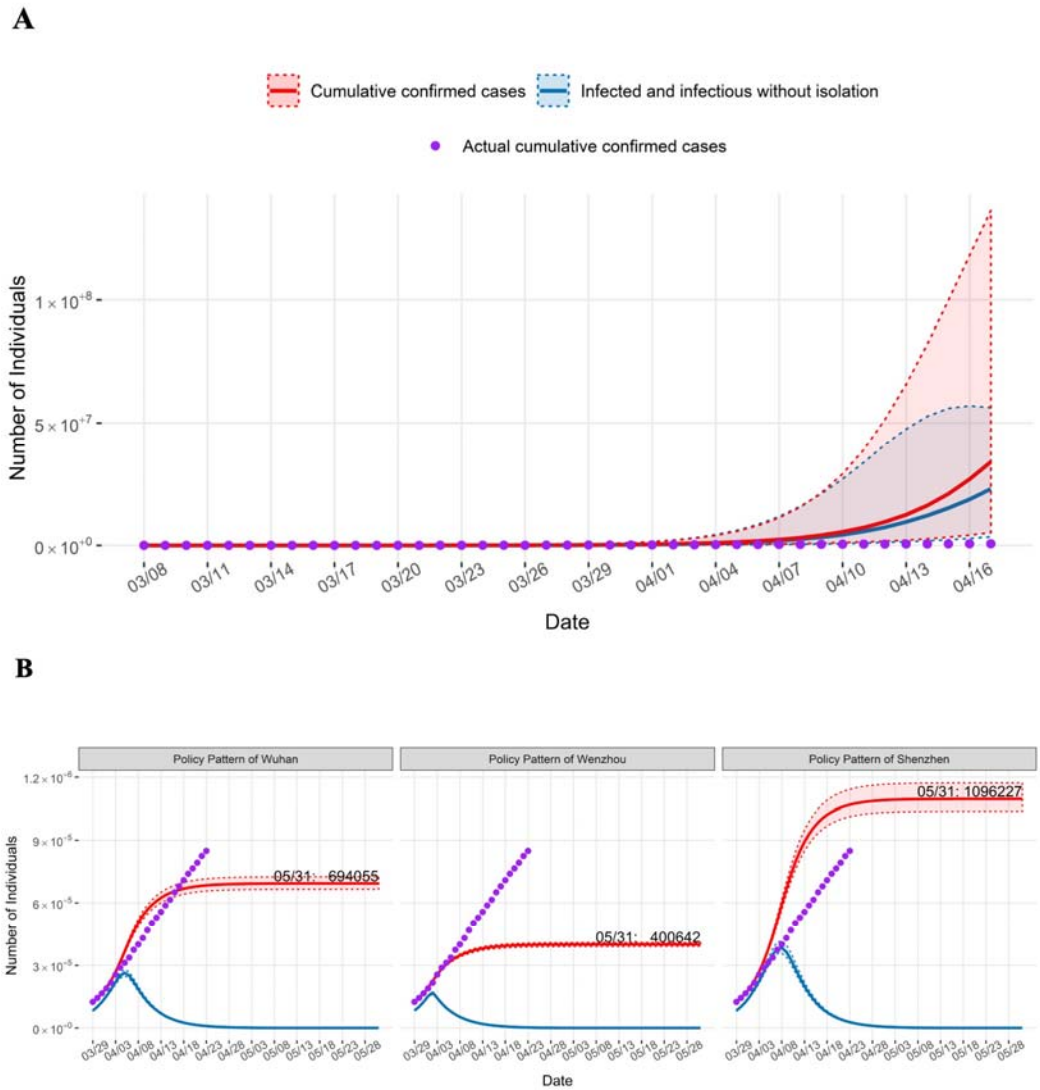

Figure 5. The simulation of COVID-19 trend for constant R_t (A) and time-varying R_t (B) of three policies in the United States. The red curve with corresponding dashed lines represents the expected cumulative confirmed cases with 95% CI and the blue curve with corresponding dashed lines represent the expected number of infected and infectious without isolation (I) with 95% CI.

Supplementary Information:

Stochastic SIHR model

COVID-19 prevalence

Model validity