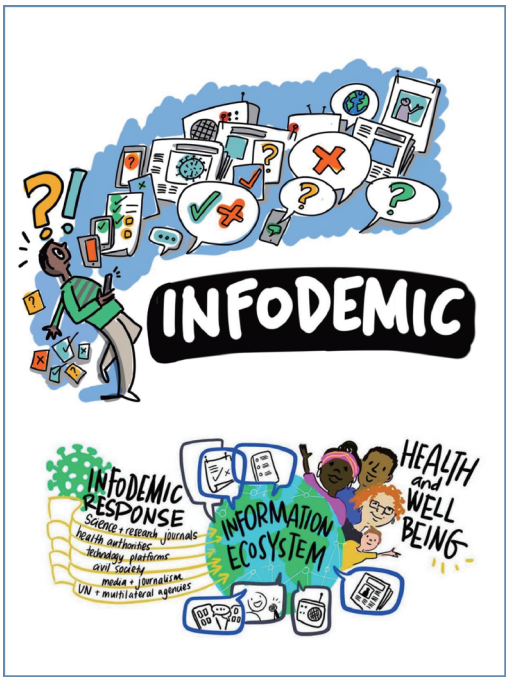
ISSN 2096-7071
CN 10-1629/R1

CHINA CDC WEEKLY

Vol. 4 No. 52 Dec. 30, 2022

中国疾病预防控制中心周报



COVID-19 ISSUE (32)

ISSN 2096-7071
9 772096 707224 52>

China CDC Weekly

Editorial Board

Editor-in-Chief Hongbing Shen

Founding Editor George F. Gao

Deputy Editor-in-Chief Liming Li Gabriel M Leung Zijian Feng

Executive Editor Feng Tan

Members of the Editorial Board

Rui Chen	Wen Chen	Xi Chen (USA)	Zhuo Chen (USA)
Gangqiang Ding	Xiaoping Dong	Pei Gao	Mengjie Han
Yuantao Hao	Na He	Yuping He	Guoqing Hu
Zhibin Hu	Yueqin Huang	Na Jia	Weihua Jia
Zhongwei Jia	Guangfu Jin	Xi Jin	Biao Kan
Haidong Kan	Ni Li	Qun Li	Ying Li
Zhenjun Li	Min Liu	Qiyong Liu	Xiangfeng Lu
Jun Lyu	Huilai Ma	Jiaqi Ma	Chen Mao
Xiaoping Miao	Ron Moolenaar (USA)	Daxin Ni	An Pan
Lance Rodewald (USA)	William W. Schluter (USA)	Yiming Shao	Xiaoming Shi
Yuelong Shu	RJ Simonds (USA)	Xuemei Su	Chengye Sun
Quanfu Sun	Xin Sun	Jinling Tang	Huaqing Wang
Hui Wang	Linhong Wang	Tong Wang	Guizhen Wu
Jing Wu	Xifeng Wu (USA)	Yongning Wu	Zunyou Wu
Min Xia	Ningshao Xia	Yankai Xia	Lin Xiao
Wenbo Xu	Hongyan Yao	Zundong Yin	Dianke Yu
Hongjie Yu	Shicheng Yu	Ben Zhang	Jun Zhang
Liubo Zhang	Wenhua Zhao	Yanlin Zhao	Xiaoying Zheng
Maigeng Zhou	Xiaonong Zhou	Guihua Zhuang	

Advisory Board

Director of the Advisory Board Jiang Lu

Vice-Director of the Advisory Board Yu Wang Jianjun Liu Jun Yan

Members of the Advisory Board

Chen Fu	Gauden Galea (Malta)	Dongfeng Gu	Qing Gu
Yan Guo	Ailan Li	Jiafa Liu	Peilong Liu
Yuanli Liu	Kai Lu	Roberta Ness (USA)	Guang Ning
Minghui Ren	Chen Wang	Hua Wang	Kean Wang
Xiaoqi Wang	Zijun Wang	Fan Wu	Xianping Wu
Jingjing Xi	Jianguo Xu	Gonghuan Yang	Tilahun Yilma (USA)
Guang Zeng	Xiaopeng Zeng	Yonghui Zhang	Bin Zou

Editorial Office

Directing Editor Feng Tan

Managing Editors	Lijie Zhang	Yu Chen	Peter Hao (USA)	
Senior Scientific Editors	Ning Wang	Ruotao Wang	Shicheng Yu	Qian Zhu
Scientific Editors	Weihong Chen	Xudong Li	Nankun Liu	Liwei Shi
	Liuying Tang	Meng Wang	Zhihui Wang	Xi Xu
	Qing Yue	Ying Zhang		

Cover Image: adapted from the World Health Organization, https://www.who.int/health-topics/infodemic#tab=tab_1.

Preplanned Studies

Longitudinal Participatory Surveillance Highlights Association Between Mask-Wearing and Lower COVID-19 Risk — United States, 2020

Makayla Swaciak[1,2]; Zachary Popp[2]; Autumn Gertz[1]; Kara Sewalk[1]; Marinanicole Schultheiss[1]; Benjamin Rader[1,2,#]; John S. Brownstein[1,3]

Summary

What is already known about this topic?

Numerous ecological and laboratory studies suggest face masks are an effective non-pharmaceutical intervention for reducing the spread of coronavirus disease 2019 (COVID-19), but cannot otherwise assess individual-level effects.

What is added by this report?

Using a prospective cohort of individuals enrolled in a participatory, syndromic surveillance tool prior to the first case of COVID-19 in the United States, we present a novel longitudinal assessment of the effectiveness of face masks.

What are the public health implications for public health practice?

Our analysis demonstrates an association between self-reported mask-wearing behavior and lower individual risk of syndromic COVID-19-like illness while adjusting for confounders at the individual level. Our results also highlight the dual utility of participatory syndromic surveillance systems as both disease trend monitors and tools that can aid in understanding the effectiveness of personal protective measures.

Mask-wearing is a non-pharmaceutical intervention that can successfully prevent the transmission of many respiratory illnesses, including coronavirus disease 2019 (COVID-19) (*1*). Throughout the COVID-19 pandemic, communities with higher proportions of self-reported mask-wearing showed better control of COVID-19 transmission (*2*) and an overall decline in COVID-19 case counts (*3*). While many ecological studies were able to support these results, they could not assess individual-level risk across all measures (*3–6*). For example, some studies have been able to collect individual self-reports of mask-wearing but can only examine their impact based on the community level COVID-19 rates that are available for use (*3*,*6*). One study was able to demonstrate an individual protective effect of mask-wearing on individual risk of COVID-19-like illness (CLI) but still used aggregated community levels for covariates such as social distancing in their analysis (*7*). Here, we collected both outcome and exposure at the individual level and used a Cox proportional hazards model to assess the effectiveness of mask-wearing on CLI incidence. Longitudinal weekly self-reports of respiratory symptoms were used to assess onset of CLI from January to June 2020 in a United States (U.S.) nationwide sample (n=4,723) of participants enrolled in FluNearYou (FNY), a web-based syndromic participatory surveillance platform, prior to the first case of COVID-19 in the U.S. Individual-level information on mask-wearing exposure and confounding variables was retrospectively gathered from an annual survey administered to FNY participants in June 2020. Overall, there were 1,310 reports of respiratory symptoms that met our primary illness definition over the study period and individuals characterized as most likely to wear masks were 44% (20%–61%) less likely to report CLI symptoms compared to individuals characterized as least likely to wear masks. Participatory surveillance systems, like FluNearYou, can provide valuable insights into the impact that personal protective measures have on individual risk of illness.

FNY is a participatory, syndromic surveillance tool that longitudinally tracks seasonal and pandemic influenza (*8–9*). To participate, FNY users submit weekly anonymous reports of any influenza-like symptoms they are experiencing through the online platform. Influenza data collected from the FNY population have historically complemented the trends observed in traditional surveillance systems, like U.S. CDC ILINET (*8*,*10*). At the end of each influenza season, a survey is administered to FNY participants to assess detailed health-related behaviors and attitudes

not queried in the weekly reports. Questions in the 2020 questionnaire evaluated mask-wearing, social distancing adoption, and other COVID-19 related attitudes and behaviors. Importantly, the 2020 survey assessed a time period before many U.S. states had implemented official mask mandates and vaccines against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) were not yet available (*4*).

The 2020 annual survey was fielded from June 18 through July 17, 2020 and received 8,018 responses from FNY users (56.2% response rate among those who answered the survey), of which there were 5,932 unique responses that could be matched to FNY weekly contributors. Longitudinal reports of respiratory illness symptoms from January 19 to June 30, 2020 were assessed in this sample of FNY participants who were enrolled in the program prior to the first confirmed case of COVID-19 reported in the U.S. on January 20, 2020 (*11*). Respondents who were reporting from outside the U.S., who reported symptoms meeting our primary illness definition in their initial survey, or who submitted fewer than 2 reports during the study period were not included in the analysis. Following the application of exclusion criteria, the resulting study cohort consisted of 4,809 unique users who submitted 87,231 weekly reports to FNY over the course of the study period. A total of 4,793 members of this cohort submitted surveys that contained answers to both of the questions used in the development of the mask-wearing likelihood variable (n=16, excluded). This study was approved by the Boston Children's Hospital Institutional Review Board and received a waiver of informed consent (IRB-P00023700).

We used a previously validated exposure variable to measure mask-wearing (*2*). In the annual survey, self-reported mask-wearing likelihood within the next week was ranked for two different settings: while grocery shopping and while visiting family and friends. For analysis, mask-wearing likelihood was categorized as "Very Likely" (very likely to wear a mask in both settings), "Somewhat Likely" (somewhat likely in one setting + very likely in the other), "Mixed Likelihood" (somewhat/very likely in one setting + not so likely/not at all likely in the other), or "Not Likely" (not so likely/not at all likely in both settings).

The primary outcome in our study was presence of CLI, defined as a self-report of shortness of breath, cough, or two or more of the following symptoms: fever, chills, sore throat, body ache, or headache (*12*). Sensitivity analyses were conducted that utilized broad and narrow definitions of respiratory illness.

Follow-up time began at the date of each participant's initial survey on or after January 19, 2020 and ended with either the final survey entry on or before June 30, 2020 or the first instance of reported symptoms meeting the respiratory illness definition. To assess the crude effect of mask-wearing on individual risk of CLI we estimated Kaplan-Meier survival curves for the four categories of mask-wearing. To adjust for factors confounding the effect of mask-wearing on individual risk of respiratory illness, we fit a Cox proportional hazards model. We controlled for the following confounding variables: age, gender, number of social distancing contacts (0–1, 1–4, 4–10, 10 or greater), date of social distancing adoption (Prior to March 1, March 1–14, March 15–31, April 1 or later), leaving home for work (binary), county population density (<500 people per square mile, ≥500 people per square mile), and time-varying county COVID-19 burden (population-adjusted incidence 1 week before survey entry). We modeled county COVID-19 burden as time-varying based on the location of each FNY weekly submission. For the regression analysis, participants who failed to provide a valid response for all model variables were excluded. Influential points, with deviance residuals greater than 3 and less than –3, were also excluded. All analyses were performed and figures were created using R software (version 4.0.2; R Core Team, Vienna, Austria).

Equation A: Cox model equation with time-varying COVID-19 burden variable

$$h(t) = \lambda_0(t) \times exp(\beta_1 gender + \beta_2 age + \beta_3 masking\ likelihood + \beta_4 number\ of\ contacts + \beta_5 social\ distancing\ adoption\ date + \beta_6 left\ home\ for\ work + \beta_7 county\ population\ density + \beta_8 county\ COVID\ burden(t))$$

The FNY population demonstrated high adherence to public health guidelines from January to June 2020, with the majority (52.2%) of individuals reporting they would be very likely to wear a mask both while grocery shopping and while visiting family and friends, and only 112 (2.3%) individuals reporting they would be not likely to wear masks in both scenarios. Self-reported mask-wearing was higher among women (69.7%) compared to men (30.3%). Those with 0–1 social distancing contacts were more likely to report wearing masks (62.1%) than those with more than 10 contacts (2.3%), and those who adopted social distancing in early to mid-March were more likely to report wearing masks (43.2%) compared to those who adopted social distancing in April or later (4.1%).

Individuals who went grocery shopping in the week prior to the questionnaire were more likely to report mask-wearing (67.1%) compared to those who did not go grocery shopping (32.9%). However, self-reported mask-wearing was lower among those who left home for work (15.7%) or to visit friends and family (15.8%). Individuals most likely to wear masks were more likely to live in counties with a population density greater than 500 people per square mile (76.5%) compared to counties with a population density of less than 500 people per square mile (22.5%). (Table 1)

TABLE 1. User demographics for the longitudinal participatory surveillance platform FluNearYou (FNY), by likelihood of mask-wearing (*n*=4,723).

Variable	Mask-wearing likelihood			
	Least likely (*n*=112)	**Mixed likelihood (*n*=1,052)**	**Somewhat likely (*n*=1,047)**	**Most likely (*n*=2,512)**
Gender, *n* (%)				
Female	64 (57.1)	662 (62.9)	731 (69.8)	1,750 (69.7)
Male	48 (42.9)	390 (37.1)	316 (30.2)	762 (30.3)
Age (years), *n* (%)				
13–48	13 (11.6)	134 (12.7)	131 (12.5)	314 (11.6)
49–64	41 (36.6)	360 (34.2)	346 (33.0)	824 (31.4)
≥65	57 (50.9)	530 (50.4)	539 (51.5)	1,289 (52.3)
Missing	1 (0.9)	28 (2.7)	31 (3.0)	85 (4.7)
Race, *n* (%)				
AIAN	3 (2.7)	11 (1.0)	11 (1.1)	36 (1.4)
Asian	0 (0.0)	10 (1.0)	22 (2.1)	63 (2.5)
Black	1 (0.9)	7 (0.7)	10 (1.0)	37 (1.5)
White	103 (92.0)	995 (94.6)	973 (92.9)	2,303 (91.7)
Other	5 (4.5)	29 (2.8)	31 (3.0)	73 (2.9)
Social distancing contacts*, *n* (%)				
0–1	39 (34.8)	482 (45.8)	544 (52.0)	1,559 (62.1)
1–4	31 (27.7)	409 (38.9)	358 (34.2)	725 (28.9)
4–10	18 (16.1)	101 (9.6)	94 (9.0)	129 (5.1)
≥10	17 (15.2)	44 (4.2)	27 (2.6)	58 (2.3)
Missing	7 (6.2)	16 (1.5)	24 (2.3)	41 (1.6)
Social distancing adoption date, *n* (%)				
Prior to March 1	6 (5.4)	58 (5.5)	53 (5.1)	212 (8.4)
March 1 to March 14	24 (21.4)	344 (32.7)	393 (37.5)	1,086 (43.2)
March 15 to March 31	47 (42.0)	529 (50.3)	479 (45.7)	953 (37.9)
April 1 or later	20 (17.9)	65 (6.2)	57 (5.4)	103 (4.1)
Missing	15 (13.4)	56 (5.3)	65 (6.2)	158 (6.3)
Left for work, *n* (%)	40 (35.7)	255 (24.2)	207 (19.8)	395 (15.7)
Left for grocery shopping, *n* (%)	98 (87.5)	846 (80.4)	786 (75.1)	1,686 (67.1)
Left to visit family and friends, *n* (%)	63 (56.2)	597 (56.7)	378 (36.1)	397 (15.8)
County population density <500 people per square mile, *n* (%)				
Yes	41 (36.6)	329 (31.3)	262 (25.0)	566 (22.5)
No	70 (62.5)	714 (67.9)	776 (74.1)	1,921 (76.5)
Missing	1 (0.9)	9 (0.9)	9 (0.9)	25 (1.0)

Abbreviation: AIAN=American Indian and Alaska Native.

* Respondents were asked how many people, including family members, coworkers etc. they saw in close contact for at least 15 minutes while practicing their strictest social distancing.

After excluding participants without valid responses for all covariates, 4,098 individuals remained in the cohort, with each individual contributing a median of 154 person-days of follow-up time [interquartile range (IQR)=69.25–161]. During the study period, 1,036 (25.3%) individuals reported symptoms meeting our definition of CLI.

Figure 1 demonstrates the unadjusted Kaplan-Meier plot for respiratory illness survival across the four categories of mask-wearing behavior. Individuals who were most likely to wear masks showed less risk of developing respiratory symptoms over the study period compared to those least likely to wear masks (*P*-value <0.001). At the end of the study period, 73.6% of individuals categorized as most likely to wear masks had not developed respiratory illness symptoms compared to only 61.4% of individuals categorized as least likely to wear masks. The difference in rates of symptom onset between categories of mask-wearing behavior increased over the study period.

The Cox model demonstrated a protective effect of mask-wearing (Figure 2). Table 2 provides the full Cox model for the primary analysis (Logrank test=96.56 on 14 df, *P*<0.0001). Individuals most likely to report mask-wearing were 44% (20%–61%) less likely to report respiratory symptoms over the study period compared to individuals least likely to report mask-wearing [hazard ratio (HR): 0.56, 95% confidence interval (*CI*): 0.39, 0.80, *P*=0.002]. Risk of CLI was also lower in the somewhat likely group (HR: 0.61, 95% *CI*: 0.42, 0.89, *P*=0.010) and in the mixed likelihood group (HR: 0.62, 95% *CI*: 0.43, 0.90, *P*=0.011) compared to the least likely group. Our sensitivity analyses using varying definitions of CLI found a similar magnitude and direction for the effect of mask-wearing on risk of CLI.

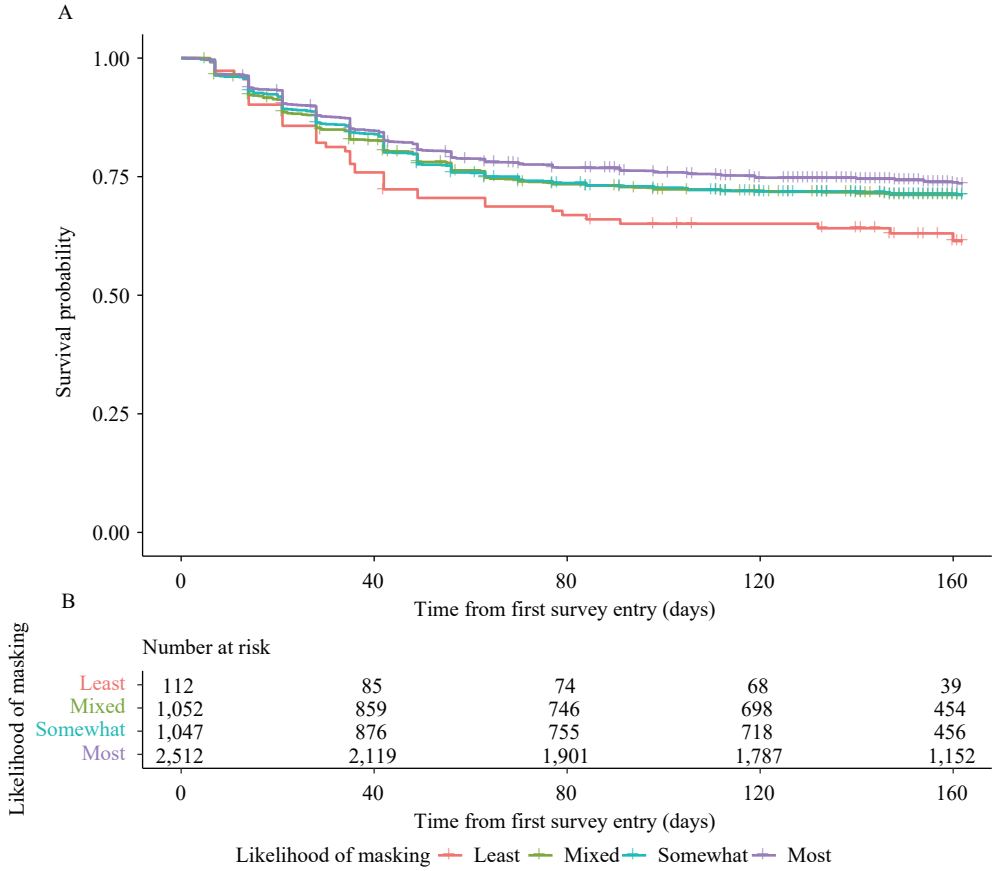


FIGURE 1. Unadjusted Kaplan-Meier Curve for mask-wearing with risk table. (A) Kaplan-Meier curves for respiratory illness symptoms; (B) Number at risk.
Note: Survival curves for users of the participatory surveillance platform FluNearYou (*n*=4,723) reported varying likelihoods of mask-wearing while grocery shopping and while visiting family and friends. Individuals reporting a higher likelihood of mask-wearing demonstrated a lower risk of respiratory illness symptoms across the study period compared to individuals reporting a lower likelihood of mask-wearing. The difference in rates of symptom onset between strata of mask-wearing likelihood increases over the study period.

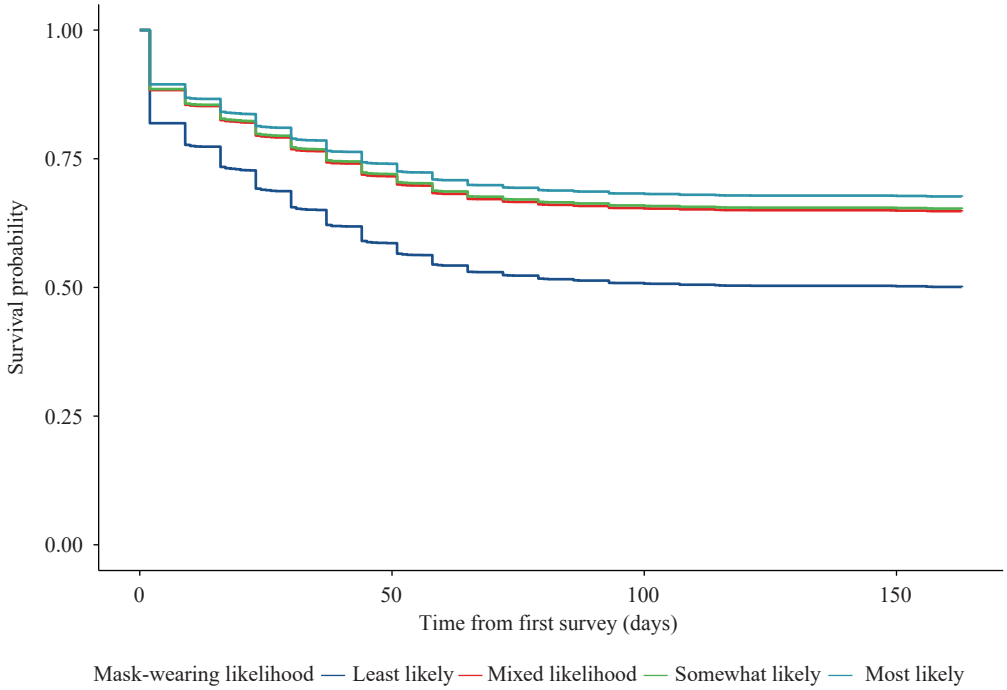


FIGURE 2. Cox-adjusted survival curves across levels of mask-wearing variable.
Note: Survival curves for respiratory illness over study period for users of the participatory surveillance platform FluNearYou (*n*=4,098) reporting varying likelihoods of mask-wearing while grocery shopping and while visiting family and friends. Results show a protective effect of mask-wearing on risk of developing respiratory illness over time, after adjusting for gender, age, number of social distancing contacts, date of social distancing, leaving home for work, home county population density, and estimated county COVID-19 burden. Individuals reporting a higher likelihood of mask-wearing demonstrated a lower risk of experiencing respiratory illness symptoms over the study period compared to individuals reporting a lower likelihood of mask-wearing. The difference in survival probability between strata of mask-wearing increases over time.

DISCUSSION

This longitudinal study of COVID-19 during the initial wave in the U.S. found that individuals who reported the greatest likelihood of mask-wearing outside of the household had a 44% reduced risk of respiratory illness compared to those least likely to wear masks during the same period. The protective effect of mask-wearing was robust after adjusting for age, gender, number of distancing contacts, date of distancing adoption, leaving home for work, home county population density, and COVID-19 burden as well as under varying definitions of respiratory illness.

The individual protective effect of mask-wearing supports ecological and randomized study findings that mask use reduces community COVID-19 cases (*2*–*3*,*6*). Previous studies have suggested that individuals who are more likely to wear masks may also be more likely to live in communities with high mask adherence due to the general influence of social norms on individual behavior (*6*). The individual protective effect we observed may reflect reduced transmission resulting from community mask-wearing behavior as well as individual mask-wearing, or other precautions taken by individuals, such as improved hand hygiene.

There are several other limitations to consider when interpreting the results of this study. The FNY population is a predominantly white, largely health-conscious group, and overrepresents retired individuals, meaning our results may not generalize to the entire U.S. population. Furthermore, our assessment of the effectiveness of mask-wearing is limited, as we do not have specific data for when individuals began to use masks or what type of masks they were using. We addressed this issue by adjusting for the date of initial self-reported social distancing. Several U.S. mask mandates were implemented much later than March 2020 (*4*), which is when the majority of our study population began social distancing, potentially leading to an underestimation of the true effect. It is also possible that people who experienced CLI in the follow-up period would have been less likely to report mask-wearing in June 2020 due to assumed protection through acquired immunity, causing an

TABLE 2. Model output for Cox proportional hazards regression for masks only and for mask-wearing and all covariates (*n*=4,098).

Variable	Model 1 (mask only), HR (95% *CI*)	Model 2 (full model), HR (95% *CI*)
Gender		
Female		REF
Male		0.85 (0.74, 0.97)
Age (years)		
<64		REF
65–		0.93 (0.82, 1.05)
Mask-wearing likelihood		
Least likely	REF	REF
Mixed likelihood	0.72 (0.52, 1.00)	0.62 (0.43, 0.90)
Somewhat likely	0.73 (0.52, 1.00)	0.61 (0.42, 0.89)
Most likely	0.65 (0.47, 0.89)	0.56 (0.39, 0.80)
Social distancing contacts		
0–1		REF
1–4		1.50 (1.31, 1.71)
4–10		1.58 (1.26, 1.98)
10 or greater		1.00 (0.67, 1.51)
Social distancing adoption date		
Prior to March 1		REF
March 1 to March 14		1.03 (0.81, 1.30)
March 15 to March 31		0.83 (0.65, 1.06)
April 1 or later		0.57 (0.39, 0.83)
Left for work		
Did not go into work		REF
Went into work		1.24 (1.07, 1.44)
County COVID-19 burden		0.98 (0.97, 0.99)
County population density <500 people per square mile		
Yes		REF
No		0.95 (0.83, 1.10)

Abbreviation: HR=hazard ratio; *CI*=confidence interval; REF=reference group; COVID-19=coronavirus disease 2019.

overestimation of the true effect.

Retrospective self-reporting of the exposure and outcome may increase the likelihood of recall bias, misclassification, and a biased effect estimate. Additionally, the use of syndromic surveillance omits asymptomatic COVID-19 infection, while inadvertently assessing infection with rhinoviruses, influenza, or other respiratory illnesses circulating in the population (*13*). The robustness of our results to both narrow and broad definitions of CLI illustrates that mask-wearing appears effective against a wide range of respiratory illnesses that were circulating in 2020, including COVID-19.

Our analysis of mask-wearing behavior in a previously enrolled symptom surveillance cohort demonstrates an association between self-reported mask-wearing behavior and a lower risk of respiratory illness at the beginning of the COVID-19 pandemic in the U.S. Our results support previous findings on the effectiveness of face masks and are especially salient given our ability to control for individual-level covariates. This study provides additional evidence in support of recommendations for mask-wearing to prevent respiratory illness. Participatory syndromic surveillance cohorts like FNY provide a valuable resource for understanding the impact of personal protective measures on COVID-19.

Conflicts of interest: No conflicts of interest.

Funding: Supported by the U.S. Centers for Disease Control and Prevention (ID 75D30121C11606).

doi: 10.46234/ccdcw2022.235

[#] Corresponding author: Benjamin Rader, benjamin.rader@childrens.harvard.edu.

[1] Computational Epidemiology Lab, Boston Children's Hospital, Boston, MA, USA; [2] Boston University School of Public Health, Boston University, Boston, MA, USA; [3] Harvard Medical School, Harvard University, Cambridge, MA, USA.

Submitted: October 29, 2022; Accepted: December 22, 2022

REFERENCES

1. Leung NHL, Chu DKW, Shiu EYC, Chan KH, McDevitt JJ, Hau BJP, et al. Respiratory virus shedding in exhaled breath and efficacy of face masks. Nat Med 2020;26(5):676 – 80. http://dx.doi.org/10.1038/s41591-020-0843-2.
2. Rader B, White LF, Burns MR, Chen J, Brilliant J, Cohen J, et al. Mask-wearing and control of SARS-CoV-2 transmission in the USA: a cross-sectional study. Lancet Digit Health 2021;3(3):E148 – 57. http://dx.doi.org/10.1016/S2589-7500(20)30293-4.
3. Aravindakshan A, Boehnke J, Gholami E, Nayak A. The impact of mask-wearing in mitigating the spread of COVID-19 during the early phases of the pandemic. PLoS Glob Public Health 2021;2(9):e0000954. http://dx.doi.org/10.1371/journal.pgph.0000954.
4. Gallaway MS, Rigler J, Robinson S, Herrick K, Livar E, Komatsu KK, et al. Trends in COVID-19 incidence after implementation of mitigation measures — Arizona, January 22–August 7, 2020. MMWR Morb Mortal Wkly Rep 2020;69(40):1460 – 3. http://dx.doi.org/10.15585/mmwr.mm6940e3externalicon.
5. Lyu W, Wehby GL. Community use of face masks and COVID-19: evidence from a natural experiment of state mandates in the US. Health Aff (Millwood) 2020;39(8):1419 – 25. http://dx.doi.org/10.1377/hlthaff.2020.00818.
6. Leech G, Rogers-Smith C, Monrad JT, Sandbrink JB, Snodin B, Zinkov R, et al. Mask wearing in community settings reduces SARS-CoV-2 transmission. Proc Natl Acad Sci USA 2022;119(23):e2119266119. http://dx.doi.org/10.1073/pnas.2119266119.
7. Kwon S, Joshi AD, Lo CH, Drew DA, Nguyen LH, Guo CG, et al. Association of social distancing and masking with risk of COVID-19. Nat Commun 12, 3737 (2021). http://dx.doi.org/10.1038/s41467-021-24115-7.
8. Smolinski MS, Crawley AW, Baltrusaitis K, Chunara R, Olsen JM, Wójcik O, et al. Flu near you: crowdsourced symptom reporting spanning 2 influenza seasons. Am J Public Health 2015;105(10):2124 – 30. http://dx.doi.org/10.2105/AJPH.2015.302696.
9. Chan AT, Brownstein JS. Putting the public back in public health — surveying symptoms of Covid-19. N Engl J Med 2020;383(7):e45. http://dx.doi.org/10.1056/NEJMp2016259.
10. Baltrusaitis K, Santillana M, Crawley AW, Chunara R, Smolinski M, Brownstein JS. Determinants of participants' follow-up and characterization of representativeness in flu near you, a participatory disease surveillance system. JMIR Public Health Surveill 2017;3(2):e18. http://dx.doi.org/10.2196/publichealth.7304.
11. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First case of 2019 novel coronavirus in the United States. N Engl J Med 2020;382(10):929 – 36. http://dx.doi.org/10.1056/NEJMoa2001191.
12. CDC. Coronavirus disease 2019 (COVID-19) 2020 interim case definition, approved April 5, 2020. 2021. https://ndc.services.cdc.gov/case-definitions/coronavirus-disease-2019-2020/. [200-6-24].
13. Maharaj AS, Parker J, Hopkins JP, Gournis E, Bogoch II, Rader B, et al. The effect of seasonal respiratory virus transmission on syndromic surveillance for COVID-19 in Ontario, Canada. Lancet Infect Dis 2021;21(5):593 – 4. http://dx.doi.org/10.1016/S1473-3099(21)00151-1.

Preplanned Studies

Hospital Strain and COVID-19 Fatality — England, April 2020–March 2022

Tengfei Lin[1]; Ziyi Zhao[1]; Zhirong Yang[1]; Bingli Li[1]; Chang Wei[1]; Fuxiao Li[1]; Yiwen Jiang[1]; Di Liu[1]; Zuyao Yang[2,#]; Feng Sha[1,#]; Jinling Tang[1,3]

Summary

What is already known about this topic?

During the coronavirus disease 2019 (COVID-19) pandemic, tremendous efforts have been made in countries to suppress epidemic peaks and strengthen hospital services to avoid hospital strain and ultimately reduce the risk of death from COVID-19. However, there is limited empirical evidence that hospital strain increases COVID-19 deaths.

What is added by this report?

We found the risk of death from COVID-19 was linearly associated with the number of patients currently in hospitals, a measure of hospital strain, before the Omicron period. This risk could be increased by a maximum of 188.0%.

What are the implications for public health practice?

These findings suggest that any (additional) effort to reduce hospital strain would be beneficial during early large COVID-19 outbreaks and possibly also others alike. During an Omicron outbreak, vigilance remains necessary to prevent excess deaths caused by hospital strain as happened in Hong Kong Special Administrative Region, China.

Hospital capacity strain arises when the demand for care resources exceeds supply in hospitals. During the coronavirus disease 2019 (COVID-19) pandemic, tremendous efforts have been made in countries and regions to prevent hospital strain, but there is limited empirical evidence that hospital strain indeed increases COVID-19 deaths. Although a few small studies showed that shortage in intensive care was associated with an increased COVID-19 fatality (*1*–*3*), strain may occur in many areas (i.e., hospital beds, drugs, devices, and staff) in the entire healthcare system besides intensive care and they may all add up to increase the risk of death from COVID-19. As any new COVID-19 cases add service demand to normal healthcare capacities, the number of COVID-19 patients currently in hospitals (PIH) can be viewed as an approximate measure of the strain on the healthcare system. Therefore, we conducted this analysis of surveillance data and used the number of PIH as a measure of hospital strain to examine its effects on the risk of death from COVID-19 using data from England before March 11, 2022. We found that the risk of death from COVID-19 was linearly associated with the number of PIH before the Omicron period and could be increased due to hospital strain by a maximum of 188.0%. This suggests that any (additional) effort to reduce hospital strain would be beneficial during early large COVID-19 outbreaks and possibly also others alike. During an Omicron outbreak, vigilance remains necessary to prevent excess deaths caused by hospital strain as happened in Hong Kong Special Administrative Region (SAR), China.

This is an analysis of surveillance data on all 147,276 COVID-19 deaths and 601,084 hospitalized COVID-19 patients in England during the period between April 9, 2020 and March 11, 2022 extracted on a daily basis from the UK Health Security Agency (*4*). The daily number of COVID-19 PIH was used as a measure of hospital strain, and daily case fatality was expressed as the ratio of the daily number of deaths from COVID-19 to the daily number of COVID-19 PIH and used as a measure of the risk of death from COVID-19. The study was divided into four periods, i.e., the wild, Alpha, Delta, and Omicron waves. The hospital strain-fatality relation in the four different periods was presented separately with a scatter plot and compared using log-linear regressions, controlling for potential confounders including proxy indicators for vaccination effect, severity of illness, error in the number of deaths, variant of the virus, improvement in hospital care, and other factors that changed over the study period. All statistical analyses were performed using R (version 4.1.0, R Development Core Team, Vienna, Austria). Details on the methods are provided in Supplementary Materials (available in http://weekly.chinacdc.cn).

Summary data including the duration of the study period, the total and median daily number of new cases and deaths, median percentage of the population vaccinated with one, two and three doses, and median daily case fatality are presented according to the four periods of the epidemic in Table 1. Notably, the median daily number of new cases increased steadily from 1,425 cases per day in period 1 to 62,303 cases per day in period 4 (mostly Omicron), a 43.7-fold increase. However, the daily number of PIH did not increase proportionally to the daily new cases diagnosed and showed a maximum of only 1.9-fold difference in the 4 periods. The median daily number of deaths was highest during period 1, resulting in a declining daily case fatality during the 4 periods from the highest 3.4% in period 1 to the lowest 1.2% in period 4, a 64.7% decrease (P=0.0137). The decline in fatality could only partially be explained by vaccination, as there were no or only a few people who completed two doses of vaccines during the first two study periods.

The 7-day moving average of daily number of new cases, PIH, and deaths, daily case fatality during the 4 periods of the epidemic in relation to the progress of vaccination and changes in lockdowns, public health measures, and variants of the virus were shown graphically in Supplementary Figure S1 (available in http://weekly.chinacdc.cn). Patterns similar to those observations described above can be visually observed regarding daily new cases, PIH, deaths, and fatality.

Importantly, the association between the daily case fatality and number of PIH, a measure of hospital strain in this study, according to the 4 periods of the epidemic is shown in Figure 1. In periods 1, 2, and 3, the fatality was positively and linearly associated with the number of PIH with a correlation coefficient of 0.95, 0.55, and 0.58, respectively (all P values <0.0001). In period 4, the fatality was sharply divided into two parts. The first part was mostly Delta and the second was predominantly Omicron, in which the fatality was the lowest and remained stable regardless of the variations in the number of PIH. The same conclusions can be drawn when patients currently in ventilation beds were used as a measure of hospital

TABLE 1. The total number of new cases and death events, average of daily new cases, death events, cases in hospitals, percentage of first, second, and third dose of vaccination, and daily case fatality during the 4 periods of the coronavirus disease 2019 epidemic in England between April 9, 2020, and March 11, 2022.

Variable	Four periods of the epidemic				Total (09/04/2020–11/03/2022)
	Period 1 (09/04/2020–11/07/2020)	Period 2 (12/07/2020–30/04/2021)	Period 3 (01/05/2021–26/11/2021)	Period 4 (27/11/2021–11/03/2022)	
No. of days in the period	94	293	210	105	702
Total No. of new cases, (%)	186,230 (1.1)	3,696,656 (22.4)	4,855,376 (29.4)	7,760,057 (47.0)	16,498,319 (100.0)
Total No. of deaths, (%)	36,335 (24.7)	83,400 (56.6)	14,554 (9.9)	12,987 (8.8)	147,276 (100.0)
Total No. of hospital admission, (%)	68,382 (11.4)	287,440 (47.8)	111,316 (18.5)	133,946 (22.3)	601,084 (100.0)
Median (quartiles) No. of new cases	1,425 (760, 3,125)	7,780 (2,224, 18,197)	25,598 (13,505, 31,601)	62,303 (39,223, 93,784)	15,156 (2,752, 31,327)
Median (quartiles) No. of deaths	240 (91, 583)	124 (28, 413)	78 (18, 106)	111 (94, 156)	102 (39, 224)
Median (quartiles) of the percentage of deaths outside hospitals (%)*	43.4 (41.2, 46.6)	29.5 (25.2, 34.2)	20.5 (17.9, 24.3)	30.2 (21.8, 32.3)	28.9 (21.7, 35.1)
Median (quartiles) No. of hospital admission	562 (300, 1,058)	528 (135, 1,467)	636 (196, 751)	1,220 (926, 1,604)	696 (218, 1,232)
Median (quartiles) No. of patients in hospital	7,360 (3,717, 12,265)	5,976 (1,378, 14,411)	5,032 (1,274, 6,066)	9,369 (7,114, 13,331)	5,996 (2,117, 11,585)
Median (quartiles) No. of patients in ventilation beds	764 (299, 1,832)	682 (178, 1,339)	683 (228, 810)	608 (355, 770)	674 (247, 910)
Percentage of population vaccinated (%)					
Median (quartiles) completed 1st dose	0.0 (0.0, 0.0)	0.0 (0.0, 27.7)	82.0 (75.3, 85.0)	90.6 (89.7, 91.3)	52.0 (0.0, 84.2)
Median (quartiles) completed 2nd dose	0.0 (0.0, 0.0)	0.0 (0.0, 1.0)	69.8 (55.4, 78.0)	83.3 (82.0, 84.6)	4.8 (0.0, 76.8)
Median (quartiles) completed 3rd dose	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.0 (0.0, 3.2)	63.3 (56.1, 65.2)	0.0 (0.0, 0.0)
Median (quartiles) daily case fatality (%)	3.4 (2.6, 4.9)	2.7 (2.1, 3.0)	1.6 (1.4, 1.8)	1.2 (1.0, 1.4)	1.9 (1.5, 2.8)

* Data were available on a weekly basis.

strain (Supplementary Figure S2, available in http://weekly.chinacdc.cn). After adjusting for vaccination score, admission rate, percentage of deaths outside hospitals, study period, and interaction term between PIH and study period, hospital strain remained statistically significantly associated with daily case fatality in the first 3 periods (all *P* values <0.0001 in study periods 1, 2, and 3, respectively) (Table 2, Supplementary Table S1, available in http://weekly.chinacdc.cn).

Finally, as the daily number of PIH increased from the lowest to the highest, the actual (or unadjusted) daily case fatality increased by 188.0% [95% confidence interval (*CI*): 165.9%, 211.6%], 69.9% (95% *CI*: 59.0%, 81.8%), and 58.2% (95% *CI*: 35.4%, 89.0%), respectively, in study periods 1, 2, and 3 (Supplementary Table S2, available in http://weekly.chinacdc.cn).

Results of additional analyses, including sensitivity analyses, are presented in the Supplementary Figures

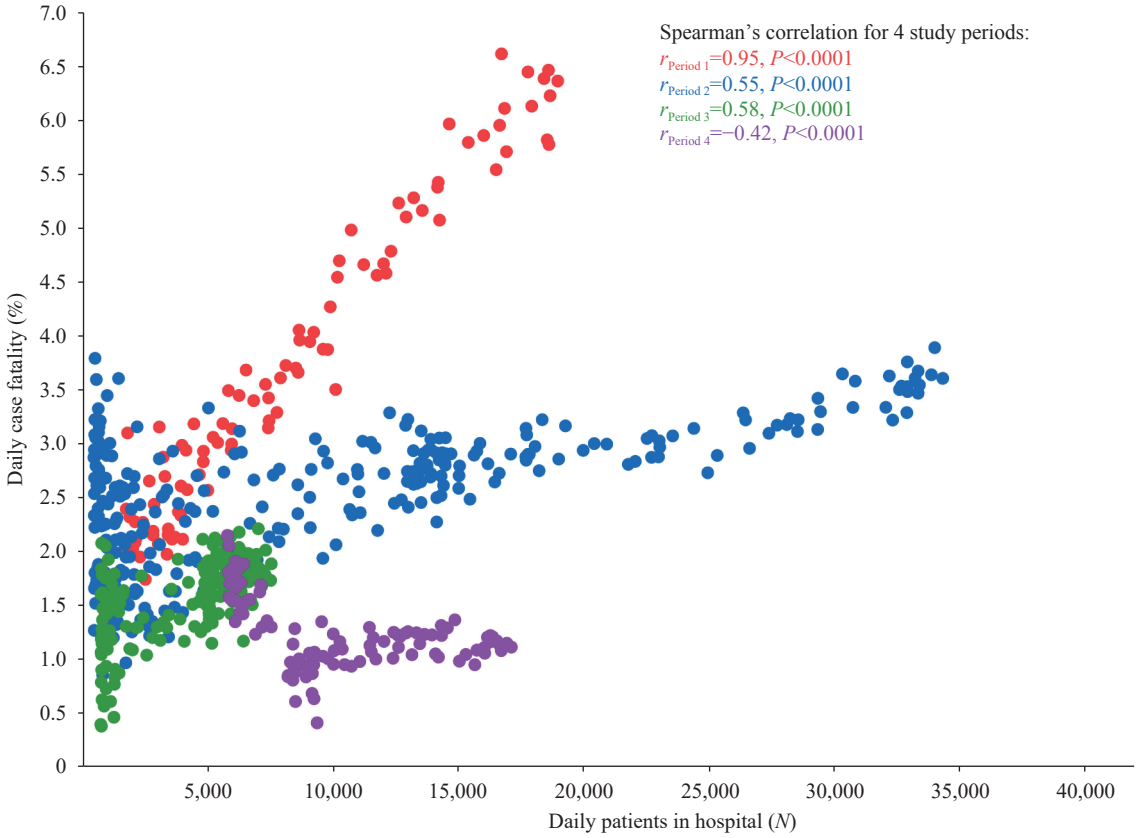


FIGURE 1. Scatter plot and Spearman's correlation between daily case fatality and daily number of patients in hospitals according to the 4 periods of epidemic in England between April 9, 2020, and March 11, 2022.

TABLE 2. Relative increase in daily case fatality for a 1,000-increase in daily number of patients in hospitals according to study period and adjusted for potential confounders.

Study period	Regression coefficients (95% *CI*)*	Relative increase (95% *CI*)*	*P*-value
Period 1	0.062 (0.057, 0.066)	1.063 (1.059, 1.068)	<0.0001
Period 2	0.014 (0.013, 0.015)	1.014 (1.013, 1.015)	<0.0001
Period 3	0.120 (0.102, 0.137)	1.127 (1.108, 1.147)	<0.0001
Period 4	−0.005 (−0.014, 0.003)	0.995 (0.986, 1.003)	0.2306

Abbreviation: *CI*=confidence interval.

* Log-linear multivariable regression was used to estimate the regression coefficients (i.e., the effect of daily number of patients in hospital on daily case fatality) and relative increase (exponential of the regression coefficient) for each period. The regression coefficients and relative increases were adjusted for vaccination score, study period, interaction term between daily number of patients in hospital and study period, admission rate, and percentage of deaths outside hospitals, and weighted by the number of patients in hospital (detailed results in Supplementary Table S1, available in https://weekly.chinacdc.cn/).

S3–S4 (available in http://weekly.chinacdc.cn), and related methods were described in Supplementary Materials.

DISCUSSION

By using authoritative English national data over 2 years of the epidemic, we found that the daily number of COVID-19 PIH as an indicator of overall hospital strain caused by the epidemic was linearly associated with the risk of death from COVID-19, except in the Omicron period, which confirmed findings from several previous studies (*1*–*3*). The largest difference in the risk of death from COVID-19 observed during an outbreak in England was 2.88-fold, suggesting that a maximum of 65.3% death risk reduction could be achieved theoretically by reducing COVID-19 PIH. However, the linear relation suggests that any (additional) effort to reduce COVID-19 PIH is related to a reduction in the risk of death and is worthwhile regardless of the total number of hospital beds available and their occupancy percentage. Our findings provide strong evidence to support efforts to ease hospital strain in order to reduce deaths during early COVID-19 outbreaks and have important implications for future infectious disease outbreaks similar to early COVID-19 variants and possibly for current Omicron outbreaks as well.

The number of COVID-19 PIH is a composite indicator for overall hospital strain, which can be caused by many complex and interrelated factors within and outside hospitals. Hospital factors include staff, facilities, equipment, drugs, ventilation beds, and preparedness. Non-pharmacological interventions (NPIs) and vaccination are major efforts that can be mobilized outside hospitals to suppress outbreak peaks and reduce hospital strain (*5*). In addition, factors such as the variant of the virus and patients' care-seeking behaviors also affect hospital strain. For example, a shortage in intensive care resources was shown to be associated with an increased risk of death from COVID-19 in the early stage of the pandemic in various countries (*2*–*3*). Our analyses with a much larger dataset also showed that the number of ventilation beds occupied by COVID-19 patients had a similar effect on fatality. However, studies on these individual determinants of hospital strain may underrate the effect of overall hospital strain on COVID-19 fatality because these studies are restricted to a small fraction of all patients who may die (*2*–*3*).

Furthermore, these factors may work together to cause difficulties for patients with severe COVID-19 and those with other diseases to be admitted to hospitals, infections in hospital staff, and inpatient cross-infections, which in turn may further increase COVID-19 fatality (*6*–*7*). Importantly, most of these factors and their interactions in each place or setting would change dynamically over time. Thus, different profiles of hospital strain determinants in a place during different periods of the epidemic may explain the different patterns of the hospital strain-fatality relation found in our study. For example, hospitals were least prepared at the beginning of the pandemic and as a result, the highest fatality was observed during period 1 in our study. As hospitals gained more experience and became more prepared, the hospital strain-fatality relation gradually became less evident. During the last period of our study, the Omicron variant caused the least severe infections, the majority of people had been vaccinated, almost all patients in the UK hospitals had been routinely tested for antigens, and hospitals, care management, and NPIs had become most prepared and efficient (*8*). Consequently, the number of PIH during this period was maintained at a relatively low level, and below it, hospital strain was not shown to be related to fatality.

Besides NPIs, including vaccination, measures can also be taken regarding or within hospitals to reduce hospital strain. For example, in the UK measures including the construction of temporary facilities (e.g. the Nightingale hospitals), the cancellation of elective admissions of patients with other diseases, and stricter triage of admissions and management of mild or moderate COVID-19 cases in communities have been implemented to ease hospital strain (*9*). Our study also showed how the relationship between hospital strain and case fatality varied with different viral variants, providing further implications for policymaking.

Having said all that, we would like to emphasize, importantly, that England's experiences with Omicron may not apply to Omicron outbreaks in all other places. If NPIs were not mobilized quickly and sufficiently, outbreaks of Omicron variants could still raise the number of patients in hospitals to a level that is high enough to cause hospital strain and increase the risk of death from COVID-19, as happened in early 2022 in Hong Kong SAR, which experienced one of the highest fatality rates from Omicron outbreaks in the world (*10*).

The study has some limitations. First, it is possible that more severe patients were admitted when the numbers of PIH were larger. Therefore, the hospital

strain-fatality relation may be a result of severity of patients admitted. Second, we used the daily numbers of PIH as the denominator of the daily case fatality, which may not be completely comparable in their severity as they may have different amalgamations of patients at different stages of disease. Third, the numerator of the daily case fatality included death events that occurred outside of hospitals, which may lead to an overestimation of the fatality at the peaks of outbreaks when a larger proportion of patients could not be admitted to hospitals and some of them died. Finally, the variant of the virus, the number of ventilation beds available and vaccination rate could all change over time and caused confounding bias in the relation between the number of PIH and fatality. However, we did sensitivity analyses and believed these limitations were unlikely to change the conclusions of this study (Supplementary Materials).

In conclusion, hospital strain is linearly associated with the risk of death from COVID-19 during early COVID-19 outbreaks, suggesting that any (additional) efforts to ease hospital strain would be beneficial in early COVID-19 outbreaks and possibly others alike. NPIs, vaccination, and hospital preparedness should be used in concert to reduce hospital strain and ultimately minimize deaths.

Conflicts of interest: No conflicts of interest.

Funding: Supported by the Shenzhen Science and Technology Programs (JSGG20220301090202005, KQTD20190929172835662).

doi: 10.46234/ccdcw2022.236

[#] Corresponding authors: Zuyao Yang, zyang@cuhk.edu.hk; Feng Sha, feng.sha@siat.ac.cn.

[1] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen City, Guangdong Province, China; [2] Division of Epidemiology, JC School of Public Health and Primary Care, the Chinese University of Hong Kong, Hong Kong Special Administrate Region, China; [3] Department of Clinical Data Center, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou City, Guangdong Province, China.

Submitted: November 21, 2022; Accepted: December 14, 2022

REFERENCES

1. Castagna F, Xue X, Saeed O, Kataria R, Puius RA, Patel SR, et al. Hospital bed occupancy rate is an independent risk factor for COVID-19 inpatient mortality: a pandemic epicentre cohort study. BMJ Open 2022;12(2):e058171. http://dx.doi.org/10.1136/bmjopen-2021-058171.
2. Wilcox ME, Rowan KM, Harrison DA, Doidge JC. Does unprecedented ICU capacity strain, as experienced during the COVID-19 pandemic, impact patient outcome. Crit Care Med 2022;50(6):e548 – e56. http://dx.doi.org/10.1097/CCM.0000000000005464.
3. Bravata DM, Perkins AJ, Myers LJ, Arling G, Zhang Y, Zillich AJ, et al. Association of intensive care unit patient load and demand with mortality rates in US department of veterans affairs hospitals during the COVID-19 pandemic. JAMA Netw Open 2021;4(1):e2034266. http://dx.doi.org/10.1001/jamanetworkopen.2020.34266.
4. GOV.UK Coronavirus (COVID-19) in the UK. https://coronavirus.data.gov.uk/. [2022-3-26].
5. Zhang Y, Quigley A, Wang Q, MacIntyre CR. Non-pharmaceutical interventions during the roll out of COVID-19 vaccines. BMJ 2021;375:n2314. http://dx.doi.org/10.1136/bmj.n2314.
6. Abbas M, Zhu NJ, Mookerjee S, Bolt F, Otter JA, Holmes AH, et al. Hospital-onset COVID-19 infection surveillance systems: a systematic review. J Hosp Infect 2021;115:44 – 50. http://dx.doi.org/10.1016/j.jhin.2021.05.016.
7. Luo QW, O'Connell DL, Yu XQ, Kahn C, Caruana M, Pesola M, et al. Cancer incidence and mortality in Australia from 2020 to 2044 and an exploratory analysis of the potential effect of treatment delays during the COVID-19 pandemic: a statistical modelling study. Lancet Public Health 2022;7(6):e537 – e48. http://dx.doi.org/10.1016/S2468-2667(22)00090-1.
8. Nyberg T, Ferguson NM, Nash SG, Webster HH, Flaxman S, Andrews N, et al. Comparative analysis of the risks of hospitalisation and death associated with SARS-CoV-2 omicron (B.1.1.529) and delta (B.1.617.2) variants in England: a cohort study. Lancet 2022;399(10332):1303 – 12. http://dx.doi.org/10.1016/S0140-6736(22)00462-7.
9. Burn S, Propper C, Stoye G, Warner M, Aylin P, Bottle A. What happened to English NHS hospital activity during the COVID-19 pandemic. The Institute for Fiscal Studies. 2021.
10. Cheung PH, Chan CP, Jin DY. Lessons learned from the fifth wave of COVID-19 in Hong Kong in early 2022. Emerg Microbes Infect 2022;11(1):1072 – 8. http://dx.doi.org/10.1080/22221751.2022.2060137.

SUPPLEMENTARY MATERIALS

Detailed Methods of the Study

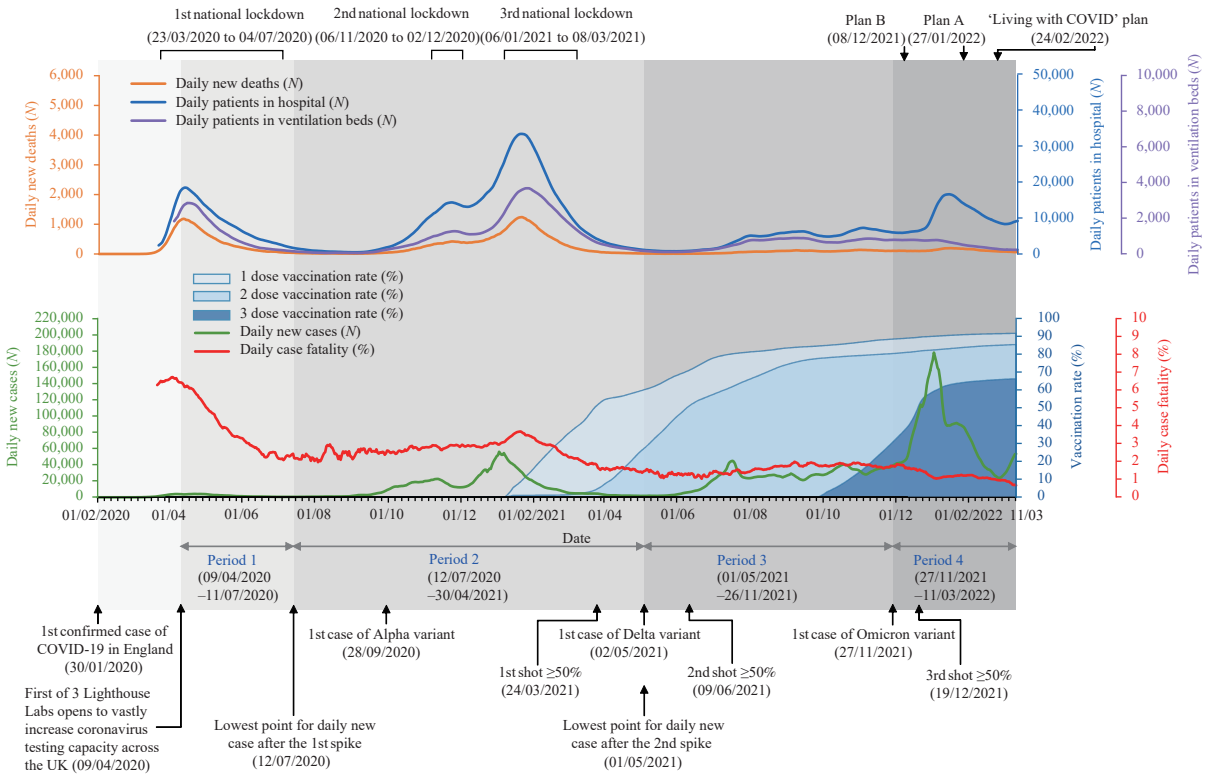
Study Design and Data Sources

This is an analysis of surveillance data extracted on a daily basis from the UK Health Security Agency (UKHSA) (*1*), the most comprehensive and authoritative source of data on coronavirus disease 2019 (COVID-19) in the country (*2*). As mentioned by the UKHSA, these COVID-19 statistics are presented in line with the Code of Practice for Statistics (the standards that producers of official statistics should commit to) (*3*) to ensure high public value, quality, and trustworthiness of the data (*4*). The baseline data included the daily number of newly diagnosed COVID-19 cases, patients currently in hospitals (PIH), patients currently in ventilation beds, and cumulative vaccination rates of 1, 2, and 3 doses in people aged 12 or above. Diagnosis of cases was confirmed by nucleic acid testing, and the date of reporting was that of sampling for the testing. The outcome of interest was those who died with COVID-19 on the same day.

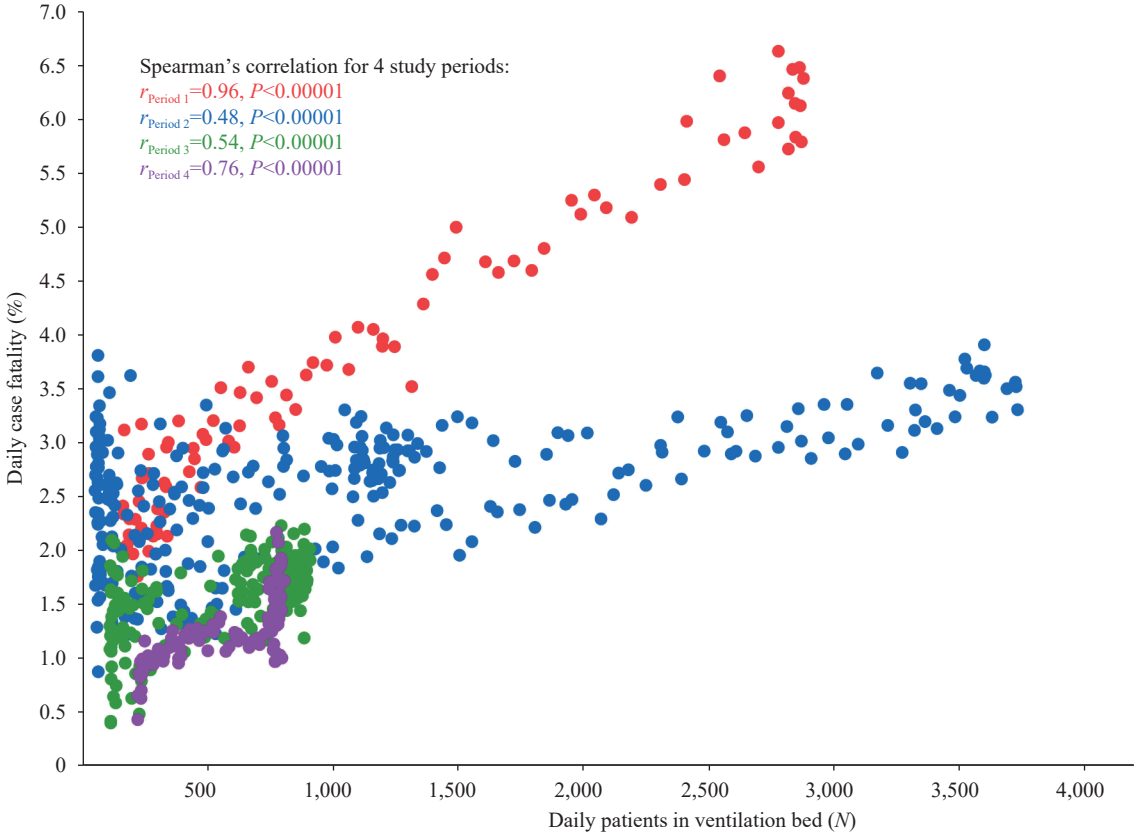
Definitions

Fatality: Daily case fatality was defined as the ratio of the daily number of deaths from COVID-19 over the daily number of COVID-19 PIH (*5*–*6*). The problem with this definition is that the number of deaths included death events that occurred outside of hospitals. This issue was addressed by examining the correlation between the percentage of deaths outside hospitals and the number of PIH, and considering the percentage of deaths outside hospitals as a potential confounder in the multiple regression analyses assessing the hospital strain-fatality relation (see the statistical analysis section).

Hospital strain: Hospital strain was measured by the total number of PIH in a day.



SUPPLEMENTARY FIGURE S1. Daily number of coronavirus disease 2019 (COVID-19) new cases, cases in hospital, death events, and daily case fatality during the 4 periods of the COVID-19 epidemic and in relation to the progress of vaccination in England between April 9, 2020 and March 11, 2022.



SUPPLEMENTARY FIGURE S2. Scatter plot and Spearman's correlation between daily coronavirus disease 2019 (COVID-19) fatality and daily number of patients currently in ventilation beds according to 4 periods of epidemic in England between April 9, 2020 and March 11, 2022.

Study periods: As the variant of the virus, vaccination coverage, and hospital strain differed considerably over time, we divided the study into four different periods and examined the effect of hospital strain on fatality separately. The first three periods were defined and divided according to the lowest numbers of cases between two major epidemic periods in England, while period 4 started from the reporting of the first Omicron case in the UK. The epidemic period before April 9, 2020 was excluded from the analyses as there was no sufficient nucleic acid testing capacity for diagnosing all infections during the period (*7*) and fatality estimates were likely biased.

Vaccination score: The vaccination rates of 1, 2, and 3 doses were converted into a single vaccination score as a measure of overall protection effect by vaccinations. Let P_1= percentage of people having received 1 dose of vaccine, P_2 for 2 doses, and P_3 for 3 doses. Then the vaccination score = $(P_1 - P_2) + 2(P_2 - P_3) + 3P_3$, which assumes that 1 dose and 2 doses give a protection approximately 1/3 and 2/3 of that of 3 doses according to current available evidence on the protection rate of vaccination (*8*).

Statistical Analysis

The 7-day moving average of daily number of new cases, PIH, deaths, vaccination rate, and daily case fatality were described chronologically in a line chart. Summary results of these variables for the 4 periods were described in a table.

The relation of hospital strain and fatality was examined graphically by scatter plots and by log-linear multivariable regression analyses weighted according to the daily number of PIH. Potential confounders adjusted included the vaccination score, admission rate (as an approximate measure of the severity of illness of patients upon hospital admission), the percentage of deaths outside hospitals (as an approximate quantification of the size of error in the number of deaths for estimating fatalities), study period (as an approximate measure of the total effect of the variants of the virus, improvements in hospital care for COVID-19 patients and other unmeasured potential

confounders that differed or changed over the four study periods), and the interaction between PIH and study period (to model different associations between PIH and fatality across four study periods). Data on the percentage of deaths outside hospitals were available on a weekly basis and acquired from the Office for National Statistics website (*9*). Additionally, we are aware that PIH acts as both the independent variable and the denominator for estimating the dependent variable (the fatality). A spurious negative association can, in theory, arise between them. As a result, a positive PIH-fatality association will be underestimated and stronger than that thus observed.

The daily case fatality was also compared for the times when hospitals were least and most strained (namely at the lowest and highest number of PIH) by using the simple regression of fatality against the number of PIH. The relative increase in fatality from the least to most strained time point was estimated and used to reflect the actual maximum increase in fatality due to hospital strain during a study period.

All analyses were conducted separately for the four study periods. P value ≤ 0.05 was considered statistically significant for all significance tests and 95% confidence intervals were constructed for all estimates. All statistical analyses and scatter plots were performed by using R software (Version 4.1.0, R Development Core Team, Vienna, Austria). The epidemic curves were drawn with OriginPro software (version 9.9.0.225, OriginLab Corporation, Northampton, USA).

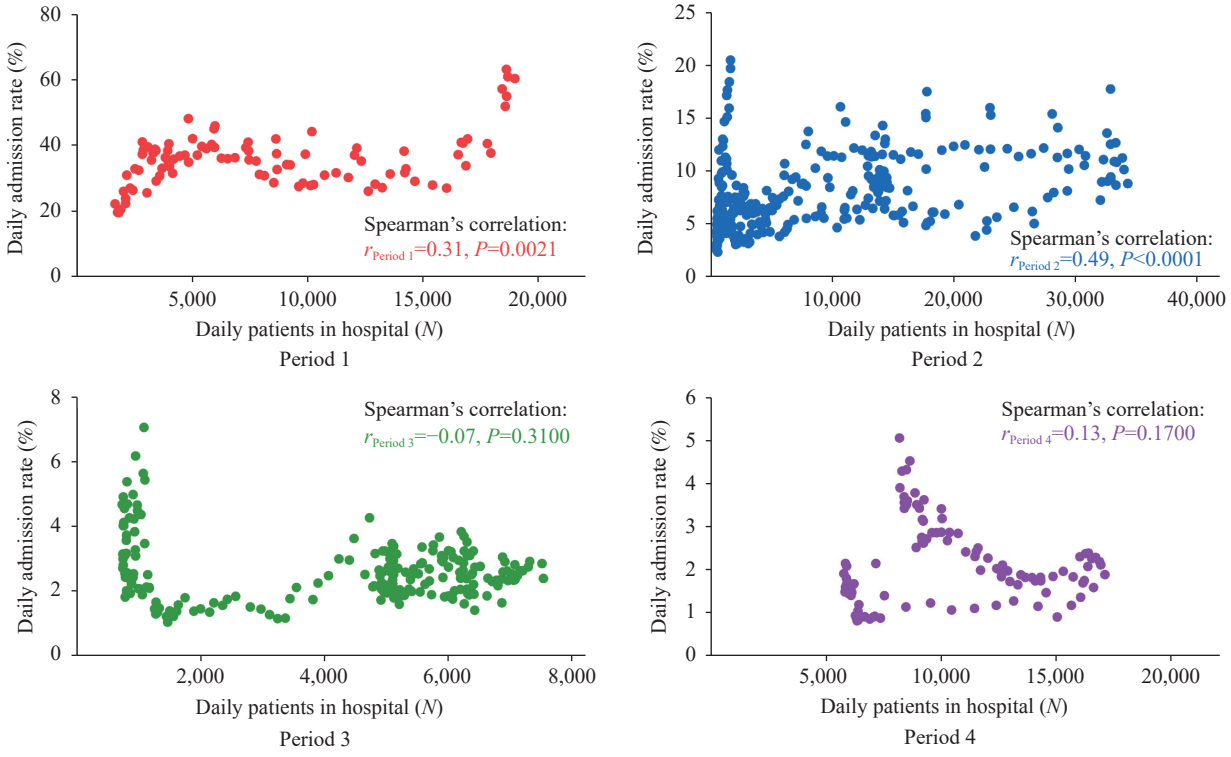
Limitations of the Study

Our study is based on a large amount of high-quality data from England on COVID-19, related deaths, and other relevant factors. Although we have made tremendous efforts to reduce biases and control for confounding, the hospital strain-fatality relation may still be fully or partly explained by residual biases and confounding. We observed that the fatality was higher when the numbers of PIH were larger. Understandably, as the total number of hospital beds was relatively stable during the period of an outbreak, only a fixed number of patients could be admitted. It is therefore possible that more severe patients were admitted when a large number of patients needed to be admitted. Consequently, the number of PIH would be positively related to the severity of patients, causing a false relation between hospital strain and fatality. The admission rate of all COVID-19 patients can be used as an indicator for the severity of patients admitted, and it indeed varied considerably, ranging from 0.8% to 63.4% during the four study periods. We found that the admission rate was not adversely associated with the number of PIH within the study periods (Supplementary Figure S3). Moreover, the hospital strain-fatality relation was affected little when admission rate was included in multiple regression analyses. Thus, we believe that the hospital strain-fatality relation is unlikely a result of the severity of patients admitted.

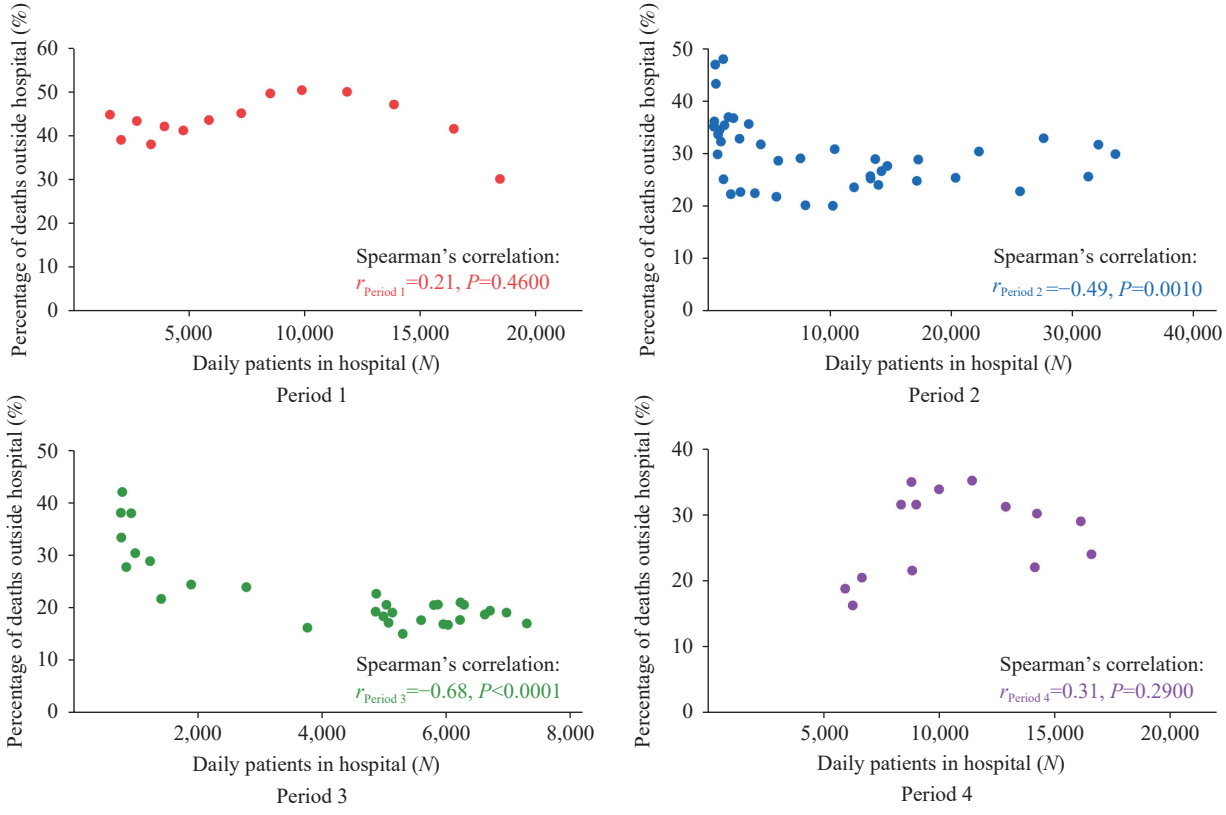
Second, our main analyses used the daily number of deaths divided by the number of PIH as the risk of death. However, the daily numbers of PIH may not be completely comparable in their severity, as they may have different amalgamations of patients at different stages of disease. It is likely that the percentage of patients admitted in earlier days and stayed on in hospitals was relatively smaller when a large number of patients need to be admitted in recent days. As patients admitted in earlier days and stayed on in hospitals were likely more severe than those admitted in recent days (*10*), the hospital strain-fatality relation in our study would have been underestimated as a result.

Third, the numerator of the daily case fatality included an average of 28.7% of deaths outside hospitals. This may lead to overestimation of the fatality at the peaks of outbreaks when a larger proportion of patients could not be admitted to hospitals and some of them died, causing a false hospital strain-fatality relation. However, the hospital strain-fatality relation was unlikely biased by the deaths outside hospitals. First, the percentage of deaths outside hospitals was not positively associated with the number of PIH, implying that it was similar regardless of the number of PIH and could not have caused a bias on the hospital strain-fatality relation (Supplementary Figure S4). In addition, the result of the hospital strain-fatality relation was not changed after the percentage of deaths outside hospitals was included in multiple regression analyses.

Finally, the variant of the virus, the number of ventilation beds available, and vaccination rate could all change over time and cause confounding bias in the relation between the number of PIH and fatality. However, we believed that by dividing into four study periods, confounding effects by the variant of the virus and the number of ventilation beds available have been reduced to a minimum as they had either not changed or changed only slightly within a study period. The confounding effect of vaccination rate was ruled out by including it in multiple regression analyses.



SUPPLEMENTARY FIGURE S3. Scatter plot and Spearman's correlation between daily admission rate and patients currently in hospital during the 4 study periods.



SUPPLEMENTARY FIGURE S4. Scatter plot and Spearman's correlation between the percentage of deaths outside hospitals and daily patients currently in hospital during the 4 study periods (on a weekly basis).

SUPPLEMENTARY TABLE S1. Regression coefficients and relative increase in daily case fatality for each predictor in the multivariable log-linear regression model in Table 2.

Variable	Regression coefficients (95% *CI*)	Relative increase (95% *CI*)	*P*-value
Study periods			
Period 1	−0.005 (−0.113, 0.103)	0.995 (0.893, 1.108)	0.9262
Period 2	Ref (0.000)	Ref (1.000)	NA
Period 3	−0.180 (−0.284, −0.076)	0.835 (0.753, 0.927)	0.0007
Period 4	0.544 (0.347, 0.741)	1.723 (1.414, 2.098)	<0.0001
Daily number of patients in hospital (per 1,000 increase) in period 2	0.014 (0.013, 0.015)	1.014 (1.013, 1.015)	<0.0001
Interaction term between daily number of patients in hospital (per 1,000 increase) and study periods			
Period 1	0.047 (0.043 ,0.052)	NA	<0.0001
Period 2	Ref (0.000)	NA	NA
Period 3	0.106 (0.088, 0.123)	NA	<0.0001
Period 4	−0.019 (−0.028, −0.011)	NA	<0.0001
Vaccination score (per 10% increase)	−0.051 (−0.059, −0.044)	0.950 (0.943, 0.957)	<0.0001
Admission rate (per 10% increase)	−0.019 (−0.040, 0.002)	0.981 (0.961, 1.002)	0.0826
Percentage of deaths outside hospital (per 10% increase)	−0.018 (−0.045, 0.009)	0.982 (0.956, 1.009)	0.1914

Note: NA=not applicable.
Abbreviation: *CI*=confidence interval.

SUPPLEMENTARY TABLE S2. Percentage increase in daily case fatality when the actual daily number of patients in hospital increased from the lowest to the highest in each of the 4 study periods.

Study period	Daily number of patients in hospitals (*N*)		Estimated daily case fatality (%)*		
	Lowest	Highest	At the lowest number patients in hospital (a)	At the highest number patients in hospital (b)	Percentage increase (b−a)/a[†]
Period 1	1,640	18,974	2.39 (2.29, 2.49)	6.87 (6.65, 7.10)	188.0 (165.9, 211.6)
Period 2	451	34,336	2.15 (2.09, 2.21)	3.65 (3.56, 3.74)	69.9 (59.0, 81.8)
Period 3	730	7,535	1.22 (1.14, 1.30)	1.93 (1.86, 2.00)	58.2 (35.4, 89.0)
Period 4	5,784	17,120	1.25 (1.15, 1.36)	1.08 (1.00, 1.17)	−13.5 (−24.0, 0.5)

* Estimated by using simple linear regression of log daily case fatality against daily number of patients in hospitals, weighted by daily number of patients in hospitals.
[†] 95% confidence interval (*CI*) was estimated by using the bootstrapping method to generate 10, 000 resampling pairs of estimated daily case fatality at the lowest and highest number patients in hospital within 95% *CI*.

REFERENCES

1. GOV.UK. Coronavirus (COVID-19) in the UK. 2022. https://coronavirus.data.gov.uk/. [2022-3-26].
2. GOV.UK. About the coronavirus (COVID-19) in the UK dashboard. 2022. https://coronavirus.data.gov.uk/about#about-the-coronavirus-COVID-19-in-the-uk-dashboard. [2022-12-11].
3. Code of practice for statistics. 2022. https://code.statisticsauthority.gov.uk/. [2022-12-11].
4. GOV.UK. Statement of voluntary application of the code of practice for statistics. 2022. https://coronavirus.data.gov.uk/details/compliance. [2022-12-11].
5. Rothman K, Greenland S. Modern epidemiology. 2nd ed. Philadelphia: Lippincott Williams & Wilkins. 1998. https://www.rti.org/publication/modern-epidemiology-2nd-edition.
6. Fletcher RH, Fletcher SW, Fletcher GS. Clinical epidemiology: the essentials. 6th ed. LWW. 2020. https://www.wolterskluwer.com/en/solutions/ovid/clinical-epidemiology-the-essentials-2532#overlay.
7. GOV.UK. Health secretary launches biggest diagnostic lab network in British history to test for coronavirus. 2020. https://www.gov.uk/government/news/health-secretary-launches-biggest-diagnostic-lab-network-in-british-history-to-test-for-coronavirus. [2022-7-18].
8. Au WY, Cheung PPH. Effectiveness of heterologous and homologous COVID-19 vaccine regimens: living systematic review with network meta-analysis. BMJ 2022;377:e069989. http://dx.doi.org/10.1136/bmj-2022-069989.
9. Office for National Statistics. Deaths registered weekly in England and Wales, provisional. 2022. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/weeklyprovisionalfiguresondeathsregisteredinenglandandwales. [2022-7-18].
10. Yao Y, Tian J, Meng X, Kan HD, Zhou L, Wang WB. Progression of severity in coronavirus disease 2019 patients before treatment and a self-assessment scale to predict disease severity. BMC Infect Dis 2022;22(1):409. http://dx.doi.org/10.1186/s12879-022-07386-3.

Perspectives

Infodemiology: The Science Studying Infodemic and Inforus

George F. Gao[1,2,3,#]

After three years of great effort, we are now facing a new challenge from the coronavirus disease 2019 (COVID-19) pandemic in China. From strict lockdown in Wuhan City (*1*–*2*), Hubei Province, to dynamic zero-COVID and subsequent precision prevention and control, the control strategies have been a good example with great achievements for preparedness and emergency response in modern settings of public health. While there are always arguments, it is clear that life-saving and time-winning processes have made essential materials available for the new challenge, with more vaccines and inhibitors/drugs available for use by now. In the past, discussions about a strategy switch raised concerns about China's response capacity and even led to rumors or mis/disinformation, which challenged the resilience and tolerance of society. China was able to explore the zero-COVID strategy because it has a strong community-level public health service (*3*) and the capacity to ensure the execution of the strategy. We have been working hard to tackle both the COVID-19 pandemic and the mis/disinformation epidemic, which was referred to as an "infodemic" as early as in 2002 (*4*). The word "infodemic" was already widely used to refer to an information epidemic when the severe acute respiratory syndrome (SARS) outbreak occurred. This term was borrowed from the real disease epidemic, but was used to refer to a wider field, including the science of humanity. An infodemic can be more exaggerated than a respiratory pathogen-caused disease because it is mainly transmitted through the internet, which allows for faster spread.

Infodemics often start suddenly whenever something new and difficult to understand occurs for the public. Recently, when the new Omicron sub-variant XBB was found in Japan, the word "hellhound" was used for this virus to scare the public and it was very effective at disturbing society. Anxiety and fear are emerging in the society and information and mis/disinformation are mixed and disseminated. While the government and professionals are working hard to control the emergence of the COVID-19 cases, they also have to work hard to deal with the infodemic.

Infodemic can sometimes be even more harmful than the disease epidemic itself. When the COVID-19 outbreak occurred in late December 2019 and early January 2020, rumors and mis/disinformation filled social media, causing serious panic around the world. Dr. Anthony Fauci, a world-renowned infectious disease expert and long-term director of National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), USA, was attacked and someone on social media even threatened to kill his two daughters. Emails of my conversation with Dr. Fauci were revealed under the US law. Bill Gates was also blamed on social media for supporting grants for infectious diseases research, claiming he supported the creation of the SARS-CoV-2 virus, the causative agent of COVID-19. Our publications from China CDC, including the identification and isolation of the SARS-CoV-2 (then called hCoV-19) and the determination of the epidemiological parameters, were published in both *the Lancet* and *The New England Journal of Medicine*. It is hard to believe that scientific research could be a target of an infodemic, even though science has played a very important role in the fight against COVID-19. Many more examples can be listed here during the early stage of the COVID-19 pandemic. Therefore, the study of infodemics needs more attention from the academic research field in the future as we face more attacks of emerging and re-emerging pathogens in the foreseeable future and other public health issues that may suddenly emerge.

We always ask ourselves if we are ready for the next potential epidemic or even pandemic. As a member of Global Preparedness Monitoring Board (GPMB) under the World Health Organization (WHO), and we have meetings twice a year to evaluate preparedness for the control and prevention of emerging pathogens. I remembered in the 2019 Annual Report, it claimed that the next pandemic might be caused by an influenza virus or coronavirus, with coronavirus at the top of the list. I was also in present in New York at the tabletop exercise of preparedness and response, called Event 201, organized by Johns Hopkins University on October 18, 2019 with an "imaginary enemy" of disease, called coronavirus associated pneumonia syndrome (CAPS), which truly sounded like the real COVID-19. As professionals, we knew that a coronavirus pandemic was a possibility but we were not ready (*5*–*6*). And indeed, we now have COVID-19 in the real world. Because of this exercise, we were

attacked and accused of knowing about or even releasing SARS-CoV-2. Again, the infodemic damaged the reputation of scientists and professionals working closely on preparedness and response for a possible pandemic. It is clear that infodemic is a real disease, which must be well studied and given more attention.

Therefore, the science of the study of infodemics is called infodemiology. Infodemiology is a new branch of epidemiology, which studies the epidemiology of infodemics. Epidemiology is a word formed from three Greek words: "epi," "demos," and "logos," which mean "on the study of population." By definition, epidemiology is a discipline of science under medicine that studies events (including diseases) that occur in a population level. In more detail, it is the study, assessment, and analysis of public health concerns in a given population; tracking the patterns and effects of diseases, environmental toxins, violence, terrorist attacks, etc. For infodemiology, as a new sub-discipline of epidemiology that typically deals with population-level questions, it studies the source and risk assessment of the mis/disinformation, public concerns, and tracks of the patterns of effects of infodemics. An infodemic is the disease. What is the causative agent of the infodemic? When I tried to figure out a good name for this, I exchanged emails with Dr. Fauci, and we both agreed to name the causative agent of the infodemic as an inforus, which is a portmanteau of information and virus. Therefore, I propose the terminology as follows: an inforus causes an infodemic, or it is the causative agent of infodemic. The study of both infodemics and inforuses is infodemiology.

In essence, an inforus is the mixture of misinformation or disinformation, plus the correct information. Unlike rumors, it is often hard to distinguish and identify an inforus because it can be a "zipped" agent of misinformation, disinformation, and information. There is a distinction between these words. Disinformation is false information spread in order to deceive people; while misinformation is wrong information or the fact that people are misinformed. Ultimately, both of them are related to people, so they fall under the category of epidemiology. Under the COVID-19 pandemic, what the virus was man-made is a typical example of inforus. In this story, mass data were "zipped" together with so many looked reasonable without scrutiny as people reply more on information of social media in the current society.

Retrospectively, we can recall so many notorious infodemics, such as vaccine hesitancy being a good example when Andrew Wakefield published his paper in *The Lancet* (*7*), linking measles vaccine with autism. Though the publication was later withdrawn, it caused several outbreaks of measles in the UK and USA, causing an increase in the death toll for children. Even now we still have serious problems with vaccine hesitancy for COVID-19. In most countries, there is only 70%–80% vaccination coverage for COVID-19. In China, though we have over 90% coverage, the coverage of both elderly population and people with underlying diseases is much lower than the average, but this population is especially vulnerable.

Infodemic are not specific diseases in public health, they can also be seen in many other fields, including the humanities. We should always bear in mind that we have to work together to tackle the infodemic problem. This is truly a global issue. I want to restate my 4 C principles (*8*) for good practice in infodemiology: Cooperation, Competition, Communication, and Coordination.

Infodemiology should be considered in the curriculum of the university level or graduate course. It should be coordinated by a joint-force to link several ministry-level offices. Let's all take the cause of infodemiology to study inforus and infodemic, for a bright future in the world.

doi: 10.46234/ccdcw2022.237

[#] Corresponding author: George F. Gao, gaofu@chinacdc.cn.

[1] Chinese Center for Disease Control and Prevention, Beijing, China; [2] Institute of Microbiology, Chinese Academy of Sciences, Beijing, China; [3] Savaid Medical School, University of Chinese Academy of Sciences, Beijing, China.

Submitted: December 24, 2022; Accepted: December 28, 2022

REFERENCES

1. Li ZJ, Chen QL, Feng LZ, Rodewald L, Xia YY, Yu HL, et al. Active case finding with case management: the key to tackling the COVID-19 pandemic. Lancet 2020;396(10243):63 – 70. http://dx.doi.org/10.1016/S0140-6736(20)31278-2.
2. Zhou L, Wu ZY, Li ZJ, Zhang YP, McGoogan JM, Li Q, et al. One hundred days of coronavirus disease 2019 prevention and control in China. Clin Infect Dis 2020;72(2):332 – 9. http://dx.doi.org/10.1093/cid/ciaa725.
3. Li ZJ, Gao GF. Strengthening public health at the community-level in China. Lancet Public Health 2020;5(12):E629 – 30. http://dx.doi.org/10.1016/S2468-2667(20)30266-8.
4. Eysenbach G. Infodemiology: the epidemiology of (mis)information. Am J Med 2002;113(9):763 – 5. http://dx.doi.org/10.1016/s0002-9343(02)01473-0.
5. Su S, Wong G, Shi WF, Liu J, Lai ACK, Zhou JY, et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. Trends Microbiol 2016;24(6):490 – 502. http://dx.doi.org/10.1016/j.tim.2016.03.003.
6. Gao GF. From "A"IV to "Z"IKV: attacks from emerging and re-emerging pathogens. Cell 2018;172(6):1157 – 9. http://dx.doi.org/10.1016/j.cell.2018.02.025.
7. Wakefield A. A statement by Dr Andrew Wakefield. Lancet 2004;363(9411):823 – 4. http://dx.doi.org/10.1016/S0140-6736(04)15710-3.
8. Gao GF, Nkengasong JN. Public health priorities for China–Africa cooperation. Lancet Public Health 2019;4(4):e177 – 8. http://dx.doi.org/10.1016/S2468-2667(19)30037-4.

Perspectives

The World Needs a "Pandamic" Solution for a Pandemic Problem

Jie Huang[1,2,#]; Ole Döring[3]; Gordon G. Liu[2]

Two decades ago, the term "infodemic" was coined right after the Severe Acute Respiratory Syndrome (SARS) emerged (*1*). The World Health Organization (WHO) defined infodemic as "*too much information including false or misleading information in digital and physical environments during a disease outbreak.*" While fully realizing the negative impact of misinformation and disinformation during an epidemic, we argue that most epidemic related information was generated not as a result of bad intention, but as a result of haste, confusion and the lack of reliable information. This is especially true for coronavirus disease 2019 (COVID-19), the emergence of a novel virus and uncertainties surrounding variants and vaccines. With a lack of timely scientific evidence, it is reasonable and even recommended for global citizens to debate and voice skepticism on certain matters. Of the past 3 years of the COVID-19 pandemic, there was some degree of infodemic during the first year. As time progressed, citizens had wider access to a range of information and more matured judgement leading to quality information trumping in an infodemic.

Therefore, we propose not to simply and pejoratively label an infodemic as the over-abundance of false or detrimental information. Instead, we propose a 3-dimensional view of an infodemic: 1) genuine information of the epidemic, 2) false information by the epidemic, and 3) intelligent information for the epidemic. The 1st dimension refers to the narrow scope of objective information concerning the epidemic itself that usually has limited circulation within government agencies and public health authorities. The 2nd dimension refers to misinformation and disinformation that are by-products of the epidemic, usually resulting from non-professionals. The 3rd dimension refers to (big) data technologies and applications for fighting a pandemic.

To further stress the importance of the 3rd dimension, we propose a new term: "pandamic" (pan-da-mic). It is similar to "pandemic" in form but conveys a completely different meaning and tone. "Pan" could be interpreted as the literal meaning of a cooking pan or in the context of Darwin's pangenesis theory and the emerging concept of pan-genome invoking all of nature (*2*); "da" refers to data applications widely used and dearly needed to fight and prevent pandemics in the most meaningful sense; "mic" means microbiology and, in particular, various omics technologies. Therefore, "pandamic" stresses the deep fusion of bio-technology (wet) and info-technology (dry). Compared to "pandemic," "pandamic" differs by one letter "a" which stands for applications. Without well-designed applications (APP), the smartphone that each of us relies on would instantaneously become a dummy. Therefore, the letter "a" used to replace letter "e" in "pandemic" is indeed powerful and meaningful.

Based on Wiktionary.org, "pandamic" is defined as "misspelling of pandemic," which showed up in the title of an article published in April 2020. From now on, "pandamic" is not the misspelling of "pandemic" any more. Rather, it is a framework to address pandemic problems. We now added a new entry onto Wiktionary.org, and defined "pandamic" as "*a broad fusion ("pan") of data applications ("da") and micro-biological research ("mic"), especially in the context of controlling and preventing pandemics.*" To facilitate the translation of this new term to other languages, we suggest the Chinese translation of "pandamic" to be "大生信" (dà shēng xìn). These three characters literally mean "big, biology, information" respectively. Interestingly, "信" also means trust. It emphasizes that information and trust should support each other. The Chinese translation for "pandamic" even rhymes with that for "pandemic" — "大流行" (dà liú xíng). Splitting "pandamic" into three widely used characters should provide a good reference for translating it into other languages using etymological approaches.

The onset of COVID-19 sparked a new wave of big data technologies to support the fight against a global pandemic. Additionally, the wave helped ensure global health security, from generating and sharing severe acute respiratory syndrome coronavirus 2 (SARS-COV-2) genome data and tracking the pandemic in real-time to facilitating diagnosis and drug repurposing (*3*). The "mic" part of "pandamic" stresses the importance of fully studying invisible

microbes in addition to visible humans. This is in line with the One Health concept. Artificial intelligence (AI) based tools such as Alpha-Fold could be used to reliably predict the micro-structure of proteins (*4*), including the SARS-CoV-2–human interactome, to explore genetic and drug perturbations (*5*). Research along these lines could get to the bottom of how virus spike proteins grab human cell receptors more and more effectively as transmissible variants emerge. This 3-dimensional research is much more powerful than the traditional 1-dimensional phylogenetic research used to assign a Greek label (i.e, Delta and Omicron) for emerging variants without a deep dive into the biology.

Another promising microbiological application is a scalable solution for wastewater genomic surveillance that allows early detection of SARS-CoV-2 variants and identification of cryptic transmission (*6*–*7*). This is more hassle-free and cost-effective than conducting mass testing on millions of people daily. Under the dynamic zero-COVID policy, China has been able to mobilize resources to conduct mass nucleic acid testing of millions on a regular basis, but this is not viable or even imaginable in many other parts of the world. Isolating and detecting pathogens from the vast sewage system is akin to finding a needle in the ocean. We choose to do this, not because it is easy, but because it is necessary. Further upgrading the technology shall push microbiology research to a new level that can eventually deliver applicable solutions to make a difference towards a more secure global public health.

About 100 years ago, Charles-Edward Winslow defined public health as "*the science and art of preventing disease, prolonging life, and promoting health*" (*8*). About 80 years ago, "global health" first appeared in the literature as a scientific term. Koplan et al. defined global health as "*an area for study, research, and practice that places a priority on improving health and achieving equity in health for all people worldwide.*" This time, the keyword "practice" is added. Through COVID-19, we see that science and art are not enough. The world needs practical and applicable science and technologies that can work effectively on the ground. Major pandemics in the 21st century brought the concept and global awareness of "One Health" (*9*). The introduction of "pandamic" comes timely and fits nicely into "One Health." After all, we propose "mic" in "pandamic" to represent "microbiology," which is a key component of One Health. Throughout the COVID-19 pandemic, the "da" part of "pandamic" has been highly visible, with data applications on both personal level and international stage. Examples of efforts include the building of a global hub for pandemic and epidemic intelligence with modern approaches to surveillance and risk assessment (*10*).

The COVID-19 pandemic is devastating, but it is now time to rethink positively and rebuild prospectively. Rephrasing "pandemic" to "pandamic" is not only a letter change, but hopefully a change of our mentality to be ready for future pandemics.

doi: 10.46234/ccdcw2022.238

[#] Corresponding author: Jie Huang, jiehuang001@hotmail.com.

[1] School of Public Health and Emergency Management, Southern University of Science and Technology, Shenzhen City, Guangdong Province, China; [2] Institute for Global Health and Development, Peking University, Beijing Municipality, China; [3] Institute for Technology Futures, Karlsruhe Institute of Technology, Karlsruhe, Germany.

Submitted: November 02, 2022; Accepted: December 25, 2022

REFERENCES

1. Coiera E. Information epidemics, economics, and immunity on the internet. We still know so little about the effect of information on public health. BMJ 1998;317(7171):1469 – 70. http://dx.doi.org/10.1136/bmj.317.7171.1469.
2. Sherman RM, Salzberg SL. Pan-genomics in the human genome era. Nat Rev Genet 2020;21(4):243 – 54. http://dx.doi.org/10.1038/s41576-020-0210-7.
3. Zhou YD, Wang F, Tang J, Nussinov R, Cheng FX. Artificial intelligence in COVID-19 drug repurposing. Lancet Digit Health 2020;2(12):e667 – 76. http://dx.doi.org/10.1016/S2589-7500(20)30192-8.
4. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. Nature 2020;577(7792):706 – 10. http://dx.doi.org/10.1038/s41586-019-1923-7.
5. Wierbowski SD, Liang SQ, Liu Y, Chen Y, Gupta S, Andre NM, et al. A 3D structural SARS-CoV-2-human interactome to explore genetic and drug perturbations. Nat Methods 2021;18(12):1477 – 88. http://dx.doi.org/10.1038/s41592-021-01318-w.
6. Karthikeyan S, Levy JI, De Hoff P, Humphrey G, Birmingham A, Jepsen K, et al. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. Nature 2022;609(7925):101 – 8. http://dx.doi.org/10.1038/s41586-022-05049-6.
7. Deng Y, Xu XQ, Zheng XW, Ding JH, Li SX, Chui HK, et al. Use of sewage surveillance for COVID-19 to guide public health response: a case study in Hong Kong. Sci Total Environ 2022;821:153250. http://dx.doi.org/10.1016/j.scitotenv.2022.153250.
8. Merrick J. Public health in a global context. Front Public Health 2013;1:9. http://dx.doi.org/10.3389/fpubh.2013.00009.
9. Huang J, McLean GR, Dubee FC, Zheng ZJ. Two pandemics in China, One Health in Chinese. BMJ Glob Health 2022;7(3):e008550. http://dx.doi.org/10.1136/bmjgh-2022-008550.
10. Morgan OW, Abdelmalik P, Perez-Gutierrez E, Fall IS, Kato M, Hamblion E, et al. How better pandemic and epidemic intelligence will prepare the world for future threats. Nat Med 2022;28(8):1526 – 8. http://dx.doi.org/10.1038/S41591-022-01900-5.

Methods and Applications

Comparing COVID-19 Case Prediction Between ARIMA Model and Compartment Model — China, December 2019–April 2020

Bangguo Qi[1,&]; Nankun Liu[1,&]; Shicheng Yu[1]; Feng Tan[1,#]

ABSTRACT

Introduction: To compare the performance between the compartment model and the autoregressive integrated moving average (ARIMA) model that were applied to the prediction of new infections during the coronavirus disease 2019 (COVID-19) epidemic.

Methods: The compartment model and the ARIMA model were established based on the daily cases of new infection reported in China from December 2, 2019 to April 8, 2020. The goodness of fit of the two models was compared using the coefficient of determination (R^2).

Results: The compartment model predicts that the number of new cases without a cordon sanitaire, i.e., a restriction of mobility to prevent spread of disease, will increase exponentially over 10 days starting from January 23, 2020, while the ARIMA model shows a linear increase. The calculated R^2 of the two models without cordon sanitaire were 0.990 and 0.981. The prediction results of the ARIMA model after February 2, 2020 have a large deviation. The R^2 of complete transmission process fit of the epidemic for the 2 models were 0.964 and 0.933, respectively.

Discussion: The two models fit well at different stages of the epidemic. The predictions of compartment model were more in line with highly contagious transmission characteristics of COVID-19. The accuracy of recent historical data had a large impact on the predictions of the ARIMA model as compared to those of the compartment model.

The outbreak of coronavirus disease 2019 (COVID-19) at the end of 2019 has caused a global pandemic and presents a major challenge to human health and survival. Accurately predicting the incidence of the COVID-19 epidemic can help distribute medicine and other health resources, take prompt and effective control measures, and suppress the spread of the epidemic. The compartment model divides the population into different compartments categorized by their epidemiological status. Ordinary differential equations were used to express the continuous dynamic changes among different compartments. Different epidemic processes of infectious diseases were simulated by adjusting the differential equations. The autoregressive integrated moving average (ARIMA) model is a time series prediction method that uses autocorrelation analysis of time series data to identify patterns of change and predict future points in the series. Previous research studies (*1*–*4*) have applied these two models in predicting COVID-19 epidemics, but few have compared them. Therefore, this study aims to compare the performance of the two models during the early COVID-19 outbreak in China. According to the timing of intervention measures and their effects, this paper divides the timeline of the epidemic into 3 stages: 1) Stage 1 from December 2, 2019, when the first case was reported, to January 22, 2020, when few interventions were taken during this stage; 2) Stage 2 from January 23 to February 1, 2022, when cordon sanitaire was implemented during this stage; 3) Stage 3 from February 2 to April 8, 2022, when centralized isolation and expanded testing were applied during this stage (details are provided in Supplementary Materials and Supplementary Figure S1, available in https://weekly.chinacdc.cn/).

METHODS

Data Source

The COVID-19 infection data was extracted from the *Infectious Disease Reporting System of Chinese Center for Disease Control and Prevention* from December 2, 2019 to April 8, 2020. The data included the reported onset date of the infection, which is the date when an infected person reported symptoms such as fever, cough, and other respiratory symptoms, and the clinical severity of each infected person, which ranged from asymptomatic, mild, moderate, severe, and critical. After excluding asymptomatic infections, a total of 81,102 confirmed cases were sorted to obtain

the number of daily new cases. This was used to construct time series models and compartment models as well as to evaluate their fit and predictive effects. The population data for the same period were collected from the official website of the National Bureau of Statistics (*5*).

Comparison Between Two Models

First, this study compared the effects of the two models in fitting the complete transmission process of the epidemic. Second, the study compared the predictions of the number of new cases without cordon sanitaire by the two models. Finally, the study compared predictions without centralized isolation and expanded testing by two models.

Comparison of Model Fitting

The coefficient of determination (R^2) was used to compare the fitting of the model. The formula is as follows:

$$R^2 = 1 - \frac{\sum (X_i^* - X_i)^2}{\sum (X_i^* - \overline{X_i})^2} \quad (1)$$

X_i^* is the true value in moment i, X_i is the predicted value in moment i, and $\overline{X_i}$ is the mean of true values.

Data Analysis

Packages "aTSA," "forecast," and "BayesianTools" in the R software (version 4.0.5, R Foundation for Statistical Computing, Vienna, Austria) were used to construct the ARIMA model and the compartment model and to predict new infections. $P<0.05$ was considered statistically significant ($\alpha = 0.05$).

RESULTS

For the compartment model, the parameters of the Stages 1–3 of the model (more details are provided in Supplementary Materials, Supplementary Table S1, and Supplementary Figure S2, available in https://weekly.chinacdc.cn/) were used to simulate the complete transmission process of the epidemic. The results are shown in Figure 1A. For the ARIMA model, the unit root test was performed on the onset sequences of Stages 1–3, and the results showed that the sequences were stationary. The autocorrelation coefficient and partial correlation coefficient of the stationary series are shown in Supplementary Figure S3A and S3B (available in https://weekly.chinacdc.cn/). The p=1–3 and q=1–3 of the onset sequence of the Stages 1–3 were preliminarily determined; the results of the residual white noise test on the 9 initially determined alternative models are shown in Supplementary Table S2 (available in https://weekly.chinacdc.cn/). According to the principle of Bayesian Information Criterion (BIC) minimization, ARIMA (1,2,1) was selected as the optimal model for the onset sequence of Stages 1–3. The optimal model was used to simulate complete transmission process of the epidemic and was compared with the compartment model (Figure 1B). The calculated R^2 of the compartment model and the ARIMA model were 0.964 ($P<0.001$) and 0.933 ($P<0.001$), respectively.

For the compartment model in Stage 1, the parameters from Stage 1 were used to predict the number of new COVID-19 cases during the 10 days starting from January 23, 2020 (i.e., first 10 days in Stage 2) with the assumption that no cordon sanitaire was implemented in China (Figure 1C). For the ARIMA model, after 3 differences in the Stage 1 incidence sequence, the unit root test showed that the sequence had been stationary. The autocorrelation coefficient and partial correlation coefficient of the stationary series are shown in Supplementary Figure S3C and S3D. The p=0 and q=1–3 of the first-stage onset sequence were preliminarily determined, and the results of the residual white noise test for the 3 preliminarily determined alternative models are shown in Supplementary Table S2. According to the principle of minimizing BIC, ARIMA (0,3,3) was chosen as the optimal model of the Stage 1 onset sequence. The optimal model was used to compare the prediction of incidence over the same period of time with that of the compartment model (Figure 1D). The prediction of the two models demonstrated that the number of new COVID-19 cases would increased if no cordon sanitaire was taken after January 23, 2020. The number of daily cases predicted by the compartment model showed an exponential increase. The ARIMA model, however, showed a linear increase, which did not reflect the high transmissibility of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus. The R^2 of the compartment model and the ARIMA model were 0.990 ($P<0.001$) and 0.981 ($P<0.001$), respectively.

For the compartment model, the parameters of the Stage 1–2 were applied to predict the number of new cases during the 10 days starting from February 2, 2020 (i.e., first 10 days in Stage 3) with the assumption of no centralized isolation and expanded testing being adopted (Figure 1E). After taking the

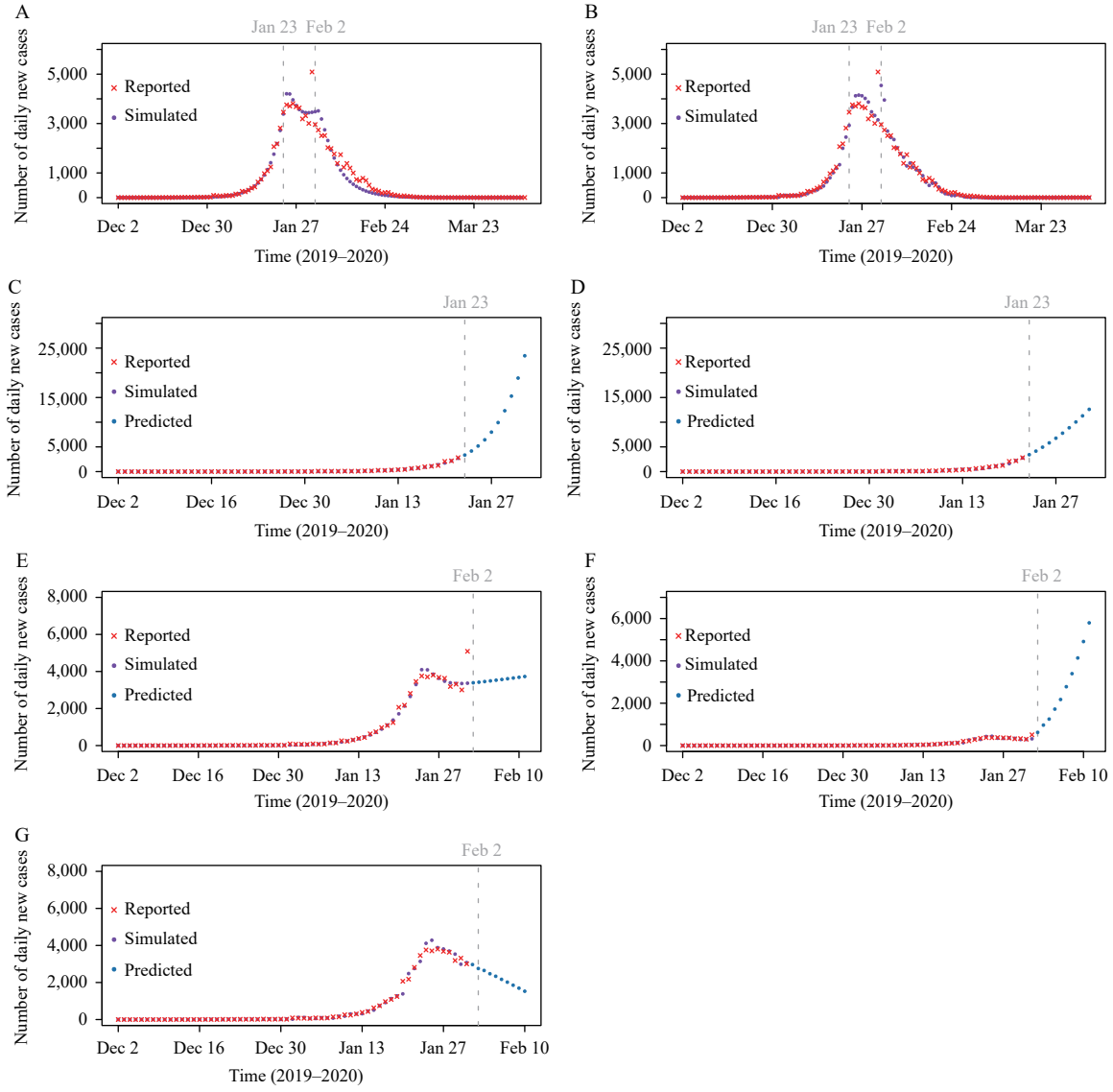


FIGURE 1. Comparision of the two models. (A) The results of the compartment model in fitting the complete transmission process of the epidemic; (B) The results of the ARIMA model in fitting the complete transmission process of the epidemic; (C) The prediction results of the compartment model without cordon sanitaire implemented; (D) The prediction results of the ARIMA model without cordon sanitaire implemented; (E) The prediction results of the compartment model without centralized isolation and expanded testing implemented; (F) The prediction results of the ARIMA model without centralized isolation and expanded testing implemented; (G) The prediction results of the ARIMA model without centralized isolation and expanded testing implemented after excluding outlier.

Note: Due to the abnormally high number of daily new cases reported on February 1, 2020, the prediction results of the ARIMA model after February 2, 2020 showed a rapid increase. After excluding the outlier, the prediction of daily new cases will decrease.

Abbreviation: ARIMA=autoregressive integrated moving average.

three differences of the Stage 1–2 onset sequence for the ARIMA model, the unit root test showed that the sequence had been stationary. The autocorrelation coefficient and partial correlation coefficient of the stationary series are shown in Supplementary Figure S3E and S3F. The p=1–3 and q=0 of the onset sequences of Stage 1–2 were preliminarily determined; the residual white noise test results of the 3 initially determined alternative models are shown in Supplementary Table S2. According to the principle of BIC minimization, ARIMA (2,3,0) was selected as the optimal model of Stage 1–2 onset sequence. This optimal model was used to predict incidence outside the modeling sequence for the same duration, and the result was compared with the compartment model (Figure 1F). Due to the abnormally high number of

cases reported in a single day on February 1, the results of the ARIMA model had a large deviation and showed a rapid increase. After excluding this outlier, the results of re-fitting the ARIMA model are shown in Figure 1G. The R^2 of the compartment model and the ARIMA model, without excluding outliers, were 0.969 (P<0.001) and 0.948 (P<0.001), respectively. After excluding outliers, the R^2 of the ARIMA model was 0.937.

DISCUSSION

Appropriate predictions can help authorities promptly adjust control strategies and allocate medical resources. The compartment model and the ARIMA model are used by numerous researchers in the prediction of COVID-19. Taking the early COVID-19 epidemic in China as an example, the predictions of the compartment model and the ARIMA model at different stages of the epidemic were compared and both models fit well at different stages of the epidemic. Furthermore, the predictions of the compartment model are in line with the highly contagious transmission characteristics of the COVID-19. In addition, since the ARIMA model is a prediction method that considers the changing trends of past values over time and predicts future values by fitting the mathematical model with historical data, the accuracy of recent historical data has a relatively large impact on the results of model extrapolation. Based on the numbers of daily new cases and parameters supported by existing literature, the compartment model can be calibrated using Markov chain Monte Carlo (MCMC) algorithm, allowing its predictions to be relatively less affected by outliers.

Although the ARIMA model does not perform as well as the compartment model in terms of predicting COVID-19, it is important to consider that the novel coronavirus is still in the process of dynamic evolution in the future. With this in mind, the parameters of the compartment model can also change accordingly and are difficult to obtain. Meanwhile, the accurate simulation of model has high requirements for the selection of parameters. Compared with the compartment model, the ARIMA model only needs time series data to build a forecasting model, which is easy to implement and has high accuracy for short-term forecasting. It can be quickly applied to forecasting COVID-19.

The compartment model divides the population into different compartments, with the dynamics of these compartments described by ordinary differential equations. Researchers can incorporate different compartments and parameters into the model to more accurately simulate transmission patterns and epidemiological characteristics of the novel coronavirus. Compared with the ARIMA model, which replaces various influencing factors with time, the compartment model can analyze the impact of population movement, vaccination, isolation measures, and other interventions on disease transmission. Therefore, when predicting COVID-19, it is necessary to comprehensively consider the advantages of different models and choose the best model based on existing conditions.

This study was subject to at least two limitations. First, there were no real-world values to compare with the models' predictions on the temporal trends of the numbers of daily new cases in specific hypothetical scenarios. Therefore, the accuracy of predictions could not be compared using mean absolute error (MAE) and root mean squared error (RMSE). Second, as a result of dynamic changes in epidemic-related influencing factors — such as prevention and control measures, medical resources, and viral transmissibility, etc. — neither the compartment model nor the ARIMA model could guarantee the accuracy of their long-term predictions. It is necessary to constantly update data to improve their prediction accuracy.

Funding: Supported by the National Natural Science Foundation of China (No. 82041023).

doi: 10.46234/ccdcw2022.239

[#] Corresponding author: Feng Tan, tanfeng@chinacdc.cn.

[1] Chinese Center for Disease Control and Prevention, Beijing, China.
[&] Joint first authors.

Submitted: December 13, 2022; Accepted: December 27, 2022

REFERENCES

1. Chintalapudi N, Battineni G, Amenta F. COVID-19 virus outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: a data driven model approach. J Microbiol Immunol Infect 2020;53(3):396 – 403. http://dx.doi.org/10.1016/j.jmii.2020.04.004.
2. Chen SM, Chen QS, Yang JT, Lin L, Li LY, Jiao LR, et al. Curbing the COVID-19 pandemic with facility-based isolation of mild cases: a mathematical modeling study. J Travel Med 2021;28(2):taaa226. http://dx.doi.org/10.1093/jtm/taaa226.
3. Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. Sci Total Environ 2020;729:138817. http://dx.doi.org/10.1016/j.scitotenv.2020.138817.
4. Hao XJ, Cheng SS, Wu DG, Wu TC, Lin XH, Wang CL. Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. Nature 2020;584(7821):420 – 4. http://dx.doi.org/10.1038/s41586-020-2554-8.
5. China NBOS. National data. 2022. https://data.stats.gov.cn/. [2022-07-11]. (In Chinese).

SUPPLEMENTARY MATERIAL

Compartment Model

Under the framework of the susceptible-exposed-infectious-recovered (SEIR) model, pre-symptomatic cases (P), asymptomatic infected cases (A), hospitalized patients (H), and shelter-isolated infected persons (Iq) were added to simulate the transmission pattern of coronavirus disease 2019 (COVID-19) at different stages (Supplementary Figure S1). According to the different interventions that were taken at different stages, different compartments and parameters were introduced to establish a multi-stage infectious disease compartment model.

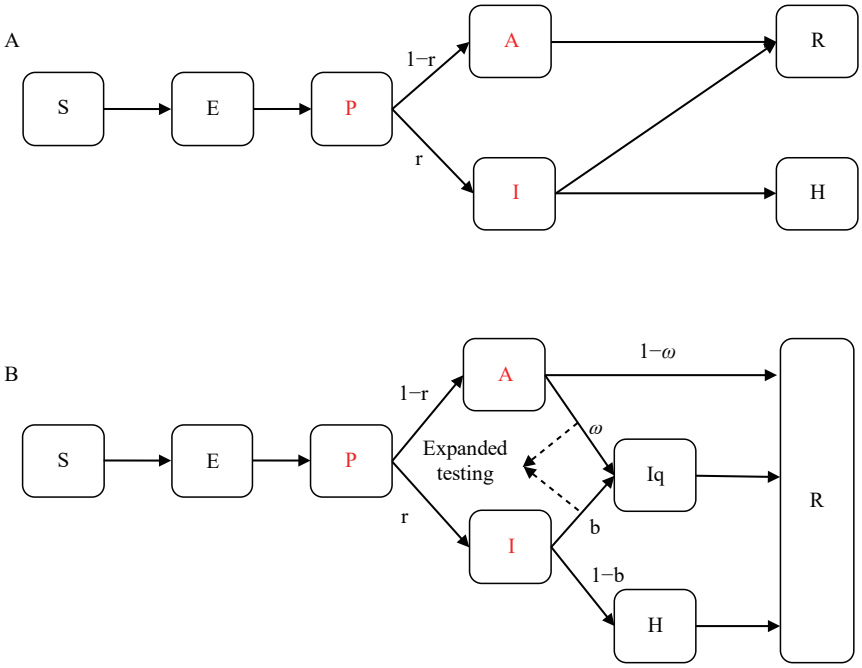
The reported onset date of the first case of COVID-19 infection, December 2, 2019, was set as the starting date of the model. The initial value of the model is set to S=59,170,000, I=1, E=P=A=H=R=0. The model parameters were obtained from two kinds of sources. Some parameters were obtained according to previous studies. Other parameters were estimated by the Markov chain Monte Carlo algorithm (MCMC) (*1*) (Supplementary Table S1).

Compartment Model Structures and Parameters

A compartment model was developed to simulate the full-spectrum dynamics of COVID-19 in China between December 2, 2019 and April 8, 2020. The resident population of Hubei Province in 2018 was used as the initial susceptible population, and the inter-provincial population movement during the Chunyun, or Spring Festival travel period, in Hubei Province was not considered in the analysis. The reasons are as follows:

1. In the period of Stage 1, most infections occurred in Hubei Province.

2. During the Chunyun (January 10 to January 22, 2020), the outflow population in Hubei Province were isolated in their homes for a long time (i.e., after January 23, 2020).



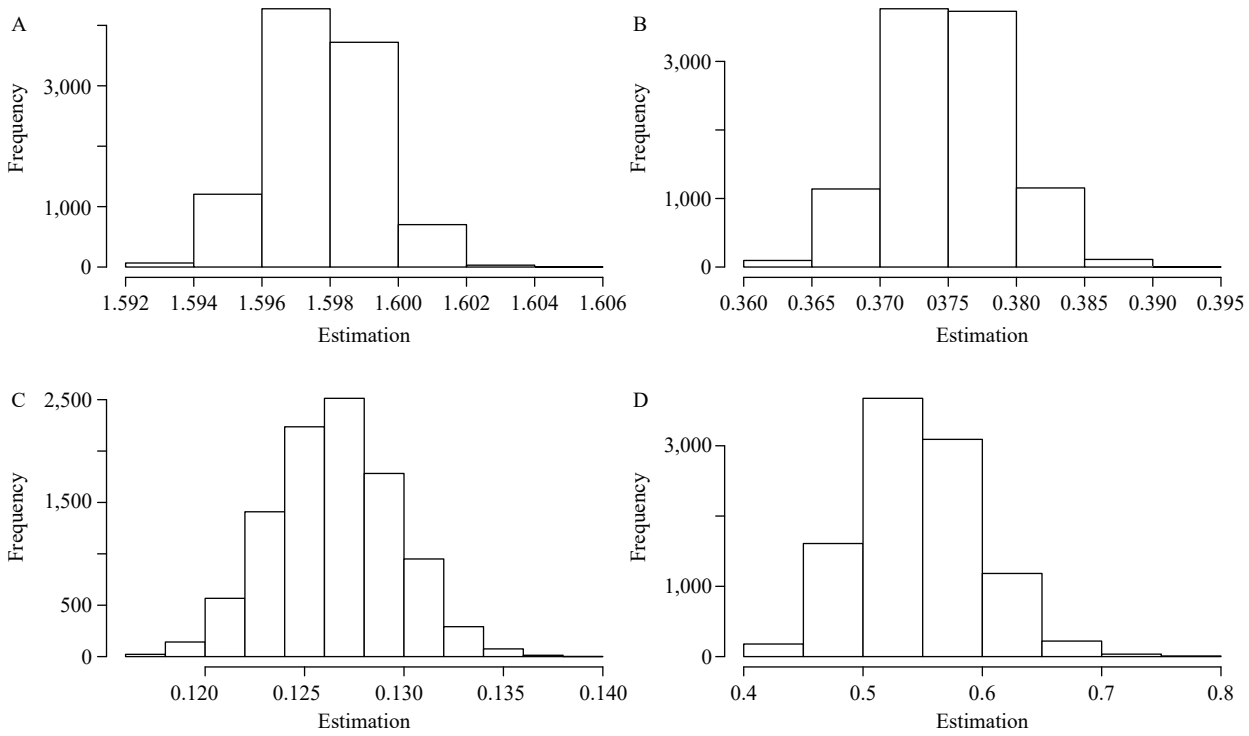
SUPPLEMENTARY FIGURE S1. Schematic diagram of the COVID-19 transmission in the compartment model. (A) At Stage 1 and Stage 2; (B) At Stage 3.

Note: S is the numbers of susceptible population. E is the exposed state with latent infection. A is the asymptomatic infected cases (i.e., people who never develop symptoms). P is the pre-symptomatic cases. Although there is no pre-symptomatic phase for asymptomatic individuals, P was treated as a transitional phase in order to distinguish between the non-infectious latent period and the infectious pre-symptomatic state. I is the confirmed cases with symptoms. H is the hospitalized patients who are hospitalized because of worsening symptoms. Iq is the confirmed cases who are detected because of expanded testing and isolated in the square cabin. R is the recovered patients.

SUPPLEMENTARY TABLE S1. Definition and value of parameters.

Model parameters	Meaning	Value	Sources
β_1	Transmission rate (Stage 1)	1.598	MCMC calibration
β_2	Transmission rate (Stage 2)	0.376	MCMC calibration
β_3	Transmission rate (Stage 3)	0.127	MCMC calibration
r	Proportion of symptomatic infected cases	20.0%	(*3*)
θ	Relative transmission risk for pre-symptomatic and asymptomatic infections	30.0%	(*4*)
D_e	Duration of latent period	2.9 days	(*5*)
D_p	Duration of infectious pre-symptomatic state	2.3 days	(*5–6*)
D_i	Infectious period for asymptomatic, mild, and moderate infections	7.0 days	(*7*)
$\gamma_{1,2}$	Duration from illness onset to hospitalization (Stage 1 and Stage 2)	6.0 days	(*5*)
γ_3	Duration from illness onset to hospitalization (Stage 3)	2.0 days	(*5*)
α	Time to recovery for the confirmed cases	10.0 days	(*8*)
δ	Duration from testing to isolation	2.0 days	(*9*)
ω	Detection rate of asymptomatic infections	54.7%	MCMC calibration
b	Proportion of isolation in the square cabin among all symptomatic infections	9.6%	(*10–11*)

Abbreviation: MCMC=Monte Carlo Markov Chain.



SUPPLEMENTARY FIGURE S2. Posterior distributions of Markov Chain Monte Carlo samples for calibrated model parameters. (A) For β_1; (B) For β_2; (C) For β_3; (D) For ω.

3. According to the report of the China-World Health Organization (WHO) joint investigation expert group, community transmission outside Hubei Province was very limited and most of it was in family clusters (*2*).

The model structure is illustrated in Supplementary Figure S1.

Model Structure

The ordinary differential equations are as follows:

$$\begin{cases} \frac{dS}{dt} = \frac{-\beta_{1,2} S(I + \theta A + \theta P)}{N} \\ \frac{dE}{dt} = \frac{\beta_{1,2} S(I + \theta A + \theta P)}{N} - \frac{E}{D_e} \\ \frac{dP}{dt} = \frac{E}{D_e} - \frac{P}{D_p} \\ \frac{dA}{dt} = \frac{(1-r)P}{D_p} - \frac{A}{D_i} \\ \frac{dI}{dt} = \frac{rP}{D_p} - \frac{I}{D_i} - \frac{I}{\gamma_{1,2}} \\ \frac{dH}{dt} = \frac{I}{\gamma_{1,2}} - \frac{H}{\alpha} \\ \frac{dR}{dt} = \frac{A+I}{D_i} + \frac{H}{\alpha} \end{cases} \tag{S1}$$

$$\begin{cases} \frac{dS}{dt} = \frac{-\beta_3 S(I + \theta A + \theta P)}{N} \\ \frac{dE}{dt} = \frac{\beta_3 S(I + \theta A + \theta P)}{N} - \frac{E}{D_e} \\ \frac{dP}{dt} = \frac{E}{D_e} - \frac{P}{D_p} \\ \frac{dA}{dt} = \frac{(1-r)P}{D_p} - \frac{(1-\omega)A}{D_i} - \frac{\omega A}{\delta} \\ \frac{dI}{dt} = \frac{rP}{D_p} - \frac{(1-b)I}{\gamma_3} - \frac{bI}{\delta} \\ \frac{dH}{dt} = \frac{(1-b)I}{\gamma_3} - \frac{H}{\alpha} \\ \frac{dI_q}{dt} = \frac{\omega A + bI}{\delta} - \frac{I_q}{\alpha} \\ \frac{dR}{dt} = \frac{I_q + H}{\alpha} + \frac{(1-\omega)A}{D_i} \end{cases} \tag{S2}$$

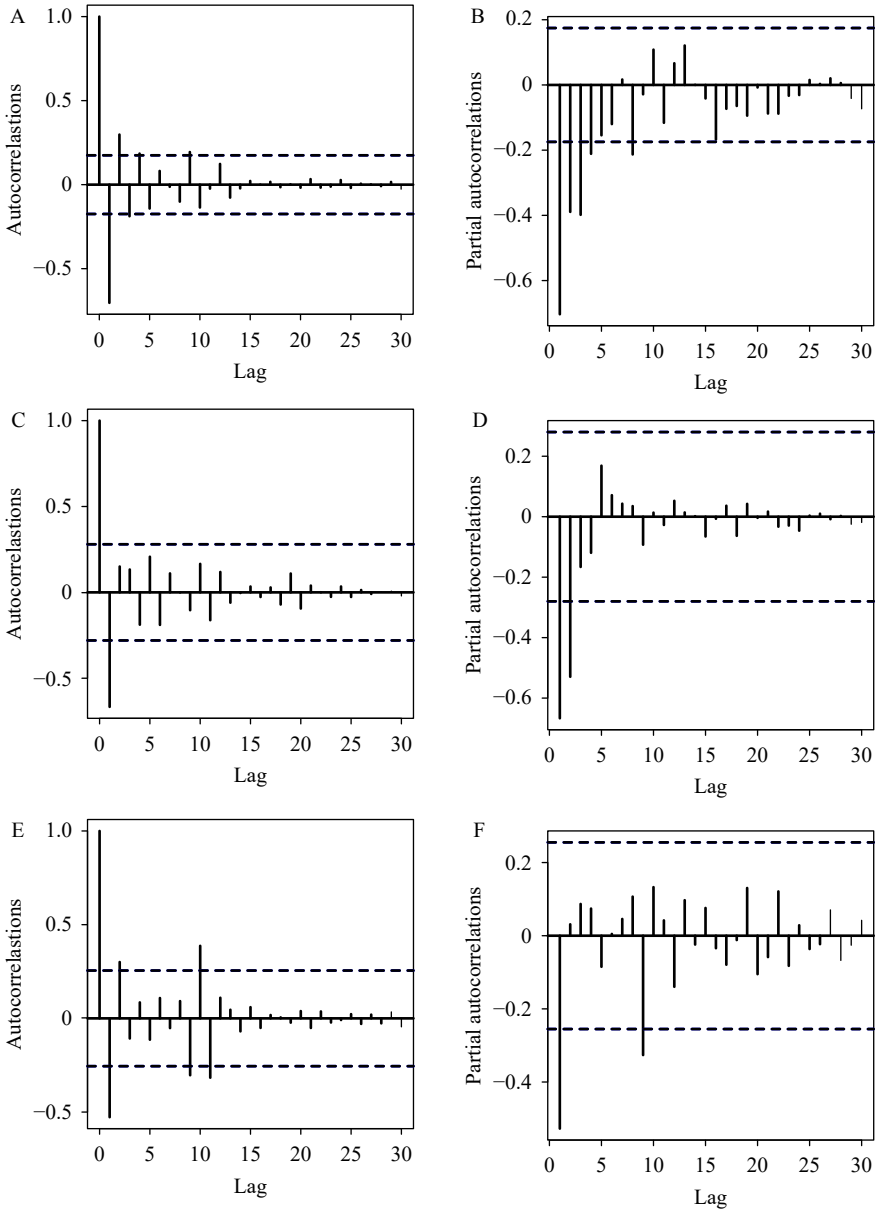
SUPPLEMENTARY TABLE S2. Residual white noise test and BIC value of alternative models.

Stages	Alternative models	Whether the model was effective	BIC value
1	ARIMA (0,3,1)	No	–
	ARIMA (0,3,2)	No	–
	ARIMA (0,3,3)	Yes	597.69
1–2	ARIMA (1,3,0)	No	–
	ARIMA (2,3,0)	Yes	862.97
	ARIMA (3,3,0)	No	–
1–3	ARIMA (1,2,1)	Yes	1,796.24
	ARIMA (2,2,1)	Yes	1,797.92
	ARIMA (3,2,1)	Yes	1797.63
	ARIMA (1,2,2)	Yes	1797.47
	ARIMA (2,2,2)	Yes	1799.20
	ARIMA (3,2,2)	Yes	1799.50
	ARIMA (1,2,3)	Yes	1798.30
	ARIMA (2,2,3)	Yes	1797.16
	ARIMA (3,2,3)	Yes	1801.48

Note: "–" means not applicable.
Abbreviation: BIC=Bayesian Information Criterion; ARIMA=autoregressive integrated moving average.

Formula S1 indicates the differential equations for COVID-19 transmission during Stage 1 and Stage 2, and Formula S2 indicates the differential equations for COVID-19 transmission during Stage 3 where parameter β is the transmission rate for the confirmed cases with symptoms. Pre-symptomatic and asymptomatic cases were assumed to be less infectious compared to people suffering from symptoms with a relative risk θ. r is proportion of symptomatic infected cases among all infected cases; D_e is the latent period; D_p is the pre-symptomatic infectious period; D_i is the asymptomatic, mild, and moderate infectious period; γ is the duration from illness onset to hospitalization; α is the time to recovery for confirmed cases; δ is the duration from testing to isolation; ω is detection rate of asymptomatic infections; and b is proportion of isolation in the square cabin among all symptomatic infections.

All model parameters are summarized in Supplementary Table S1. This system dynamics model is implemented in the R software (version 4.0.5; R Core Team, Vienna, Austria).



SUPPLEMENTARY FIGURE S3. The ACF and PACF graphs. (A) ACF for Stages 1–3; (B) PACF for Stages 1–3; (C) ACF for Stage 1; (D) PACF for Stage 1; (E) ACF for Stages 1–2; (F) PACF for Stages 1–2.
Abbreviation: ACF=autocorrelation function; PACF=partial autocorrelation function.

Model Calibration

To estimate parameters $(\beta_1, \beta_2, \beta_3, \omega)$, the Metropolis–Hastings Markov Chain Monte Carlo (MCMC) algorithm was used. The Delayed Rejection Adaptive Metropolis (DRAM) sampler from the BayesianTools R package was also used. Calibration target is numbers of daily new cases, which was extracted from *The Infectious Disease Reporting System of the Chinese Center for Disease Control and Prevention* from December 2, 2019 (i.e., date of onset of the first reported infection) to April 8, 2020. It was assumed that the observed number of daily new cases in which individuals experienced symptom onset on day d — denoted as k_d — follows a Poisson distribution with rate $\mu_d = \frac{rP_{d-1}}{D_p}$, in which P_{d-1} is the expected number of pre-symptomatic cases on day (d−1). The likelihood function is

$$L(\beta_1, \beta_2, \beta_3, \omega) = \prod_{d=1}^{D} \frac{e^{-\mu_d} \mu_d^{k_d}}{k_d!} \quad \text{(S3)}$$

In MCMC sampling, a non-informative flat was set prior of Unif (0,2) for $\beta_1, \beta_2, \beta_3$ and Unif (0,1) for ω. After a burn-in period of 50,000 iterations, MCMC sampling was continued for an additional 100,000 iterations and MCMC samples were selected at every 10 iterations to avoid auto-correlation. Means of the model parameters are presented in Supplementary Table S1. Supplementary Figure S2 shows a histogram of the posterior distributions.

ARIMA Model

The ARIMA model was established in 4 steps (*12*). 1) Time series stabilization. The model requires that the fitted time series be stable, that is, the mean and variance of the series do not change over time. If the original series is not stable, it needs to be made into a stationary series by means of difference. The time series diagram and unit root test can be used to judge whether the series is stationary. 2) Model identification. Autocorrelation function (ACF) and partial autocorrelation function (PACF) were drawn from the sequence that meets the stationarity requirement after the difference. The values of p and q in the model were preliminarily determined according to its truncation or tailing situation, and multiple alternative models were fitted. 3) Model diagnosis. To check whether the model was effective, the sufficiency of information extraction was tested. The Ljung-Box residual white noise test was carried out on the candidate model. A non-white noise fitting residual sequence indicated that there were still relevant factors that had not been extracted and needed to be excluded. 4) Model optimization. According to the Bayesian Information Criterion (BIC), the model with the minimum value of BIC function was the optimal model.

REFERENCES

1. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. 3rd ed. New York: Chapman and Hall. 2013. https://www.taylorfrancis.com/books/mono/10.1201/b16018/bayesian-data-analysis-david-dunson-donald-rubin-john-carlin-andrew-gelman-hal-stern-aki-vehtari.
2. Shi YC. China-WHO joint investigation expert group: community transmission is limited in other provinces outside Hubei, mostly family clustered outbreaks. 2022. https://news.ifeng.com/c/7uKcoSNfqfQ. [2022-07-11]. (In Chinese).
3. He ZY, Ren LL, Yang JT, Guo L, Feng LZ, Ma C, et al. Seroprevalence and humoral immune durability of anti-SARS-CoV-2 antibodies in Wuhan, China: a longitudinal, population-level, cross-sectional study. Lancet 2021;397(10279):1075 – 84. http://dx.doi.org/10.1016/S0140-6736(21)00238-5.
4. Chen Y, Wang AH, Yi B, Ding KQ, Wang HB, Wang JM, et al. Epidemiological characteristics of infection in COVID-19 close contacts in Ningbo city. Chin J Epidemiol 2020;41(5):667 – 71. http://dx.doi.org/10.3760/cma.j.cn112338-20200304-00251. (In Chinese).
5. Hao XJ, Cheng SS, Wu DG, Wu TC, Lin XH, Wang CL. Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. Nature 2020;584(7821):420 – 4. http://dx.doi.org/10.1038/s41586-020-2554-8.
6. Li Q, Guan XH, Wu P, Wang XY, Zhou L, Tong YQ, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020;382(13):1199 – 207. http://dx.doi.org/10.1056/NEJMoa2001316.
7. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. Lancet Public Health 2020;5(5):e261 – 70. http://dx.doi.org/10.1016/S2468-2667(20)30073-6.
8. Shao P, Shan YJ. Beware of asymptomatic transmission: study on 2019-nCoV prevention and control measures based on extended SEIR model. bioRxiv. 20202020 – 1. http://dx.doi.org/10.1101/2020.01.28.923169.
9. Kucharski AJ, Klepac P, Conlan AJK, Kissler SM, Tang ML, Fry H, et al. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. Lancet Infect Dis 2020;20(10):1151 – 60. http://dx.doi.org/10.1016/S1473-3099(20)30457-6.
10. China TSCI. Fighting covid-19: China in action. People's Daily. http://to.china-embassy.gov.cn/eng/sgxw/202006/t20200608_57786.htm. [2020-6-8].
11. Bai M, Liu XQ, Wu WQ, Sun K, Huang X, Jin HZ. Clinical characteristics of 472 cases of novel coronavirus pneumonia in Wuhan Jiangan Makeshift (Fangcang) Hospital. Clin Focus 2020;35(4):297-301. https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2020&filename=LCFC202004002. (In Chinese).
12. Wang Y. Time series analysis with R. Beijing: China Renmin University Press. 2015. http://find.nlc.cn/search/showDocDetails?docId=-4764597708139708653&dataSource=ucs01&query=%E6%97%B6%E9%97%B4%E5%BA%8F%E5%88%97%E5%88%86%E6%9E%90. (In Chinese).

Notifiable Infectious Diseases Reports

Reported Cases and Deaths of National Notifiable Infectious Diseases — China, October 2022

Diseases	Cases	Deaths
Plague	0	0
Cholera	3	0
SARS-CoV	0	0
Acquired immune deficiency syndrome*	3,965	1,549
Hepatitis	109,020	71
Hepatitis A	752	0
Hepatitis B	90,015	33
Hepatitis C	16,020	36
Hepatitis D	10	0
Hepatitis E	1,683	2
Other hepatitis	540	0
Poliomyelitis	0	0
Human infection with H5N1 virus	0	0
Measles	111	0
Epidemic hemorrhagic fever	400	2
Rabies	12	6
Japanese encephalitis	27	0
Dengue	326	0
Anthrax	35	0
Dysentery	2,559	0
Tuberculosis	51,125	304
Typhoid fever and paratyphoid fever	494	1
Meningococcal meningitis	4	0
Pertussis	2,594	0
Diphtheria	0	0
Neonatal tetanus	1	0
Scarlet fever	1,383	0
Brucellosis	2,535	0
Gonorrhea	7,959	1
Syphilis	39,054	2
Leptospirosis	28	0
Schistosomiasis	11	0
Malaria	92	0
Human infection with H7N9 virus	0	0
COVID-19[†]	9,354	0
Influenza	69,072	1
Mumps	9,537	0

Continued

Diseases	Cases	Deaths
Rubella	118	0
Acute hemorrhagic conjunctivitis	2,009	0
Leprosy	13	0
Typhus	118	0
Kala azar	16	0
Echinococcosis	115	0
Filariasis	0	0
Infectious diarrhea[§]	61,743	0
Hand, foot and mouth disease	47,395	0
Total	421,228	1,937

* The number of deaths of acquired immune deficiency syndrome (AIDS) is the number of all-cause deaths reported in the month by cumulative reported AIDS patients.

[†] According to the data from website of the National Health Commission of the People's Republic of China, the number of COVID-19 cases in the whole country in October were 9,885 cases, which included 366 cases from Hong Kong Special Administrative Regions, Macao Special Administrative Regions, and Taiwan, and 165 imported foreign cases. No death were reported.

[§] Infectious diarrhea excludes cholera, dysentery, typhoid fever and paratyphoid fever.

The number of cases and cause-specific deaths refer to data recorded in National Notifiable Disease Reporting System in China, which includes both clinically-diagnosed cases and laboratory-confirmed cases. Only reported cases of the 31 provincial-level administrative divisions in the mainland of China are included in the table, whereas data of Hong Kong Special Administrative Region, Macau Special Administrative Region, and Taiwan, China are not included. Monthly statistics are calculated without annual verification, which were usually conducted in February of the next year for de-duplication and verification of reported cases in annual statistics. Therefore, 12-month cases could not be added together directly to calculate the cumulative cases because the individual information might be verified via National Notifiable Disease Reporting System according to information verification or field investigations by local CDCs.

doi: 10.46234/ccdcw2022.220

Indexed by Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI), PubMed Central (PMC), Scopus, Chinese Scientific and Technical Papers and Citations, and Chinese Science Citation Database (CSCD)

Copyright © 2022 by Chinese Center for Disease Control and Prevention

All Rights Reserved. No part of the publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise without the prior permission of *CCDC Weekly*. Authors are required to grant *CCDC Weekly* an exclusive license to publish.

All material in *CCDC Weekly* Series is in the public domain and may be used and reprinted without permission; citation to source, however, is appreciated.

References to non-China-CDC sites on the Internet are provided as a service to *CCDC Weekly* readers and do not constitute or imply endorsement of these organizations or their programs by China CDC or National Health Commission of the People's Republic of China. China CDC is not responsible for the content of non-China-CDC sites.

The inauguration of *China CDC Weekly* is in part supported by Project for Enhancing International Impact of China STM Journals Category D (PIIJ2-D-04-(2018)) of China Association for Science and Technology (CAST).

CCDC weekly

Vol. 4 No. 52 Dec. 30, 2022

Responsible Authority
National Health Commission of the People's Republic of China

Sponsor
Chinese Center for Disease Control and Prevention

Editing and Publishing
China CDC Weekly Editorial Office
No.155 Changbai Road, Changping District, Beijing, China
Tel: 86-10-63150501, 63150701
Email: weekly@chinacdc.cn

CSSN
ISSN 2096-7071
CN 10-1629/R1