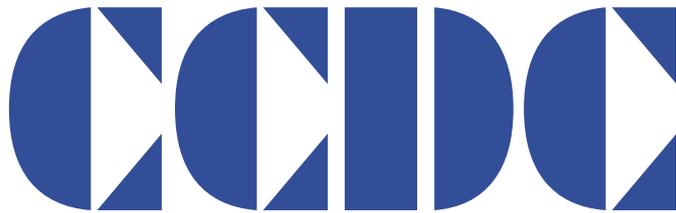


## CHINA CDC WEEKLY


**Vol. 5 No. 4 Jan. 27, 2023**  
 weekly

## 中国疾病预防控制中心周报



## COVID-19 ISSUE (33)

## Preplanned Studies

Characterizing Human Collective Behaviors During COVID-19 — Hong Kong SAR, China, 2020 71

## Methods and Applications

High-Resolution Data on Human Behavior for Effective COVID-19 Policy-Making — Wuhan City, Hubei Province, China, January 1–February 29, 2020 76

Optimization of Population-Level Testing, Contact Tracing, and Isolation in Emerging COVID-19 Outbreaks: a Mathematical Modeling Study — Tonghua City and Beijing Municipality, China, 2021–2022 82

Knowledge Graph: Applications in Tracing the Source of Large-Scale Outbreak — Beijing Municipality, China, 2020–2021 90



ISSN 2096-7071



## 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
Yuantaao 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

**Senior Scientific Editors** Ning Wang

**Scientific Editors** Weihong Chen

Liuying Tang

Qing Yue

Yu Chen

Ruotao Wang

Xudong Li

Meng Wang

Ying Zhang

Peter Hao (USA)

Shicheng Yu

Nankun Liu

Zhihui Wang

Qian Zhu

Liwei Shi

Xi Xu

## Preplanned Studies

## Characterizing Human Collective Behaviors During COVID-19 — Hong Kong SAR, China, 2020

Zhanwei Du<sup>1</sup>; Xiao Zhang<sup>2</sup>; Lin Wang<sup>3</sup>; Sidan Yao<sup>4</sup>; Yuan Bai<sup>1,2</sup>; Qi Tan<sup>1,2</sup>; Xiaoke Xu<sup>5</sup>; Sen Pei<sup>6</sup>; Jingyi Xiao<sup>1</sup>; Tim K. Tsang<sup>1</sup>; Qiuyan Liao<sup>1</sup>; Eric H. Y. Lau<sup>1,2</sup>; Peng Wu<sup>1,2</sup>; Chao Gao<sup>7,#</sup>; Benjamin J. Cowling<sup>1,2,#</sup>

### Summary

#### What is already known about this topic?

People are likely to engage in collective behaviors online during extreme events, such as the coronavirus disease 2019 (COVID-19) crisis, to express awareness, take action, and work through concerns.

#### What is added by this report?

This study offers a framework for evaluating interactions among individuals' emotions, perceptions, and online behaviors in Hong Kong Special Administrative Region (SAR) during the first two waves of COVID-19 (February to June 2020). Its results indicate a strong correlation between online behaviors, such as Google searches, and the real-time reproduction numbers. To validate the model's output of risk perception, this investigation conducted 10 rounds of cross-sectional telephone surveys on 8,593 local adult residents from February 1 through June 20 in 2020 to quantify risk perception levels over time.

#### What are the implications for public health practice?

Compared to the survey results, the estimates of the risk perception of individuals using our network-based mechanistic model capture 80% of the trend of people's risk perception (individuals who are worried about being infected) during the studied period. We may need to reinvigorate the public by involving people as part of the solution that reduced the risk to their lives.

Countries have adopted public health and social measures to control coronavirus disease 2019 (COVID-19) transmission, loss of jobs, education impediments, and other critical social and cultural activities affected during crises (1–3). During such extreme events, people are likely to engage in a miscellaneous set of behaviors (henceforth collective behavior), such as exchanging information in social situations (4). For instance, people are inclined to swiftly spread messages via social media about natural disasters in order to gain more knowledge and decrease

unforeseen worries (5). Recent studies suggest people would likely similarly share such COVID-19 content on their social media (6–7).

The first COVID-19 case was confirmed in Hong Kong Special Administrative Region (SAR) on January 22, 2020 (8). Since then, Hong Kong has put in place strong measures to prevent COVID-19, including wearing masks in all public areas; the closure of schools, bars and social venues; work-at-home policies; and restaurant disease control measures (9). Alongside this, social media has predictably become an all-embracing part of daily life for rapid knowledge dissemination during the isolationism of the COVID-19 pandemic. These platforms have been used by Hong Kong people to, among other uses, express their emotions (e.g., depression) during the pandemic (10). To study human collective behaviors during the COVID-19 response, this study evaluates interactions among Hong Kong residents' emotions, perceptions, and behaviors using a network-based mechanistic model that links together external situations of COVID-19 prevalence and social networks (Figure 1A).

This study's stochastic, network-based, agent-based model incorporates environment, agents, local behaviors, and ever-updating rules by combining the mapping between multiagent systems and social networks (11). Following external situation reports, individuals perceive risks, experience different emotional reactions, and further change their behaviors — usually by following the strengthening process (i.e., risk perception drives emotional reactions, and emotional reactions affect collective behaviors, Figure 1A). Inversely, resulting behaviors should reduce individuals' emotions (e.g., anxiety and stress) and risk perception.

This study gives a general framework for network-based, agent-based models — extending this study's authors' prior proposed study of collective behaviors during extreme events (5). In this framework, each individual has a profile of six attributes, of which one is

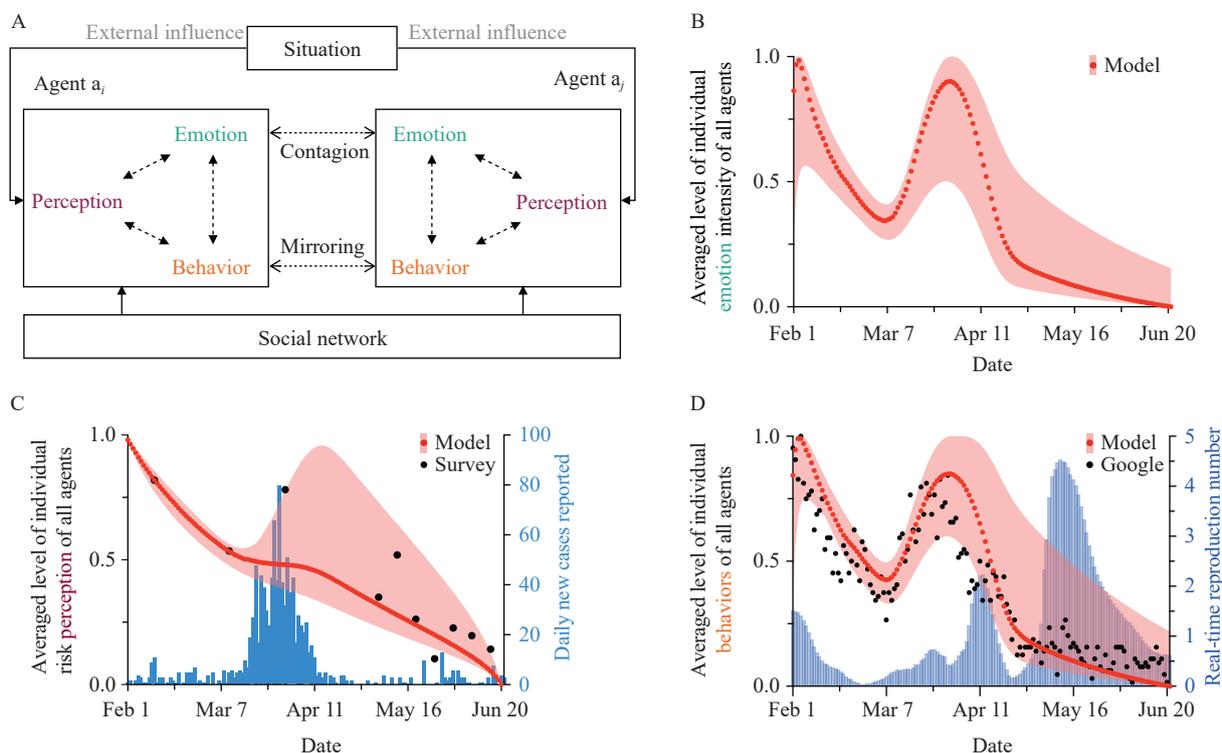


FIGURE 1. Reconstruction of human collective behaviors during the first and second waves of the coronavirus disease 2019 (COVID-19) pandemic in Hong Kong from February 1 to June 20 in 2020, using a human collective behavior model that incorporates daily external situation reports of new infections. (A) Structure of the individual-based model with influences from the external situation and individual-based social networks. (B) Results for daily individual emotional intensity levels. The red dots and shading indicate the median and 95% credibility interval (CrI) of square root of normalized averaged levels of individual emotional intensity across all agents in the model. (C) Results for daily individual risk perception levels. (D) Results for normalized daily search behavior levels.

Note: We projected the daily time series of the observed behaviors of individuals by tracking external situation reports of new infections. In response to the external situation reports, each individual can experience the strengthening process that first changes the perceived risk of infections and then changes the emotional reactions (e.g., anxiety and stress), which in turn leads to the adjustment of protective behaviors. Each individual will also experience the weakening process, in which the changes in protective behaviors can reduce the emotional reactions and perceived risk of infections. Due to daily interactions on social networks (e.g., messaging in Facebook), the emotion and behaviors of an individual can influence the emotions and behaviors of other connected individuals in the network, denoted as emotion contagion and behavior mirroring. These dotted lines denote the interactions of psychological factors within and between individuals. Our human collective behavior model is well matched to the observed time series in our survey data (black dots). The black dots indicate the normalized daily logarithmic percentage of individuals who are worried about being infected in our survey to indicate people's risk perception in Hong Kong. The red dots and shading indicate the median and 95% CrI of normalized averaged levels of individual risk perception across all agents in the model. The blue bars indicate daily new cases reported in Hong Kong. Our human collective behavior model incorporating the interactions among agents is well fitted to the observed time series data (black dots). The black dots indicate the observation of normalized daily Google search behaviors of residents in Hong Kong. The red dots and shading indicate the median and 95% CrI of averaged levels of individual behaviors across all agents in the model.

individual behavior, two psychological factors (emotional intensity and risk perception) and three intrinsic characteristics (personality characteristics, open level, and expression level). This study then projected the population-scale human collective behaviors by the overall behaviors of all agents in the study's multiagent system. More details of methods and data can be found in the Supplementary Material (available in <http://weekly.chinacdc.cn>). Lastly,

although this study analyzed the model with a focus on Hong Kong, the results can be applied to other cities in general. All analyses were conducted by Matlab software (version R2020a, The MathWorks Inc., Natick, MA, USA).

Following reports of external situations, individuals perceive modified risks, experience different emotional reactions, and change their behaviors. Those resulting protective behaviors inversely reduce people's anxiety

and perceptions of susceptibility, resulting in less decreases in perceived risk during March and April 2020 (Figure 1C and D). This study then projected the daily time series of the observed behaviors of individuals by tracking external situation reports of new infections (Figure 1B–D). When comparing observed behaviors of Google search data, this study's human collective behavior model incorporating the interactions among agents is well fitted. About 79% of the observed Google search data are included in the 95% credibility interval (CrI) ranging of normalized averaged levels of individual behaviors across all agents in the model (Figure 1D). The Pearson's linear correlation coefficient between the real-time reproduction number ( $R_t$ ) and the averaged levels of individual behaviors is  $-0.40$  with a  $P$ -value of  $0.0001$ . After the individual behavior reached its bottom,  $R_t$  began to increase after April 23, 2020.

To validate the model output of risk perception, this investigation conducted 10 rounds of cross-sectional telephone surveys from February 1 through June 20 in 2020. A total of 8,593 local adult residents have been interviewed via these surveys (Supplementary Material). Such large-scale longitudinal data provides an opportunity to quantify risk perception levels over time. This study estimates the daily time series of the risk perception of individuals using the network-based mechanistic model informed by external situation reports of new infections. Compared to the survey results, the estimated values of median and 95% CrI capture 80% (8 out of 10 surveys) of the trend of people's risk perception (individuals who worried about being infected) during the studied period (Figure 1C). The average level of individual risk perception continues to decrease, but slows down as case numbers start to soar in March 2020.

## DISCUSSION

Hong Kong has been following a zero-COVID strategy, through which it has implemented stringent social distancing measures (including unprecedented movement restrictions, quarantines on inbound travellers, universal masking, closure of schools, bars and social venues, work-at-home policies, and restaurant measures) to curb COVID-19 transmission since January 2020 and bring case numbers down to low levels in each wave (9,12–13). Given the continued threat of COVID-19 in Hong Kong (14), pandemic fatigue is a natural response due to the complex interplay of cultural and social factors (e.g.,

the risk perception of threats) (15–18), which has been observed in many countries (19–22). This study's results of decreasing individual risk perception indicate that the gradual emergence of pandemic fatigue in Hong Kong arose as demotivation from a series of related behaviors.

The theory behind this study is the interplay of human psychological factors and external influences on psychological factors, as indicated in Figure 1 (5). This study has several limitations. First, it analyzed self-reported behavior but did not validate this against actual behaviors — although self-reported surveys have been widely used to study human behavior, such as contact patterns (23) and hospital attendance (24). Second, this study focuses on the period of the first two waves in Hong Kong, which are taken as extreme events, rather than subsequent waves. Informed by the daily Google search interest in Google Trend for COVID-19 in Hong Kong, this study finds it has decreased 20% to 50% on average since the third wave, perhaps due to people having gained enough knowledge and having gotten used to the COVID-19 situation. Third, other social activities may affect the risk perception and protective behaviors. Prolonged financial stress due to job loss, mask costs, and the distrust of governmental policies may also contribute to the emergence of pandemic fatigue in the studied period. Fourth, the risk perception is represented in this study by the proportion of weekly surveyed residents who are beyond moderately worried, while in reality, it is a complex concept involving many emotions-including anxiety, depression, post-traumatic stress disorder, and psychological stress triggered by extreme events (e.g., COVID-19, weather disasters) (25–26). Importantly, excepting worry, other sentiments related to risk perception may have different temporal dynamics and attributes. This study cautions researchers to be cautious about their conclusions when using the framework in this study. Fifth, this study runs 100 stochastic simulations and uses medians of risk perception across all agents across simulations to calibrate the model (Supplementary Material, section parameter calibration). More simulations may have a direct impact on the range of simulation outputs, but a limited impact on the medians. As such, this study reminds researchers to be careful when they use the range of the study model, rather than only medians. Despite these limitations, the close matching of model output with human collective behavior of Google search data and surveyed risk perception data suggests that the capacity of

individual-based human collective behavior model in capturing the actual population behaviors.

The current socio-political and economic dilemma caused by a pandemic call for decision-makers to focus beyond the number of cases reported. The fluctuation of human collective behaviors online reflects people's social, emotional, and mental health needs, impacted by external situations. To maintain people's risk perception of COVID-19 on a high level, global leadership may need to reinvigorate the public by engaging people as part of the solution, understanding their needs, acknowledging their hardship, and empowering them to live their lives with reduced risk (19,27).

**Conflicts of interest:** BJC consults for AstraZeneca, Fosun Pharma, GlaxoSmithKline, Moderna, Pfizer, Roche and Sanofi Pasteur. BJC is supported by the AIR@innoHK program of the Innovation and Technology Commission of the Hong Kong SAR Government. No other conflicts of interest reported.

**Funding:** Supported by the Key Program for International Science and Technology Cooperation Projects of China (No. 2022YFE0112300), AIR@InnoHK administered by Innovation and Technology Commission of the Research Grants Council of the Hong Kong SAR Government, the Health and Medical Research Fund, Food and Health Bureau, Government of the Hong Kong Special Administrative Region (Grant No. COVID190118, 21200632), the Research Grants Council of the Hong Kong Special Administrative Region, China (GRF 17110221), National Natural Science Foundation of China (Grant No. 72104208). The weekly telephone surveys were supported by the Government of the Hong Kong Special Administrative Region.

doi: 10.46234/ccdcw2023.014

# Corresponding authors: Benjamin J. Cowling, bcowling@hku.hk; Chao Gao, cgao@nwpu.edu.cn.

<sup>1</sup> WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China; <sup>2</sup> Laboratory of Data Discovery for Health, Hong Kong Science and Technology Park, New Territories, Hong Kong Special Administrative Region, China; <sup>3</sup> Department of Genetics, University of Cambridge, Cambridge, CB2 3EH, UK; <sup>4</sup> Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A\*STAR), Singapore; <sup>5</sup> College of Information and Communication Engineering, Dalian Minzu University, Dalian City, Liaoning Province, China; <sup>6</sup> Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York City, NY, USA; <sup>7</sup> School of Artificial Intelligence, Optics, and Electronics (iOpen), Northwestern Polytechnical University, Xi'an City, Shaanxi Province, China.

Submitted: October 08, 2022; Accepted: January 11, 2023

## REFERENCES

1. World Health Organization. Survey tool and guidance: rapid, simple, flexible behavioural insights on COVID-19. 2020. <https://www.medbox.org/document/survey-tool-and-guidance-rapid-simple-flexible-behavioural-insights-on-covid-19#GO>. [2023-1-9].
2. Bai Y, Xu MD, Liu CF, Shen MW, Wang L, Tian LW, et al. Travel-related importation and exportation risks of SARS-CoV-2 Omicron variant in 367 prefectures (cities) — China, 2022. *China CDC Wkly* 2022;4(40):885 – 9. <http://dx.doi.org/10.46234/ccdcw2022.184>.
3. Bai Y, Du ZW, Xu MD, Wang L, Wu P, Lau EHY, et al. International risk of SARS-CoV-2 Omicron variant importations originating in South Africa. *J Travel Med* 2022;29(6):taac073. <http://dx.doi.org/10.1093/jtm/taac073>.
4. Adar E, Weld DS, Bershad BN, Gribble SS. Why we search: visualizing and predicting user behavior. In: Proceedings of the 16th international conference on world wide web (WWW'07). ACM, 2007;161-70. <http://dx.doi.org/10.1145/1242572.1242595>.
5. Gao C, Liu JM. Network-based modeling for characterizing human collective behaviors during extreme events. *IEEE Trans Syst Man Cybern: Syst* 2017;47(1):171 – 83. <http://dx.doi.org/10.1109/TSMC.2016.2608658>.
6. World Health Organization. Social media & COVID-19: a global study of digital crisis interaction among Gen Z and Millennials. 2021. <https://www.who.int/news-room/feature-stories/detail/social-media-covid-19-a-global-study-of-digital-crisis-interaction-among-gen-z-and-millennials>. [2023-1-9].
7. Tsao SF, Chen H, Tisseverasinghe T, Yang Y, Li LH, Butt ZA. What social media told us in the time of COVID-19: a scoping review. *Lancet Digit Health* 2021;3(3):e175 – 94. [http://dx.doi.org/10.1016/S2589-7500\(20\)30315-0](http://dx.doi.org/10.1016/S2589-7500(20)30315-0).
8. CHP investigates highly suspected imported case of novel coronavirus infection. 2020. <https://www.info.gov.hk/gia/general/202001/22/P2020012200982.htm>. [2023-1-9].
9. The government of the Hong Kong special administrative region. Together, we fight the virus. <https://www.coronavirus.gov.hk/eng/index.html>. [2023-1-9].
10. Yang X, Yip BHK, Mak ADP, Zhang DX, Lee EKP, Wong SYS. The differential effects of social media on depressive symptoms and suicidal ideation among the younger and older adult population in Hong Kong during the COVID-19 pandemic: population-based cross-sectional survey study. *JMIR Public Health Surveill* 2021;7(5):e24623. <http://dx.doi.org/10.2196/24623>.
11. Jiang YC, Hu J, Lin DH. Decision making of networked multiagent systems for interaction structures. *IEEE Trans Syst Man Cybern Part A: Syst Humans* 2011;41(6):1107 – 21. <http://dx.doi.org/10.1109/TSMCA.2011.2114343>.
12. Du ZW, Wang CY, Liu CF, Bai Y, Pei S, Adam DC, et al. Systematic review and meta-analyses of superspreading of SARS-CoV-2 infections. *Transbound Emerg Dis* 2022;69(5):e3007 – 14. <http://dx.doi.org/10.1111/tbed.14655>.
13. Du ZW, Wang SQ, Bai Y, Gao C, Lau EHY, Cowling BJ. Within-host dynamics of SARS-CoV-2 infection: a systematic review and meta-analysis. *Transbound Emerg Dis* 2022. <http://dx.doi.org/10.1111/tbed.14673>.
14. Wang SQ, Zhang FD, Wang Z, Du ZW, Gao C. Reproduction numbers of SARS-CoV-2 Omicron subvariants. *J Travel Med* 2022;29(8):taac108. <http://dx.doi.org/10.1093/jtm/taac108>.
15. Morrison M, Parton K, Hine DW. Increasing belief but issue fatigue: changes in Australian Household Climate Change Segments between 2011 and 2016. *PLoS One* 2018;13(6):e0197988. <http://dx.doi.org/10.1371/journal.pone.0197988>.
16. Masten AS, Motti-Stefanidi F. Multisystem resilience for children and youth in disaster: reflections in the context of COVID-19. *Adv Res Sci* 2020;1(2):95 – 106. <http://dx.doi.org/10.1007/s42844-020-00010-w>.

17. Du ZW, Wang L, Shan SW, Lam D, Tsang TK, Xiao JY, et al. Pandemic fatigue impedes mitigation of COVID-19 in Hong Kong. *Proc Natl Acad Sci USA* 2022;119(48):e2213313119. <http://dx.doi.org/10.1073/pnas.2213313119>.
18. Gao C, Liu JM. Uncovering spatiotemporal characteristics of human online behaviors during extreme events. *PLoS One* 2015;10(10):e0138673. <http://dx.doi.org/10.1371/journal.pone.0138673>.
19. Habersaat KB, Scheel AE. Pandemic fatigue - Reinvigorating the public to prevent COVID-19. World Health Organization-Regional Office for Europe. 2020. <https://www.preventionweb.net/publications/view/74208>. [2023-1-9].
20. Ilesanmi OS, Bello AE, Afolabi AA. COVID-19 pandemic response fatigue in Africa: causes, consequences, and counter-measures. *Pan Afr Med J* 2020;37(Suppl 1):37. <http://dx.doi.org/10.11604/pamj.suppl.2020.37.37.26742>.
21. World Health Organization. Statement - Rising COVID-19 fatigue and a pan-regional response. 2020. <https://www.who.int/europe/news/item/06-10-2020-statement-rising-covid-19-fatigue-and-a-pan-regional-response>. [2023-1-9].
22. Crane MA, Shermock KM, Omer SB, Romley JA. Change in reported adherence to nonpharmaceutical interventions during the COVID-19 pandemic, April-November 2020. *JAMA* 2021;325(9):883 - 5. <http://dx.doi.org/10.1001/jama.2021.0286>.
23. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med* 2008;5(3):e74. <http://dx.doi.org/10.1371/journal.pmed.0050074>.
24. Hegde ST, Salje H, Sazzad HMS, Hossain MJ, Rahman M, Daszak P, et al. Using healthcare-seeking behaviour to estimate the number of Nipah outbreaks missed by hospital-based surveillance in Bangladesh. *Int J Epidemiol* 2019;48(4):1219 - 27. <http://dx.doi.org/10.1093/ije/dyz057>.
25. Yu SB, Eisenman D, Han ZQ. Temporal dynamics of public emotions during the COVID-19 pandemic at the epicenter of the outbreak: sentiment analysis of weibo posts from Wuhan. *J Med Internet Res* 2021;23(3):e27078. <http://dx.doi.org/10.2196/27078>.
26. Han ZQ, Shen MF, Liu HB, Peng YF. Topical and emotional expressions regarding extreme weather disasters on social media: a comparison of posts from official media and the public. *Humanit Soc Sci Commun* 2022;9(1):421. <http://dx.doi.org/10.1057/S41599-022-01457-1>.
27. Du ZW, Tian LW, Jin DY. Understanding the impact of rapid antigen tests on SARS-CoV-2 transmission in the fifth wave of COVID-19 in Hong Kong in early 2022. *Emerg Microbes Infect* 2022;11(1):1394 - 401. <http://dx.doi.org/10.1080/22221751.2022.2076616>.

## SUPPLEMENTARY MATERIAL

### Data

**Epidemic data.** This study collected the daily data of all newly confirmed coronavirus cases (by reporting date) in Hong Kong from January 31 to June 28, 2020 (*I*) to denote the impact of external situations on Hong Kong residents.

**Google search data.** This study collected the daily data of search interest for the disease of COVID-19 in Hong Kong during the period of January 31 to June 28, 2020 from Google Trend (*I*) using URL (<https://trends.google.com/trends/explore?geo=HK&q=coronavirus>) to denote the online behavior dynamics of Hong Kong residents. This study normalized the daily values between 0 and 1 and used them for model parameter calibration.

**Survey data.** In each monthly/weekly survey from February 1 to June 20 in 2020, this study contacted either 500 or 1,000 local residents through random digit dialing of landlines and mobile telephones (using age, gender, education, and employment information to weight response frequencies relative to the adult population in Hong Kong) (*I*). Then, 8,593 local residents were interviewed through these 10 cross-sectional telephone surveys. During these calls, this study asked each participant about their perception of the risk of being infected by COVID-19. Specifically, to assess the risk perception, participants were asked whether they were worried about being infected with COVID-19 (with a spectrum of response options including: not at all, mildly, moderately, very much, and extremely worried). Then, this study defined the overall risk perception each week as the proportion of weekly surveyed residents who are beyond moderately worried.

### Methods

**General formulation for modeling collective behaviors.** Combining the mapping between multiagent systems and social networks (*I*), we extend multiagent systems to a network-based agent-based model, which contains three basic parts: 1) an environment for both serving as a platform and communication channels for agents, 2) autonomous agents that react to each other in their local environment based on local behaviors and updating rules involved in their own decision-making process, following the general framework of collective behaviors during extreme events (*I*).

Modeling human responses, especially two types of feedback loops underlying the decision-making process in an uncertain environment, is an open and non-equilibrium system. Through implicitly incorporating certain global influences or biases (e.g., uncertainty reduction theory in our study) into local behaviors, the whole system can achieve a desired global state (i.e., collective regulation of information-related behaviors and anxious emotion in our study). Here, we would highlight the effects of two conditions on collective regulation based on the methodology of complex behavior characterization.

As a sufficient condition to realize self-organized computation, agents should implicitly incorporate certain global influences into local autonomy. When agents perform certain behaviors in an uncertainty situation, their profiles (measured by parameters) can be dynamically updated based on the interactions with other agents. Each local agent  $a_i$  makes a decision for a certain behavior in terms of a design-making mechanism for each neighbor of  $a_i$ . Before  $a_i$  makes a decision, it will first estimate the influences from its neighbors  $a_j$  because of the content diffusion. Then,  $a_i$  makes a decision by aggregating all impact factors in order to implement the exploit or explore behaviors with different probabilities based on their own decision-making mechanism.

**Modeling collective behaviors during extreme events.** Following external situation reports, individuals perceive risks, experience different emotional reactions, and further affect their behaviors following the strengthening process. Inversely, resulting behaviors would reduce individuals' emotions (e.g., anxiety and stress) and risk perception, with respect to the weakening process.

In our case study of Hong Kong, we consider 1,000 agents with weighted links. Each agent has a profile of six attributes, including individual behavior (*IB*), emotional intensity (*EI*), risk perception (*RP*), personality characteristic (*PC*), open level (*OL*) and expression level (*EL*), represented by  $ib_i(t)$ ,  $ei_i(t)$ ,  $rp_i(t)$ ,  $pc_i(t)$ ,  $ol_i(t)$ , and  $el_i(t)$ , respectively, which are all in the range between 0 and 1 for agent  $a_i$  at time step  $t$ . Agent  $a_i$  with larger  $ib_i(t)$  and  $ei_i(t)$  would perform more behaviors and emotions to other agents respectively.  $rp_i(t)$  indicates how sensitive

agent  $a_i$  is to external situation at time step  $t$ .  $pc_i(t)$ ,  $ol_i(t)$ , and  $el_i(t)$  characterize how much this agent would weaken the impact of negative information, be impacted by other agents, and affect others, respectively. The collective responses, represented by  $ib_c(t)$ ,  $ei_c(t)$ , and  $rp_c(t)$  at time step  $t$ , denote the mean estimates of  $ib_i(t)$ ,  $ei_i(t)$ , and  $rp_i(t)$  of all agents.

Given  $IB$  and  $EI$  have the same updating equations, we use  $S$  to denote  $IB$  or  $EI$ , and  $s$  to represent the  $ib$  or  $ei$ . Agents are affected by the local environment/social network (denoted as  $F_{SN}$ ) and the psychological mechanism after obtaining the latest public news (denoted as  $F_{PM}$ ). The agent  $a_i$  is updated over time by

$$s_i(t+1) = \eta F_{SN}(s_i(t)) + (1-\eta) F_{PM}(s_i(t)), \quad (1)$$

$$F_{SN}(s_i(t)) = \sum_{j \neq i} w_{ij} \times [f(\bar{s}_i(t), s_i(t)) - s_i(t)], \quad (2)$$

where  $\eta$  is a scaling factor.  $w_{ij}$  indicates the link weight between agents  $a_i$  and  $a_j$ .  $\bar{s}_i(t)$  represent the normalized diffusion strength of the neighbors of agent  $a_i$  at time step  $t$ , estimated by

$$\bar{s}_i(t) = \frac{\sum_{j \neq i} w_{ij} \times s_j(t)}{\sum_{j \neq i} w_{ij}}. \quad (3)$$

The coupling function  $f(\bar{s}_i(t), s_i(t))$  is typically used to estimate the influence of selective parameters on a decision-making process (6):

$$f(\bar{s}_i(t), s_i(t)) = rp_i(t) \times [1 - (1 - \bar{s}_i(t)) \times (1 - s_i(t))] + (1 - rp_i(t)) \times \bar{s}_i(t) \times s_i(t). \quad (4)$$

$F_{PM}(ib_i(t+1))$  and  $F_{PM}(ei_i(t+1))$  are estimated by:

$$F_{PM}(ib_i(t+1)) = \sqrt{ib_i(t-1)} [1 - (1 - I(u)) \times (1 - ei_i(t)) \times (1 - rp_i(t))], \quad (5)$$

$$F_{PM}(ei_i(t+1)) = \sqrt{ei_i(t-1)} [1 - (1 - I(u)) \times ib_i(t) \times (1 - rp_i(t))], \quad (6)$$

where  $u = (r(t))^\beta$  and  $\beta$  is a scaling factor. Higher  $RP$  and external influence  $I(u)$  denote the impact of public reports  $r(t)$  on agents, thus capturing the gradually unfolding COVID-19 events in the studied period. We give the square root to  $ib_i(t-1)$  and  $ei_i(t-1)$ , resulting in better fitting performance than not.  $I(u)$  is given by:

$$I(u) = \frac{\ln(1 + \lambda u)}{1 + \lambda}. \quad (7)$$

The  $RP$  of agents is affected by  $EI$  and the external related information and the constant  $\nu$  is used to measure the weight of these two factors.  $rp_i(t+1)$  is updated as:

$$rp_i(t+1) = (1 - \nu) F_{rp} + \nu \frac{1}{1 + e^{-\delta(ei_i(t) - \tau)}} F_{rp}, \quad (8)$$

where the value of the second part of the formula will rise rapidly when  $ei_i(t)$  exceeds the threshold  $\tau$ . The  $F_{rp}$  is used to measure the impact of external information on the  $RP$  of agent  $a_i$ .  $I(u)$  would increase individuals' risk perception and reduce uncertainty of extreme events. And pessimists (i.e.,  $pc_i$  is low) will amplify the impact of negative public news (i.e.,  $p$  is low), and vice versa.  $F_{rp}$  is therefore expressed as follows:

$$F_{rp} = rp_i(t) \times [1 - ib_i(t) \times (1 - I(u)) \times (1 - pc_i \times p - (1 - pc_i) \times (1 - p))]. \quad (9)$$

The strengthening and weakening processes are reflected by the positive and negative relationships in the above equations.

Parameter calibration. To reduce the impact of data noise, we would fit a curve  $r_f(t)$  from model to the external situation reports following ref. (2). To introduce the stochasticity into simulations,  $r_f(t)$  is based on  $r(t)$  with random noise following the uniform distribution  $u(-\phi, \phi)$ :

$$r_f(t) = r(t) + u(-\phi, \phi). \quad (10)$$

We ran 100 simulations, each with a different time series of  $r(t)$  following this sampling method.  $\beta$  is chosen as the best value of resulting in the least root mean square deviation between the normalized daily Google search behaviors in Hong Kong and the medians of normalized averaged levels of individual risk perception across all agents across 100 stochastic simulations.  $\phi$  is chosen as 0.14 by experience to balance the width of 95% CrI of individual behaviors in the model.

## REFERENCES

1. Our World in Data. Coronavirus (COVID-19) cases. <http://ourworldindata.org/covid-cases>. [2023-1-9].
2. Google. Google trends. <https://trends.google.com/trends/explore?date=2020-01-31%202020-06-28&geo=HK&q=%2Fm%2F01cppy>. [2023-1-9].
3. Cowling BJ, Ali ST, Ng TWY, Tsang TK, Li JCM, Fong MW, et al. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. *Lancet Public Health* 2020;5(5):e279 – 88. [http://dx.doi.org/10.1016/S2468-2667\(20\)30090-6](http://dx.doi.org/10.1016/S2468-2667(20)30090-6).
4. Jiang YC, Hu J, Lin DH. Decision making of networked multiagent systems for interaction structures. *IEEE Trans Syst Man Cybern Part A: Syst Humans* 2011;41(6):1107 – 21. <http://dx.doi.org/10.1109/TSMCA.2011.2114343>.
5. Gao C, Liu JM. Network-based modeling for characterizing human collective behaviors during extreme events. *IEEE Trans Syst Man Cybern: Syst* 2017;47(1):171 – 83. <http://dx.doi.org/10.1109/TSMC.2016.2608658>.
6. Lanham MJ, Morgan GP, Carley KM. Social network modeling and agent-based simulation in support of crisis de-escalation. *IEEE Trans Syst Man Cybern: Syst* 2014;44(1):103 – 10. <http://dx.doi.org/10.1109/TSMCC.2012.2230255>.

## Methods and Applications

# High-Resolution Data on Human Behavior for Effective COVID-19 Policy-Making — Wuhan City, Hubei Province, China, January 1–February 29, 2020

Jingyuan Wang<sup>1,2,3,#</sup>; Honghao Shi<sup>1</sup>; Jiahao Ji<sup>1</sup>; Xin Lin<sup>1</sup>; Huaiyu Tian<sup>4,#</sup>

## ABSTRACT

**Introduction:** High-resolution data is essential for understanding the complexity of the relationship between the spread of coronavirus disease 2019 (COVID-19), resident behavior, and interventions, which could be used to inform policy responses for future prevention and control.

**Methods:** We obtained high-resolution human mobility data and epidemiological data at the community level. We propose a metapopulation Susceptible-Exposed-Presymptomatic-Infectious-Removal (SEPIR) compartment model to utilize the available data and explore the internal driving forces of COVID-19 transmission dynamics in the city of Wuhan. Additionally, we will assess the effectiveness of the interventions implemented in the smallest administrative units (subdistricts) during the lockdown.

**Results:** In the Wuhan epidemic of March 2020, intra-subdistrict transmission caused 7.6 times more infections than inter-subdistrict transmission. After the city was closed, this ratio increased to 199 times. The main transmission path was dominated by population activity during peak evening hours.

**Discussion:** Restricting the movement of people within cities is an essential measure for controlling the spread of COVID-19. However, it is difficult to contain intra-street transmission solely through city-wide mobility restriction policies. This can only be accomplished by quarantining communities or buildings with confirmed cases, and conducting mass nucleic acid testing and enforcing strict isolation protocols for close contacts.

In the ongoing coronavirus disease 2019 (COVID-19) pandemic (1), human mobility has been identified as a key factor in the spread of the disease (2) and in

shaping its transmission dynamics (3–4). From a global perspective, cross-border travel can be used to predict the potential trajectory of global transmission (5–6). From a country-wide scale, studies have shown that human mobility from Wuhan City to other cities in China had a significant impact on the epidemics in these cities during the first wave of the outbreak (7–8). Control measures implemented in China, as well as in other countries, were successful in substantially suppressing the transmission of COVID-19. The transmission between subdistricts in a city is usually responsible for most of the disease transmission across spatial scales, but it is rarely measured (9).

We will demonstrate how different types of human mobility can affect transmission dynamics in a city. Therefore, we can identify social behaviors that are strongly associated with the epidemic trajectory in a metropolis of 10 million residents.

## METHOD

The calibration of parameters is performed with the Python (version 3.6.0, Python Software Foundation, Wilmington, US) and the Python package PyMC (version 2.3.8). The data of COVID-19 cases (high-resolution) were sourced from the large epidemic network of China Electronics Technology Group Corporation (CETC), which was obtained indirectly from the front-line hospitals and disease control departments in Wuhan. Population mobility data was derived from China Mobile's cell phone signaling records.

## Model Development

We adopt a metapopulation model to simulate the transmission of COVID-19 in Wuhan City. Supplementary Figure S1 (available in <https://weekly.chinacdc.cn/>) shows the schematic diagram of the model. Our model treats each subdistrict as a metapopulation. For each subdistrict, the model

divides the whole population into five compartments, i.e., the susceptible population  $S$ , the exposed population  $E$ , the pre-symptomatic infectious population  $P$ , the infectious population  $I$ , and the effectively removed population  $R$ . Therefore, our model is named as a metapopulation Susceptible-Exposed-Presymptomatic-Infectious-Removal (SEPIR) model. The equations of transition relationships between the five populations are given as follows.

$$\frac{dS_i}{dt} = -\sum_{b=1}^3 \left( \beta_{1,b}^i \frac{S_i}{N_i} \sum_{j=1}^l \frac{T_{bjit} \cdot C_j}{N_j} I_j \right) - \beta_2^i \frac{S_i}{N_i} I_i - \sum_{b=1}^3 \left( q\beta_{1,b}^i \frac{S_i}{N_i} \sum_{j=1}^l \frac{T_{bjit} \cdot C_j}{N_j} P_j \right) - q\beta_2^i \frac{S_i}{N_i} P_i \quad (1)$$

$$\frac{dE_i}{dt} = \sum_{b=1}^3 \left( \beta_{1,b}^i \frac{S_i}{N_i} \sum_{j=1}^l \frac{T_{bjit} \cdot C_j}{N_j} I_j \right) + \beta_2^i \frac{S_i}{N_i} I_i + \sum_{b=1}^3 \left( q\beta_{1,b}^i \frac{S_i}{N_i} \sum_{j=1}^l \frac{T_{bjit} \cdot C_j}{N_j} P_j \right) + q\beta_2^i \frac{S_i}{N_i} P_i - \alpha_e E_i \quad (2)$$

$$\frac{dP_i}{dt} = \alpha_e E_i - \alpha_p P_i \quad (3)$$

$$\frac{dI_i}{dt} = \alpha_p P_i - \gamma I_i \quad (4)$$

$$\frac{dR_i}{dt} = \gamma I_i \quad (5)$$

where  $i$  denotes the subdistrict index. The variables  $S_i$ ,  $E_i$ ,  $P_i$ ,  $I_i$ ,  $R_i$  denote the corresponding compartments' population of the  $i$ -th subdistrict. The parameter  $\beta$  is the transmission rate between susceptible and infectious populations.  $\alpha_e$  is the transition rate from the population  $E$  to  $P$ ,  $\alpha_p$  is the incidence rate, and  $\gamma$  is the removal rate.

### Parameter Setting

The transition rate  $\alpha_e$  is set as the inverse of the average period between exposure and presymptomatic infectious (the incubation period minus 2.3 days), and the incidence rate  $\alpha_p$  is set as the inverse of the average presymptomatic infectious period (2.3 days). The removal rate  $\gamma$  was dynamically set as the inverse of the average duration from symptom onset to confirmation for every day. As shown in Supplementary Figure S2 (available in <https://weekly.chinacdc.cn/>), this duration substantially reduces as a result of intervention policies.

In our model, we set the transmission rates  $\beta$  to be dynamic. Given a subdistrict  $i$ , there were four transmission rates, namely  $\beta_{1,1}^i$ ,  $\beta_{1,2}^i$ ,  $\beta_{1,3}^i$ ,  $\beta_{1,4}^i$ . For any one of the four transmission rates, denoted as  $\beta_{*}^i$ , we set it as  $\beta_{*}^i = \tilde{\beta}_{*}^i \cdot M_t^i$  on the day  $t$ , where  $\tilde{\beta}_{*}^i$  was a basic transmission rate and  $M_t^i$  was the total volume of

resident mobility in the subdistrict  $i$  on the day  $t$ . The  $M_t^i$  was calculated as  $M_t^i = (\sum_{ij} w_{ijt} + h_{ijt} + r_{ijt}) / N_i$ , where  $w_{ijt}$  was the amount of inter-subdistrict mobility from subdistrict  $i$  to subdistrict  $j$  during the morning-peak (to-workplace) period,  $h_{ijt}$  was during the evening-peak (to-home) period, and  $r_{ijt}$  was during the off-peak period.

We derived the effective reproduction number  $R_e$  of the metapopulation SEPIR model by the next generation matrix. Suppose a model with  $m$  metapopulations, let  $x = (E_1, E_2, \dots, E_m, P_1, P_2, \dots, P_m, I_1, I_2, \dots, I_m)^T$  be the number of individuals for each infected compartments,  $u_i = \frac{S_i}{N_i} \beta_{2,1}^i$ ,  $v_{ji} = \sum_b \beta_{1,b}^i \frac{S_i}{N_i} \frac{T_{bji}}{N_j} C_j$ . The detailed calculation process is shown in Supplementary Material (available in <https://weekly.chinacdc.cn/>).

### Evaluation Experiments

We adjusted the parameters of the calibrated model to estimate the effectiveness of different interventions and their interactions with the effective reproduction number  $R_e$ . Using the calibrated model parameters on January 23, 2020 as the benchmark, we adjusted the resident mobility intensity, i.e.,  $w_{ijt}$ ,  $h_{ijt}$ , and  $r_{ijt}$ , to simulate the effectiveness of the mobility restriction policy, as well as adjusted the average duration from symptom onset to isolation, i.e.,  $1/\gamma$ , to simulate the effectiveness of the policies aiming to reduce the infectious period. We construct the contour plot of  $R_e$  in Figure 1A through traversing the relative resident mobility intensity and the average duration from symptom onset to isolation to generate corresponding effective reproduction numbers.

We designed two scenarios to evaluate the effectiveness of non-pharmaceutical interventions. In the first scenario, we set the mobility volume after the Wuhan lockdown to the same level as the last day before the lockdown (January 22, 2020), while the infectious period is reduced to reflect reality. This scenario aims to simulate the condition where only interventions to reduce the infectious period are implemented (Scenario 1). Alternatively, in Scenario 2, we simulate the condition where only the intra-city mobility restriction is implemented. In this scenario, the duration from symptom onset to isolation after the Wuhan lockdown was set as 15.7 days (the average time on January 22, 2020) and the mobility volume was reduced to its lowest level.

We re-conducted experiments in Figure 1A with the

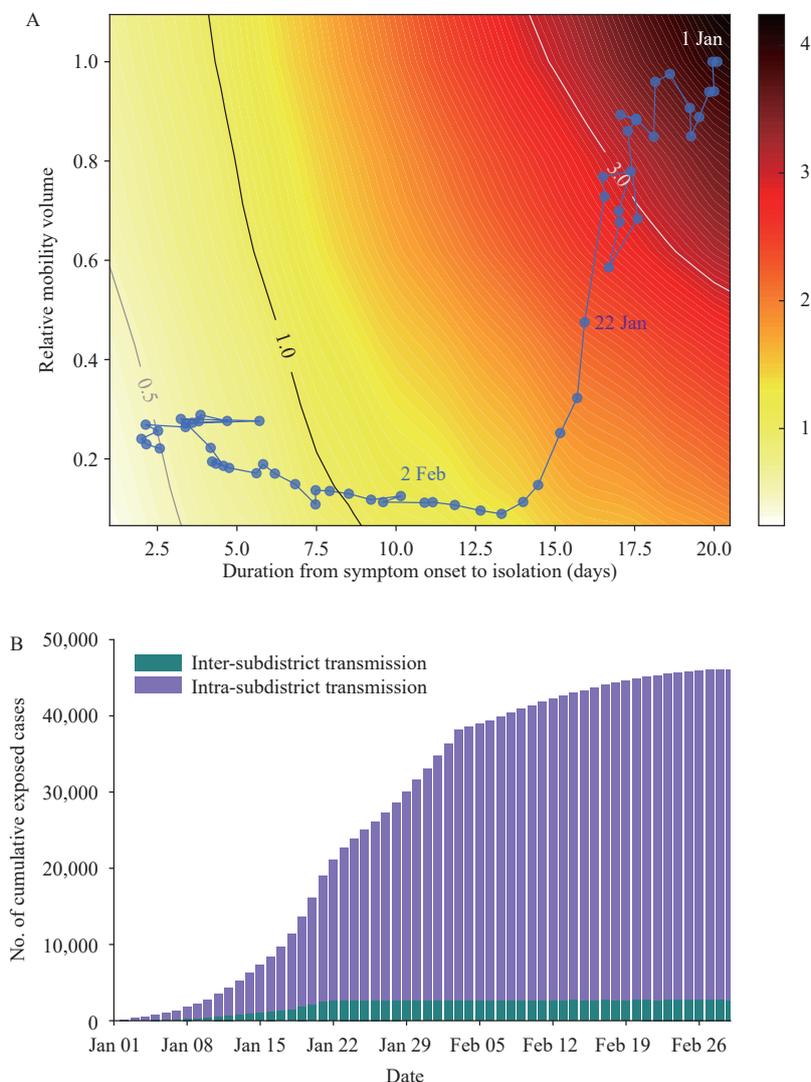


FIGURE 1. Simulation experiments on transmission dynamics and non-pharmaceutical interventions from January 1 to February 29, 2020 in Wuhan, China. (A) The contour plot between effective reproduction number  $R_e$  and the two categories of interventions implemented. (B) The number of cumulative exposed cases caused by intra- and inter-subdistrict transmissions.

Note: two categories of interventions implemented in Wuhan included mobility restriction (corresponding to relative mobility volume) and infectious period reduction (corresponding to duration from symptom onset to isolation). The color on the contour plot represents the value of  $R_e$  of corresponding relative mobility volume and duration from symptom onset to isolation. The line formed by blue dots reflects the  $R_e$  from January 1 to February 29, 2020.

parameters from Zhang et al. (10) to simulate the impact of the COVID-19 Delta variant B.1.617.2. Specifically, the incubation period was uniformly set to 4.4 days, and all the transmission rates were set to twice those fitted by the data. Supplementary Figure S3 (available in <https://weekly.chinacdc.cn/>) showed the contour plot of  $R_e$  under the Delta variant.

## RESULTS

We first analyzed the intra-city human mobility and

transmission dynamics using a multi-phase framework. The first wave of COVID-19 in Wuhan can be divided into three phases: 1) From January 1 to January 23, 2020, there were nearly no interventions; 2) On January 23, 2020, mobility restrictions were implemented; and 3) On February 3, 2020, in addition to mobility restrictions, large-scale centralized isolation policies for suspected, mild patients, and close contacts were implemented to reduce the duration of the infectious period (11).

Based on the above framework, we further analyzed

the role of resident mobility in intra- and inter-subdistrict transmission. We used the mobility network measured by cell phone data to establish a metapopulation model to simulate the spread of the disease within and between subdistricts (Supplementary Material and Supplementary Figure S1). Since different travel purposes may lead to different behaviors that could impact transmission, we further divided residents' inter-subdistrict mobility into three categories based on the hours of a day (Supplementary Figure S2C and Supplementary Figure S4, available in <https://weekly.chinacdc.cn/>): the morning-peak period (7 a.m. to 9 a.m.), the evening-peak period (4 p.m. to 6 p.m.), and the off-peak period (the remaining hours of the day). Therefore, in our model, the infection rate of a contact is determined by subdistricts, mobility type (intra- and inter-subdistrict), and mobility purpose (during morning-peak, evening-peak, and off-peak periods). Our model accurately captured the daily number of onset cases in all 99 subdistricts, with a mean absolute percentage error (MAPE) of 7.04% (see Supplementary Figures S5 and S6, available in <https://weekly.chinacdc.cn/>). We investigated the influence of intra- and inter-subdistrict mobility on COVID-19 transmission using the model. Before intra-city mobility was restricted, the volume of intra- and inter-subdistrict mobility was 71.9% and

28.1%, respectively. This indicates that intra-subdistrict mobility was about 2.5 times higher than inter-subdistrict mobility. Our model indicated that the number of infections caused by intra-subdistrict transmission in the first phase was 20,011 [95% confidence interval (CI): 18,556–21,967], which was approximately 7.6 times (95% CI: 6.9–8.4) that caused by inter-subdistrict transmission (2,650, 95% CI: 2,209–3,164).

We analyzed the relationship between inter- and intra-subdistrict transmission. In the second phase, the inter-subdistrict mobility was suppressed by 98.9% due to mobility restrictions (Figure 2), resulting in almost complete termination of the inter-subdistrict transmission (Figure 1B). The intra-subdistrict mobility decreased by 84.0% (Figure 2), yet the intra-subdistrict transmission persisted until the centralized isolation policy was implemented (Figure 1B). Our model showed that the intra-subdistrict transmission caused 23,321 (95% CI: 21,097–25,350) infections after the mobility restrictions, accounting for 99.5% (95% CI: 99.5%–99.5%) of the new infections. According to individual-level clinical data, the average time from onset to isolation of a case was more than 15 days in the first phase, reducing to less than 3 days in the third phase (Supplementary Figure S2).

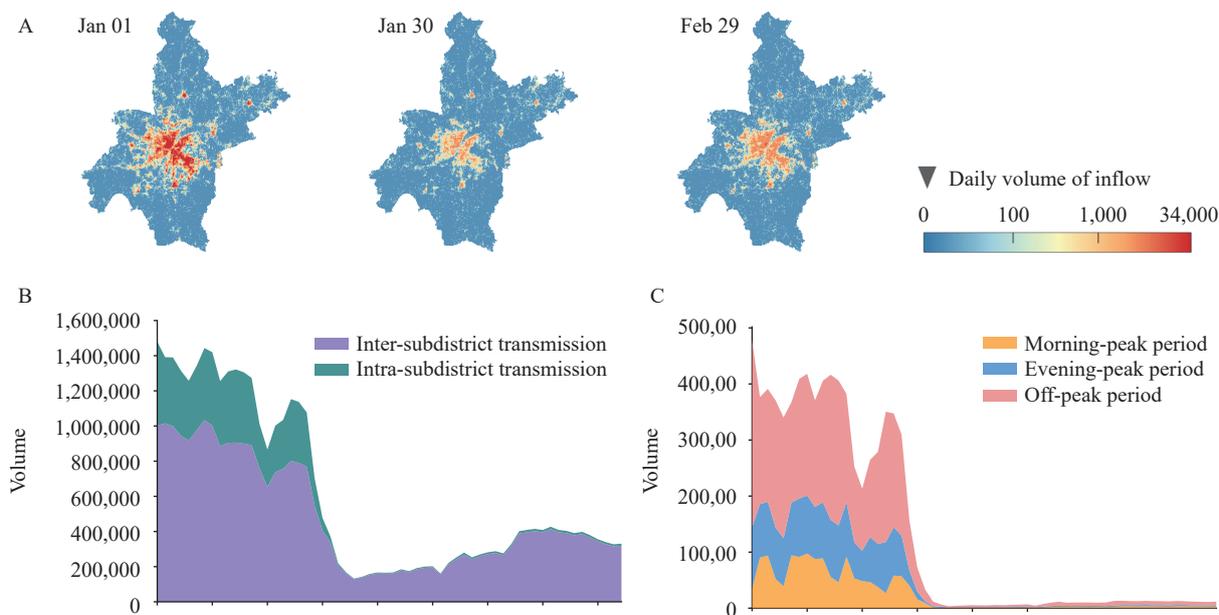


FIGURE 2. Dynamics of intra-city mobility from January 1 to February 29, 2020 in Wuhan, China. (A) The heatmaps of average daily inflow mobility volume for each subdistrict in Wuhan with different dates among the three phases. (B) Changes of intra- and inter-subdistrict mobility volume in Wuhan during the outbreak. (C) Changes of inter-subdistrict mobility during different peak periods of one day.

## DISCUSSION

Since 2021, the COVID-19 Delta variant B.1.617.2 has spread rapidly around the world, posing a serious challenge to containing the pandemic. We designed a scenario to simulate the case of the Delta variant transmitting in Wuhan in early 2020, and investigated the impacts of the interventions discussed above using parameters obtained from Zhang et al. (10) (Supplementary Material). Under this scenario, mobility restriction alone is unable to reduce  $R_e$  to 1 due to the high transmissibility of the Delta variant, and containment of the epidemic could only be achieved by a 30% relative mobility volume, together with a short infectious period (less than 2.5 days) (Supplementary Figure S3). There would be an estimated 3.81 million (95% CI: 3.54–4.02 million) cases as of March 1, 2020, if the same interventions were implemented in Wuhan under the Delta variant. This result indicates the difficulty of containing this new variant, and underscores the importance of reducing the infectious period.

Our work also investigated the effectiveness of non-pharmaceutical interventions implemented in Wuhan. Although travel restrictions could reduce the number of new cases in the short term, they were not sufficient to terminate transmission. Strict isolation policies in exchange for a relaxation of traffic control have been helpful in restoring the economy damaged by the epidemic. In fact, this was the policy that the Chinese government adopted to reduce the spread of the virus. During several rounds of cluster outbreaks after May 2020 in China, the government blocked communities and buildings with confirmed cases, implemented large-scale nucleic acid tests, and enforced strict isolation policies to reduce the duration of the infection (12–13). Comprehensive and precise control measures can contain the outbreak while minimizing its impact on people's daily lives and the economy.

In summary, we completed a review of the Wuhan COVID-19 outbreak using a refined metapopulation model. Based on this, we can make counterfactual inferences about policies that are more beneficial for decision-making in advance than the predictions and analyses of similar work (10,14).

However, the metapopulation model used in this work has limitations in terms of generalizability. First, this model requires high-quality raw data and refined population flow data. Second, the number of parameters is large, and when the preset parameters are significantly different from the original data, effective

fitting cannot be achieved.

**Conflicts of interest:** No conflicts of interest.

**Funding:** Supported by the National Key R&D Program of China (2021ZD0111201) and National Natural Science Foundation of China (No. 82161148011, 72222022, 72171013).

doi: 10.46234/ccdcw2023.015

# Corresponding authors: Jingyuan Wang, jywang@buaa.edu.cn; Huaiyu Tian, tianhuaiyu@gmail.com.

<sup>1</sup> School of Computer Science and Engineering, Beihang University, Beijing, China; <sup>2</sup> Pengcheng Laboratory, Shenzhen City, Guangdong Province, China; <sup>3</sup> School of Economics and Management, Beihang University, Beijing, China; <sup>4</sup> State Key Laboratory of Remote Sensing Science, Center for Global Change and Public Health, College of Global Change and Earth System Science, Beijing Normal University, Beijing, China.

Submitted: December 24, 2022; Accepted: January 16, 2023

## REFERENCES

1. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. *Nature* 2020;579(7798):265–9. <http://dx.doi.org/10.1038/s41586-020-2008-3>.
2. Wells CR, Sah P, Moghadas SM, Pandey A, Shoukat A, Wang YN, et al. Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proc Natl Acad Sci USA* 2020;117(13):7504–9. <http://dx.doi.org/10.1073/pnas.2002616117>.
3. Changruenngam S, Bicout DJ, Modchang C. How the individual human mobility spatio-temporally shapes the disease transmission dynamics. *Sci Rep* 2020;10(1):11325. <http://dx.doi.org/10.1038/s41598-020-68230-9>.
4. Xiong CF, Hu SH, Yang MF, Luo WY, Zhang L. Mobile device data reveal the dynamics in a positive relationship between human mobility and COVID-19 infections. *Proc Natl Acad Sci USA* 2020;117(44):27087–9. <http://dx.doi.org/10.1073/pnas.2010836117>.
5. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* 2020;395(10225):689–97. [http://dx.doi.org/10.1016/S0140-6736\(20\)30260-9](http://dx.doi.org/10.1016/S0140-6736(20)30260-9).
6. Nouvellet P, Bhatia S, Cori A, Ainslie KEC, Baguelin M, Bhatt S, et al. Reduction in mobility and COVID-19 transmission. *Nat Commun* 2021;12(1):1090. <http://dx.doi.org/10.1038/s41467-021-21358-2>.
7. Jia JS, Lu X, Yuan Y, Xu G, Jia JM, Christakis NA. Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* 2020;582(7812):389–94. <http://dx.doi.org/10.1038/s41586-020-2284-y>.
8. Tian HY, Liu YH, Li YD, Wu CH, Chen B, Kraemer MUG, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* 2020;368(6491):638–42. <http://dx.doi.org/10.1126/science.abb6105>.
9. Lee EC, Wada NI, Grabowski MK, Gurley ES, Lessler J. The engines of SARS-CoV-2 spread. *Science* 2020;370(6515):406–7. <http://dx.doi.org/10.1126/science.abd8755>.
10. Zhang M, Xiao JP, Deng AP, Zhang YT, Zhuang YL, Hu T, et al. Transmission Dynamics of an Outbreak of the COVID-19 Delta Variant B.1.617.2—Guangdong province, China, May–June 2021. *China CDC Wkly* 2021;3(27):584–6. <http://dx.doi.org/10.46234/ccdcw2021.148>.
11. Pan A, Liu L, Wang CL, Guo H, Hao XJ, Wang Q, et al. Association of

- public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. *JAMA* 2020;323(19):1915 – 23. <http://dx.doi.org/10.1001/jama.2020.6130>.
12. Xing YH, Wong GWK, Ni W, Hu XW, Xing QS. Rapid response to an outbreak in Qingdao, China. *N Engl J Med* 2020;383(23):e129. <http://dx.doi.org/10.1056/NEJMc2032361>.
  13. Li ZJ, Liu FF, Cui JZ, Peng ZB, Chang ZR, Lai SJ, et al. Comprehensive large-scale nucleic acid-testing strategies support China's sustained containment of COVID-19. *Nat Med* 2021;27(5):740 – 2. <http://dx.doi.org/10.1038/s41591-021-01308-7>.
  14. 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>.

## SUPPLEMENTARY MATERIAL

### Epidemiological and Demographic Data of Wuhan

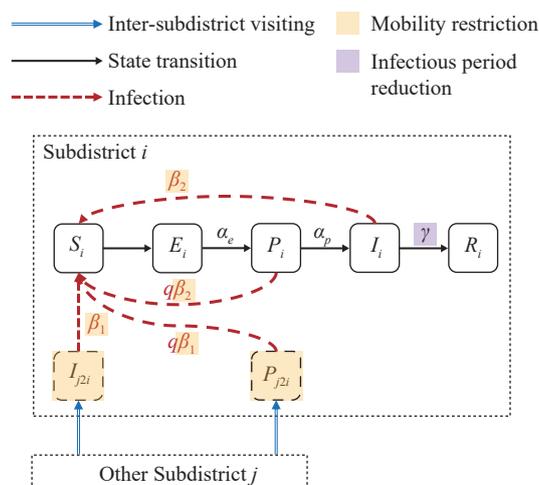
The epidemiological data of Wuhan were extracted from the Notifiable Disease Report System of China. This study used anonymous individual coronavirus disease 2019 (COVID-19) case data, including residential subdistrict, date of onset, and date of confirmation, from January 1, 2020 to March 1, 2020, for analysis. The demographic data for the subdistricts in Wuhan City, including population and geographical boundaries, were obtained from the Sixth Census conducted by the National Bureau of Statistics. We matched the epidemiological cases to the subdistricts. A total of 161 subdistricts with COVID-19 cases were used for further analysis. Since there were subdistricts with insufficient cases of COVID-19 to be modeled, we merged the epidemiological and demographic data of these subdistricts into the geographically closest subdistricts. Lastly, there are  $n=99$  subdistricts for model simulation.

### Proxies for Human Mobility Data in Wuhan

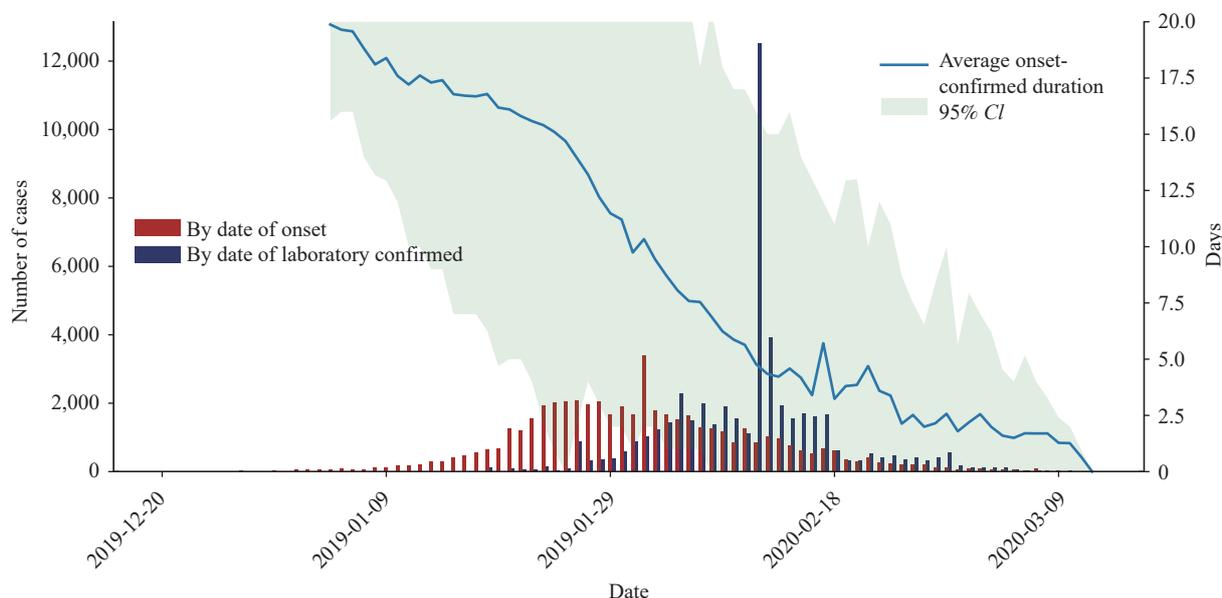
We used cell phone signaling data as a proxy to measure population mobility in Wuhan during the epidemic. The anonymous cell phone mobility data, provided by a major mobile carrier in China, covered approximately 51.9% (5.82 million/11.21 million) of the population in Wuhan. The raw cell phone signaling data records the visiting trajectories of cell phone users at each cellular base station. We integrated the raw data as travel flow of phone users between  $500\text{ m} \times 500\text{ m}$  grids for each hour. We further integrated the data as travel flow between subdistricts by merging the flows of grids in the same subdistrict together.

### Periods Division of Residents' Mobility in One Day

Supplementary Figure S4 illustrates the average hourly volume of inter-subdistrict mobility on workdays prior to January 23, 2020. As shown in the figure, there is a morning peak at 8 a.m. and an evening peak at 5 p.m., reflecting the temporal rhythm pattern of residents' mobility behaviors on workdays. Based on this, we classified residential mobility in one day into three categories based on the time of departure. The first category is the mobility from 7 a.m. to 9 a.m., i.e., the morning rush hour when people commute to work from their homes. The second category is the mobility from 4 p.m. to 6 p.m., i.e., the evening peak period when people are returning home from their workplaces. The last one is mobility during off-peak periods, excluding morning and evening rush hours. The mobility during the off-peak period is relatively random.

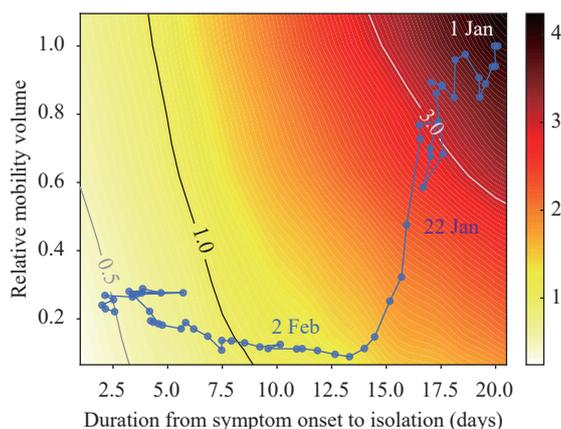


SUPPLEMENTARY FIGURE S1. Schematic diagram of the metapopulation SEPIR model. Abbreviation: SEPIR=Susceptible-Exposed-Presymptomatic-Infectious-Removal



SUPPLEMENTARY FIGURE S2. COVID-19 cases by date of symptom onset and by date of diagnosis from January 1 to February 29, 2020 in Wuhan, China.

Note: Changes in the average duration between symptom onset and laboratory-confirmed. Abbreviation: COVID-19=coronavirus disease 2019; CI=confidence interval.

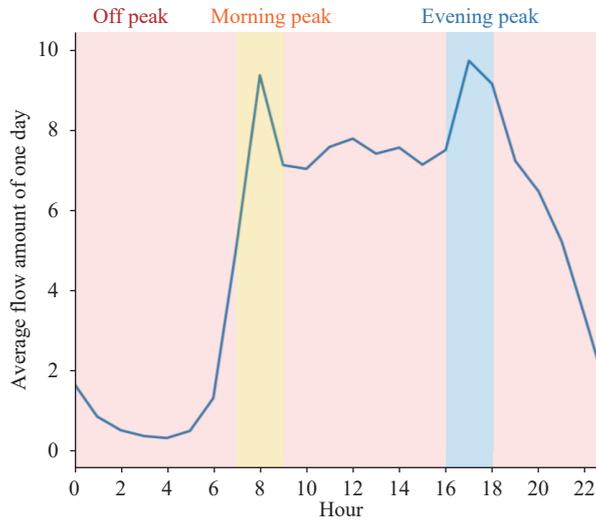


SUPPLEMENTARY FIGURE S3. Simulation experiments on transmission dynamics and non-pharmaceutical interventions under the Delta variant.

Note: The contour plot between effective reproduce number  $R_e$  and the two categories of interventions implemented in Wuhan, i.e., mobility restriction (corresponding to relative mobility volume) and infectious period reduction (corresponding to duration from symptom onset to isolation) with the parameter of Delta variant. The color on the contour plot represents the value of  $R_e$  of corresponding relative mobility volume and duration from symptom onset to isolation. The line formed by blue dots reflects the  $R_e$  from January 1 to February 29, 2020 in Wuhan.

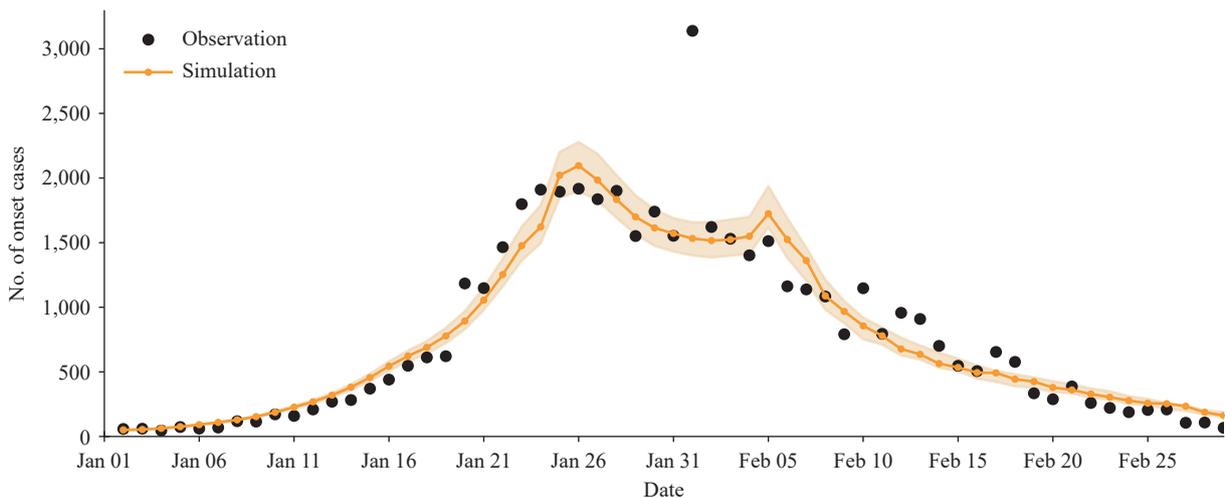
### Metapopulation SEPIR Model

In order to study the impacts of different patterns of resident mobility on intra-city epidemic transmission, our model refines the transmission process into two parts, namely the intra-subdistrict transmission and the inter-subdistrict transmission, and further divides the inter-subdistrict transmission into three categories, i.e., transmission in the evening-peak, morning-peak, and off-peak periods. As shown in Formula 1, for the intra-subdistrict transmission, the number of newly exposed population for metapopulation  $i$  in one day is  $\beta \frac{S_i}{N_i} I_i$ , which is the same as the definition of the standard SEIR model. For the inter-subdistrict transmission, the increment of the exposed



SUPPLEMENTARY FIGURE S4. Changes in the average amount of inter-subdistrict traffic at different times of the workday before January 23, 2020.

Note: Shaded regions in different colors denote the split of time.



SUPPLEMENTARY FIGURE S5. Observation and model simulation of onset cases for Wuhan City.

Note: The shadowed regions represent the 95% confidence interval of model simulation.

population in the metapopulation  $i$  caused by the inflow mobility from the metapopulation  $j$  is expressed as  $\beta \frac{S_i}{N_i} I_{j2i}$ , where  $I_{j2i} = \frac{T_{hjit} \cdot C_j}{N_j} I_j$ , i.e., the infectious population traveled from the metapopulation  $j$  to  $i$ , is calculated using  $I_j$  and scaled by human mobility data. Here,  $N_j$  represents the population of subdistrict  $j$ , which is obtained from the census data.  $T_{hjit}$  is the amount of inter-subdistrict mobility from the subdistrict  $j$  to  $i$  in the period  $h$  of the day  $t$ , where  $h=1$  for the morning-peak,  $h=2$  for the evening-peak, and  $h=3$  for the off-peak period. The parameter  $C_i$  the ratio of  $N_i$  and the number of cell phone users in subdistrict  $i$ , which is used to calibrate the mismatch between cell phone users and the population.

As different patterns of mobility should have different effects to transmission of COVID-19, we set the transmission rate  $\beta$  as four types in Formula 1. Specifically,  $\beta_{1,1}^i, \beta_{1,2}^i, \beta_{1,3}^i$  denote the transmission rates for the inter-subdistrict transmission in the evening-peak, morning-peak, and off-peak periods in the metapopulation  $i$ , respectively, while  $\beta_2^i$  denotes the transmission rate of intra-subdistrict transmissions in the metapopulation  $i$ . The

presymptomatic infectious population may have different infectiousness with infectious population ( $I$ ), we multiply  $\beta$  with a factor  $q$  for the transmission between presymptomatic infectious population and susceptible population.

### Parameter Setting, Calibration and Epidemic Dynamic Simulation

In the epidemiological data, the original record of each infected case includes two dates: the date of symptom onset and the date of laboratory confirmation. We sampled an incubation period from a Weibull distribution, as reported in a previous study (2). By using the sampled incubation period and the symptomatic onset date, we can approximate the exposure date for each case. Moreover, we set the last 2.3 days of the incubation period as the presymptomatic infectious period according to previous studies (3). In this way, the timeline for an infected case is divided as five periods, i.e., the **S**usceptible period (before the date of exposure), the **E**xposed period (from the date of exposure to 2.3 days before the date of onset), the **P**resymptomatic infectious periods (the last 2.3 days before the date of onset), the **I**nfectious periods (from the date of onset to the date of confirmation), and the **R**emoval periods (after the date of confirmation). We set a confirmed case as a removed one since all infected persons will be immediately quarantined once they get confirmation in China and therefore would not cause secondary infections anymore. We calculate the size of population  $E_i, P_i, I_i, R_i$  in Formula (1) using the number of cases on each day for each subdistrict  $i$ , and calculate the size of the susceptible population as  $S_i = N_i - E_i - P_i - I_i - R_i$ .

In Formula (1), the transition rate  $\alpha_e$  is set as the inverse of the average period between exposure and presymptomatic infectious (the incubation period minus 2.3 days), and the incidence rate  $\alpha_p$  is set as the inverse of the average presymptomatic infectious period (2.3 days). The removal rate  $\gamma$  is dynamically set as the inverse of the average duration from symptom onset to confirmation for every day. As shown in Supplementary Figure S2, this duration substantially reduces as a result of intervention policies.

In our model, we set the transmission rates  $\beta$  in a dynamic way. Given a subdistrict  $i$ , there are four transmission rates, namely  $\beta_{1,1}^i, \beta_{1,2}^i, \beta_{1,3}^i, \beta_2^i$ . For any one of the four transmission rates, denoted as  $\beta_*^i$ , we set it as  $\beta_*^i = \hat{\beta}_*^i \cdot M_t^i$  on the day  $t$ , where  $\hat{\beta}_*^i$  is a basic transmission rate and  $M_t^i$  is the total volume of resident mobility in the subdistrict  $i$  on the day  $t$ . The  $M_t^i$  is calculated as  $M_t^i = \frac{\sum_{ij} w_{ijt} + h_{ijt} + r_{ijt}}{N_i}$ , where  $w_{ijt}$  is the amount of inter-subdistrict mobility from subdistrict  $i$  to subdistrict  $j$  during the morning-peak (to-workplace) period,  $h_{ijt}$  is during the evening-peak (to-home) period, and  $r_{ijt}$  is during the off-peak period.

The basic transmission rates  $\hat{\beta}_{1,1}^i, \hat{\beta}_{1,2}^i, \hat{\beta}_{1,3}^i, \hat{\beta}_2^i$ , and the presymptomatic infectiousness discount factor  $q$  are calibrated by the Metropolis-Hastings Markov Chain Monte Carlo (MCMC) algorithm (4), with the state  $P_i, I_i, R_i$  for each day as supervisions. The process of the parameter generation is performed separately for each subdistrict and for three phases. For each phase, after a burn-in of 1,000 iterations, we run the MCMC simulation for 10,000 times, with sampling at every 50th step. The average root mean square error (RMSE) for each subdistrict is 4.35, and the simulation results are shown in Supplementary Figure S5. The calibration of parameters is performed with the Python (version 3.6.0, Python Software Foundation, Wilmington, US) and the Python package *PyMC* (version 2.3.8) (5).

### Estimation of Effective Reproduce Number from Model Parameters

We derive the effective reproduce number  $R_e$  of the metapopulation SEPIR model by the next generation matrix (6). Suppose a model with  $m$  metapopulations, let  $x = (E_1, E_2, \dots, E_m, P_1, P_2, \dots, P_m, I_1, I_2, \dots, I_m)^T$  be the number of individuals for each infected compartments,  $u_i = \frac{S_i}{N_i} \beta_{2,1}^i, v_{ji} = \sum_b \beta_{1,b}^i \frac{S_i}{N_i} \frac{T_{bji}}{N_j} c_j$ .

Furthermore, we have

$$\frac{dx_i}{dt} = F_i(x) - V_i(x)$$

Where  $F_i(x)$  is the rate of generating new infections in the  $i$ -th compartments of vector  $x$ ,  $V_i(x)$  is the transition rate of infections in the  $i$ -th compartments of vector  $x$  by all other means,  $F(x), V(x) \in R^{3m \times 1}$ , and based on the ordinary differential equations in Formula (1), we can derive the formulation of  $F(x)$  and  $V(x)$  as

$$F(x) = ([F_E(x)], [F_P(x)], [F_I(x)])^T$$

With  $F_E(x) = [\sum_{j \neq i} qv_{ji}P_j + qu_iP_i + \sum_{j \neq i} v_{ji}I_j + u_iI_i]_{i=1}^m$ ,  $F_P(x) = F_I(x) = O^{m \times 1}$ ,  
 $V(x) = ([V_E(x)], [V_P(x)], [V_I(x)])^T$

With  $V_E(x) = [\alpha_e E_i]_{i=1}^m$ ,  $V_P(x) = [-\alpha_p P_i + \alpha_p P_i]_{i=1}^m$ ,  $V_I(x) = [-\alpha_p P_i + \gamma I_i]_{i=1}^m$ .

Next, we have the matrix

$$F(x) = \left[ \frac{\partial F_i(x)}{\partial x_j} \right] = \left[ \frac{\partial F_E^i(x)}{\partial E_j} \quad \frac{\partial F_E^i(x)}{\partial P_j} \quad \frac{\partial F_E^i(x)}{\partial I_j} \quad \frac{\partial F_P^i(x)}{\partial E_j} \quad \frac{\partial F_P^i(x)}{\partial P_j} \quad \frac{\partial F_P^i(x)}{\partial I_j} \quad \frac{\partial F_I^i(x)}{\partial E_j} \quad \frac{\partial F_I^i(x)}{\partial P_j} \quad \frac{\partial F_I^i(x)}{\partial I_j} \right]$$

Where  $\left[ \frac{\partial F_E^i(x)}{\partial P_j} \right] = \{qu_i, i = jqv_{ji}, i \neq j, \left[ \frac{\partial F_E^i(x)}{\partial I_j} \right] = \{u_i, i = jv_{ji}, i \neq j, \text{ and other sub-matrixes are } O^{m \times m}$ .

Similarly, we have

$$V(x) = \left[ \frac{\partial V_i(x)}{\partial x_j} \right] = \left[ \frac{\partial V_E^i(x)}{\partial E_j} \quad \frac{\partial V_E^i(x)}{\partial P_j} \quad \frac{\partial V_E^i(x)}{\partial I_j} \quad \frac{\partial V_P^i(x)}{\partial E_j} \quad \frac{\partial V_P^i(x)}{\partial P_j} \quad \frac{\partial V_P^i(x)}{\partial I_j} \quad \frac{\partial V_I^i(x)}{\partial E_j} \quad \frac{\partial V_I^i(x)}{\partial P_j} \quad \frac{\partial V_I^i(x)}{\partial I_j} \right] =$$

$$\left[ [\alpha_e 1]^{m \times m} \quad O^{m \times m} \quad O^{m \times m} \quad [-\alpha_p 1]^{m \times m} \quad [\alpha_p 1]^{m \times m} \quad O^{m \times m} \quad O^{m \times m} \quad [-\alpha_p 1]^{m \times m} \quad [\gamma 1]^{m \times m} \right]$$

and

$$V^{-1}(x) = \left[ [\alpha_e^{-1} \cdot 1]^{m \times m} \quad O^{m \times m} \quad O^{m \times m} \quad [\alpha_p^{-1} \cdot 1]^{m \times m} \quad [\alpha_p^{-1} \cdot 1]^{m \times m} \quad O^{m \times m} \quad [\gamma^{-1} \cdot 1]^{m \times m} \quad [\gamma^{-1} \cdot 1]^{m \times m} \quad [\gamma^{-1} \cdot 1]^{m \times m} \right]$$

Where 1 denotes identity matrix.

Based on this, we can derive the next generation matrix for the metapopulation SEPIR model as

$$FV^{-1} = [ABCDEFGH I]$$

Where  $A = \left\{ \frac{qu_i}{\alpha_p} + \frac{u_i}{\gamma}, i = j \frac{qv_{ji}}{\alpha_p} + \frac{v_{ji}}{\gamma}, i \neq j, B = \left\{ \frac{qu_i}{\alpha_p} + \frac{u_i}{\gamma}, i = j \frac{qv_{ji}}{\alpha_p} + \frac{v_{ji}}{\gamma}, i \neq j, C = \left\{ \frac{u_i}{\gamma}, i = j \frac{v_{ji}}{\gamma}, i \neq j, \text{ and other sub-matrixes equal to } O^{m \times m} \right. \right.$

Finally, by Driessche and Watmough (6), the effective reproduce number  $R_e$  can be derived as

$$R_e = \rho(FV^{-1}),$$

Where  $\rho(A)$  represents the spectral radius of a matrix  $A$ . According to the property of matrix computation, this is equivalent to the maximum of absolute eigenvalues of the matrix

$$A = \left( \frac{q}{\alpha_p} + \frac{1}{\gamma} \right) [u_1 \cdots v_m; \vdots; v_1 \cdots u_m].$$

### Evaluate on the Effectiveness of Non-Pharmaceutical Intervention

We adjust the parameters of the calibrated model to estimate the effectiveness of different interventions and their interactions to the effective reproduce number  $R_e$ . Using the calibrated model parameters on January 23, 2020 as the benchmark, we adjust the resident mobility intensity, i.e.,  $w_{ijt}$ ,  $h_{ijt}$ , and  $r_{ijt}$ , to simulate the effectiveness of the mobility restriction policy, as well as adjust the average duration from symptom onset to isolation, i.e.,  $\frac{1}{\gamma}$ , to simulate the effectiveness of the infectious period reduction policies. We construct the contour plot of  $R_e$  in Figure 2 through traversing the relative resident mobility intensity and average duration from symptom onset to isolation to generate corresponding effective reproduction numbers.

We design two scenarios to evaluate the effectiveness of non-pharmaceutical interventions. In the first scenario, we set mobility volume after the Wuhan lockdown to be the same as the last day before the lockdown (January 22, 2020), while the infectious period declines as the reality. This scenario is set to simulate the condition where only the interventions to reduce the infectious period are implemented, which is called *Scenario 1*. Oppositely, in the second scenario (*Scenario 2*), we simulate the condition where only the intra-city mobility restriction is implemented, where the duration from symptom onset to isolation after the Wuhan lockdown is set as 15.7 days (the average time on January 22, 2020) in the model, and the mobility volume drops to its lowest level as the reality.

To simulate the impact of the COVID-19 Delta variant B.1.617.2, we reconducted the experiments in Figure 2 with the parameters from Zhang et al. (7). Specifically, the incubation period was set to 4.4 days uniformly, and all the transmission rates were set to 2 times of those fitted by data. Supplementary Figure S3 shows the contour plot of  $R_e$  under the Delta variant.

## REFERENCES

1. He X, Lau EHY, Wu P, Deng XL, Wang J, Hao XX, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med* 2020;26(5):672 – 5. <http://dx.doi.org/10.1038/s41591-020-0869-5>.
2. Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020. *Euro Surveill* 2020;25(5):2000062. <http://dx.doi.org/10.2807/1560-7917.ES.2020.25.5.2000062>.
3. 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>.
4. Haario H, Saksman E, Tamminen J. An adaptive Metropolis algorithm. *Bernoulli* 2001;7(2):223 – 42. <http://dx.doi.org/10.2307/3318737>.
5. Patil A, Huard D, Fonnesbeck CJ. PyMC: Bayesian stochastic modelling in python. *J Stat Softw* 2010;35(4):1-81. <https://pubmed.ncbi.nlm.nih.gov/21603108/>.
6. van den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math Biosci* 2002;180(1 – 2):29 – 48. [http://dx.doi.org/10.1016/S0025-5564\(02\)00108-6](http://dx.doi.org/10.1016/S0025-5564(02)00108-6).
7. Zhang M, Xiao JP, Deng AP, Zhang YT, Zhuang YL, Hu T, et al. Transmission dynamics of an outbreak of the COVID-19 Delta variant B.1.617.2 — Guangdong province, China, May–June 2021. *China CDC Wkly* 2021;3(27):584-6. <http://dx.doi.org/10.46234/ccdcw2021.148>.

## Methods and Applications

# Optimization of Population-Level Testing, Contact Tracing, and Isolation in Emerging COVID-19 Outbreaks: a Mathematical Modeling Study — Tonghua City and Beijing Municipality, China, 2021–2022

Zengmiao Wang<sup>1,✉</sup>; Ruixue Wang<sup>2,✉</sup>; Peiyi Wu<sup>1</sup>; Bingying Li<sup>1</sup>; Yidan Li<sup>1</sup>;  
Yonghong Liu<sup>3</sup>; Xiaoli Wang<sup>3</sup>; Peng Yang<sup>3</sup>; Huaiyu Tian<sup>1,✉</sup>

## ABSTRACT

**Introduction:** The transmissibility of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Omicron variant poses challenges for the existing measures containing the virus in China. In response, this study investigates the effectiveness of population-level testing (PLT) and contact tracing (CT) to help curb coronavirus disease 2019 (COVID-19) resurgences in China.

**Methods:** Two transmission dynamic models (i.e. with and without age structure) were developed to evaluate the effectiveness of PLT and CT. Extensive simulations were conducted to optimize PLT and CT strategies for COVID-19 control and surveillance.

**Results:** Urban Omicron resurgences can be controlled by multiple rounds of PLT, supplemented by CT — as long as testing is frequent. This study also evaluated the time needed to detect COVID-19 cases for surveillance under different routine testing rates. The results show that there is a 90% probability of detecting COVID-19 cases within 3 days through daily testing. Otherwise, it takes around 7 days to detect COVID-19 cases at a 90% probability level if biweekly testing is used. Routine testing applied to the age group 21–60 for COVID-19 surveillance would achieve similar performance to that applied to all populations.

**Discussion:** Our analysis evaluates potential PLT and CT strategies for COVID-19 control and surveillance.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is one of the most dangerous infectious diseases of the 21st century. Its rapid and global

emergence is due in part to its large reproduction number as well as its significant levels of transmission by pre-symptomatic and asymptomatic hosts (1). Undetected asymptomatic cases are dangerous because they may trigger flare-ups that circulate in the community (2). All of this was greatly exacerbated by Omicron, a variant that emerged in 2021 with a high degree of transmissibility (3). To effectively identify asymptomatic infections and prevent rampant disease transmission, it is critical to broadly test all at-risk communities (4).

SARS-CoV-2 testing has been emphasized since the beginning of 2020. Although many studies showed the positive impacts of testing on coronavirus disease 2019 (COVID-19) control (5–9), they initially primarily examined cost-effective rapid antigen testing (5), mass-testing methods that only cover 5% of the population (6), routine PCR testing for specific subsets of at-risk groups [e.g., health workers (7) or quarantined persons (9)], symptomatic cases (8), and the effect of testing on reducing quarantine lengths (10). However, as knowledge increased about SARS-CoV-2, testing was expanded to cover a broader population: testing to detect symptoms (e.g., fever), testing regardless of symptoms, community-testing (11), population-level testing (12), and mass-testing (6).

Many countries employed community-level and/or population-level testing to better prevent COVID-19 transmission. In England, 8 rounds of community-level PCR testing were carried out to investigate symptom profiles at different ages (11). Slovakia conducted population-wide rapid antigen testing and found that two rounds of testing reduced the prevalence of COVID-19 by 58% (12). However, the investigation of PLT on the suppression of COVID-19 flare-ups has been scant, especially for the Omicron variant. Preventing COVID-19 resurgence is a moving question in the face of emerging variants and the many

possible interventions.

The COVID-19 resurgences in Tonghua City, Jilin Province (B.1.1 variant) and Beijing Municipality (Omicron variant) provide a valuable opportunity to study the effectiveness of PLT and CT, as multiple rounds of PLT and CT were performed in both cities. PLT and CT facilitated fast case identification and alleviated the effects of underreporting in China. With these features in this dataset and transmission-dynamic models of infectious diseases, the strategies for COVID-19 control and surveillance are quantified.

## METHODS

### Data Collection

The daily infection data were collected from the Beijing and Tonghua health commission websites. The population size was obtained from the local Statistics Bureau or government census data.

### Transmission Models

To evaluate the effectiveness of population-level testing (PLT) and contact tracing (CT) strategies, two transmission models were developed in this study. First, a transmission model incorporating PLT and CT was introduced to set up the context of modeling. Then, the model was extended to take age structure into account. Please see the Supplementary Materials (available in <http://weekly.chinacdc.cn/>) for more details on this.

### Modeling the Probability of Detecting COVID-19 Cases for Surveillance Under Routine Testing

The probability of detecting COVID-19 cases under routine testing is a function of the sensitivity of PCR tests, the testing rate, and the particularized dynamics of the outbreak. To model the dynamic of an outbreak, an extended Susceptible-Exposed-Infectious-Removed (SEIR) model was developed. Please see the Supplementary Materials for more details on this.

### Simulation Set-up

A PLT strategy was one of the combinations: the time lag between the date of the first case identified and the date of the PLT launched (ranging anywhere from 1 to 7 days), population-level testing intervals (i.e. the time to complete 1 round of testing, including sample collection and reporting results, set to be either

2, 3, 4, 5, 7, or 10 days) and the break intervals between sequential rounds of testing (set to be either 0, 1, or 2 days) — which are often needed as a break for the testing staff. The total days considered to complete 1 round of testing was the testing interval plus the break interval. The outbreak duration divided by the total days needed to complete 1 round of testing was calculated as the rounds of PLT.

## RESULTS

### COVID-19 Resurgence in Tonghua and Beijing

Resurgences of SARS-CoV-2 occurred in Tonghua (B.1.1 variant) in Jan-Feb of 2021, and Beijing (Omicron variant) in April-Jun of 2022. Once an index case was identified, CT was launched. To rapidly detect SARS-CoV-2 infections, the cities launched population-level PCR tests. To contain the transmission, Tonghua performed 3 rounds of testing — whereas Beijing conducted 26 rounds. After multiple rounds of PLT, there were no new cases reported, with the recurrence ultimately seeing 318 cases in Tonghua and 2,230 cases in Beijing.

### The Population-level Testing and Contact Tracing Model Without Age Structure

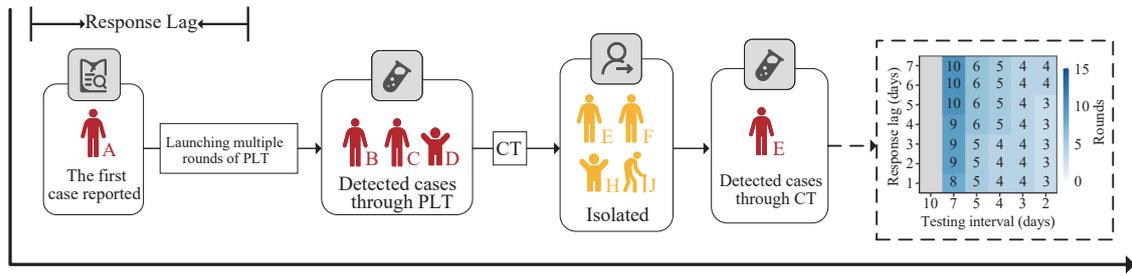
A transmission-dynamic model without age structure (Figure 1) was fitted using the daily new infections identified from both PLT and CT in Tonghua. The model assumed all infected individuals were identified either from CT or PLT, and were then quarantined and removed from the transmission chain.

Using the estimated parameters for Tonghua, this study evaluated the effectiveness of different PLT strategies in the containment of COVID-19 flare-ups. The required number of rounds of testing increased with the decreasing success fraction of CT if the testing interval remained unchanged (Figure 2A). If the success fraction of tracing remained unchanged, decreasing the testing interval not only reduced the necessary number of testing rounds, but also shortened the outbreak duration (Figure 2B).

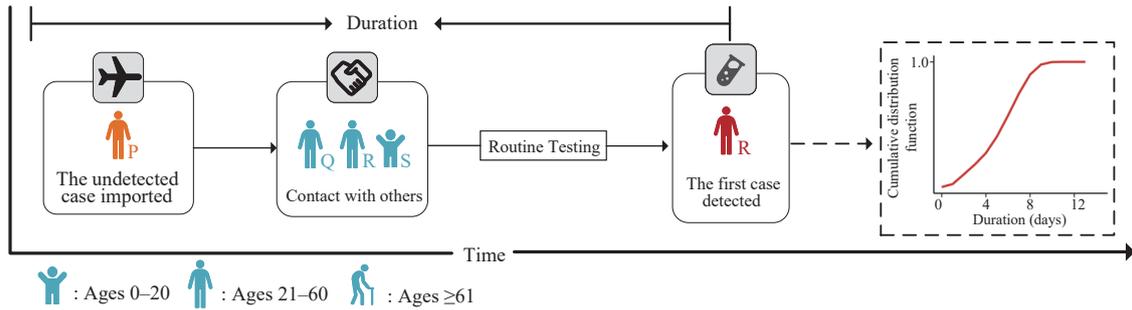
Our analyses show that the time lag, the testing interval, and the break interval have important effects on outbreak control. For a given testing interval, longer time lags necessitate more rounds of testing and result in longer durations of flare-ups (Figure 3A and

A

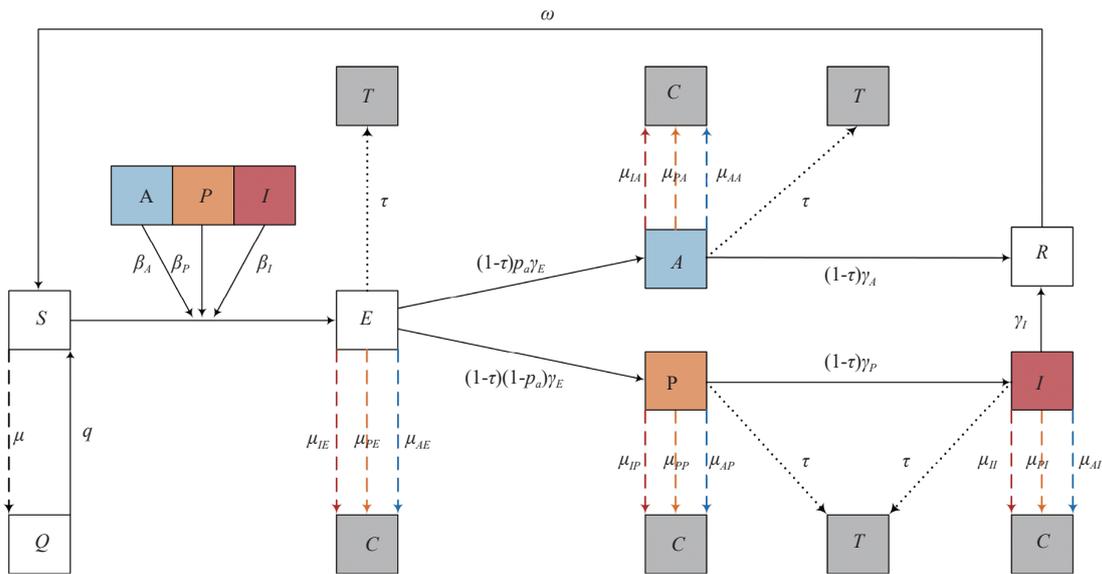
COVID-19 outbreak control



COVID-19 surveillance



B



$\mu_{ij}$ : removing rate for a contact traced through compartment  $i$  is in compartment  $j$  at the time of tracing,  $i \in \{A, P, I\}, j \in \{E, A, P, I\}$ .  
 $\mu$ : removing rate for a traced contact is in compartment  $S$  at the time of tracing.

FIGURE 1. Simplified illustration of models. (A) A schematic of coronavirus disease 2019 (COVID-19) outbreak control and surveillance in China. (B) The transmission-dynamic model.

Note: In the transmission dynamic model, the following compartments are considered: susceptible ( $S$ ), exposed ( $E$ ), infectious pre-symptomatic ( $P$ ), infectious asymptomatic ( $A$ ), infectious symptomatic ( $I$ ), and recovered ( $R$ ). Compartments for infections identified through population-level testing ( $T$ ) or contact tracing ( $C$ ) as well as healthy individuals in quarantine ( $Q$ ) are also included. The infections in  $T$  and  $C$  are isolated.  $\beta_A, \beta_P, \beta_I$  are the transmission rates for infectious asymptomatic, pre-symptomatic, and symptomatic cases, respectively.  $p_a$  is the proportion of asymptomatic cases.  $1/\gamma_E$  is the latent period.  $1/\gamma_P$  is the pre-symptomatic period for symptomatic cases.  $1/\gamma_A$  and  $1/\gamma_I$  are the time to recover for asymptomatic and symptomatic cases, respectively.  $\tau$  represents the rate of population-level PCR tests.  $1/q$  is the time for quarantine.  $\omega$  is the decay rate of antibodies for the individual in  $R$ . For further details, please refer to the methods section.

Abbreviation: PLT=population-level testing; CT=contact tracing.

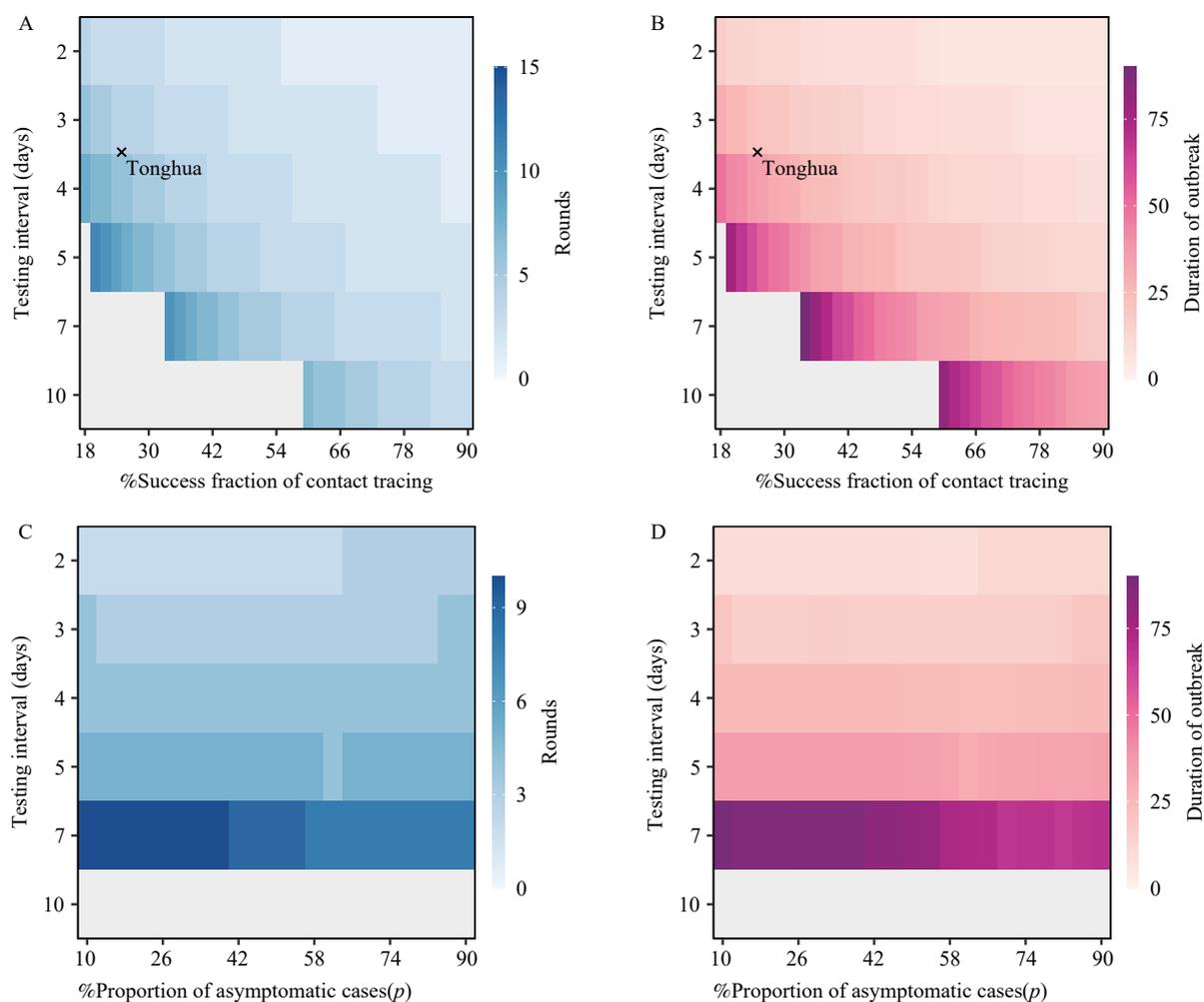


FIGURE 2. The effects of PLT strategies, CT, and asymptomatic proportion on mitigation of COVID-19 flare-up outbreaks in Tonghua City. (A) The number of rounds of tests required to contain transmission for the B.1.1 lineage. (B) The corresponding outbreak duration (days) in (A). (C) The effects of the proportion of asymptomatic cases on the required rounds of testing for the B.1.1 lineage. (D) The corresponding duration of flare-up outbreaks in (C).

Note: For panels A–D, the break interval is 2 days and the time lag is 3 days. For panel B, columns correspond to the success fraction of CT: the fraction of contacts that were successfully traced ( $\kappa$ ). Rows correspond to the number of days needed to complete a population-level round of testing. For panels C–D, the success fraction of contact tracing is set to 0.35. Grey areas represent parameter combinations by which outbreaks would not be controlled. This study defined that the outbreak was under control if the daily new infections were zero for 2 successive days. The number of rounds of PLT was calculated as the days to control the outbreak divided by the total days to complete 1 round of testing.

Abbreviation: COVID-19=coronavirus disease 2019; PLT=population-level testing; CT=contact tracing.

D). Similar issues are predicted for increasing testing intervals if the time lag is fixed (Figure 3).

### The Population-Level Testing and Contact Tracing Model with Age Structure

We then extended the previous model to include age structure using the Omicron infection data from Beijing.

Based on the estimated parameters for Beijing, the effects of different testing rates for different age groups

were explored (Figure 4). Overall, the simulation showed that the Omicron resurgence in a city could be controlled by multiple rounds of PLT. However, variations in the average number of tests needed per individual under different testing strategies were observed (Figure 4). The 21–60 age group has an important role in COVID-19 transmissions (Figure 4B). The modeling results demonstrate that, taking the average number of tests per individual as a benchmark, an appropriate frequency of tests for all age groups would be the best testing strategy to combat Omicron outbreaks.

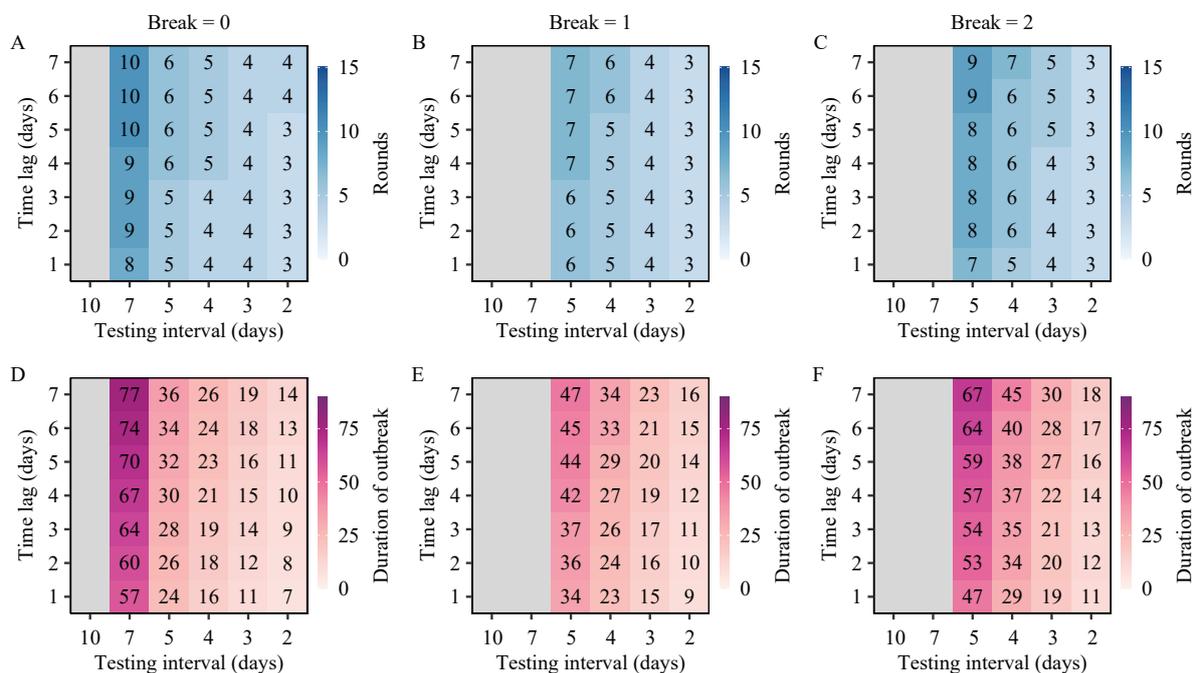


FIGURE 3. The effects of time lag, testing interval, and break interval on mitigation of B.1.1 outbreaks in Tonghua. (A) Rounds of PLT required to contain transmission at Break interval=0. (B) Rounds of PLT required to contain transmission at Break interval=1. (C) Rounds of PLT required to contain transmission at Break interval=2. (D) The corresponding outbreak duration (days) in (A). (E) The corresponding outbreak duration (days) in (B). (F) The corresponding outbreak duration (days) in (C).

Note: In panels A–C, the number in each cell represents the rounds of PLT needed to control the outbreak. The success fraction of CT is set to 26% for (A) to (F). Rows correspond to the time lag between the date of the first case identified and the date of launching PLT. Columns correspond to the population-level testing interval. The break interval represents the break time between 2 sequential PLTs. The total time to complete 1 round of testing is the testing interval plus the break interval. The grey represents the PLT strategy by which the outbreak would not be sustained.

Abbreviation: PLT=population-level testing; CT=contact tracing.

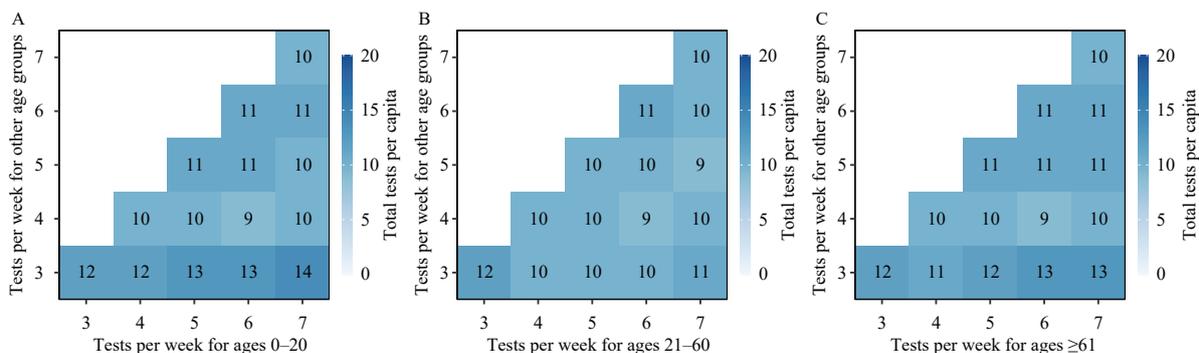


FIGURE 4. The needed total tests per capita to control Omicron wave under different testing strategies in Beijing Municipality. (A) The testing frequency for ages 0–20 years. (B) The testing frequency for ages 21–60 years. (C) The testing frequency for ages ≥61.

Note: The testing frequency for other age groups is smaller or equal to that for ages 0–20. Each cell represents the total tests per capita.

### The Probability of Detecting COVID-19 Cases for Surveillance Under Different Routine Testing Rates for Beijing

Routine testing is critical for COVID-19

surveillance. The routine testing rate and the targeted population for testing have impacts on the time it takes to detect a COVID-19 flare-up. We developed a model to estimate the cumulative distribution of time needed to detect COVID-19 cases since an undetected

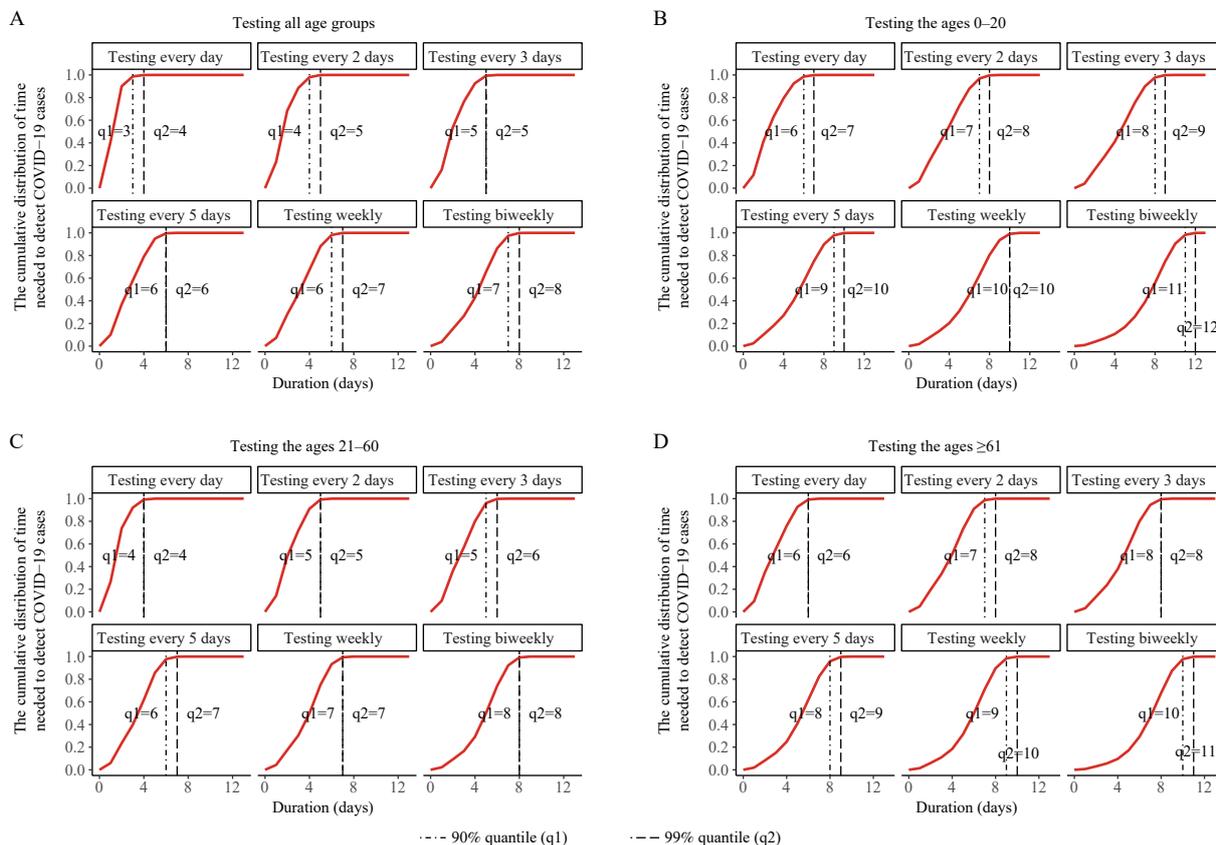


FIGURE 5. The cumulative distribution of time (days) needed to detect coronavirus disease 2019 (COVID-19) cases under different routine testing strategies in Beijing for surveillance. (A) Testing all age groups. (B) Testing the ages 0–20. (C) Testing the ages 21–60. (D) Testing those aged  $\geq 61$ .

Note: This study assumes that the first imported case was under exposed status and no reported SARS-CoV-2 infections occurred before the imported case. The duration is defined as the time interval from the date of importation of a COVID-19 case to the date of detecting at least 1 COVID-19 case by routine testing.

SARS-CoV-2 infection was imported. If routine testing is applied to all age groups, there is a 90% probability of detecting COVID-19 cases within 3 days using daily testing. Otherwise, it takes 7 days to detect COVID-19 cases with a 90% probability level under biweekly testing schemas (Figure 5A). If routine testing is applied to the 0–20 and  $\geq 61$  age groups, respectively, a longer delay to detect COVID-19 cases is observed (Figure 5B and 5D). However, the cumulative distribution of time needed to identify COVID-19 cases when testing the 21–60 age group is similar to that when testing all age groups (Figure 5C). This indicates that routine testing applied to the 21–60 age group for COVID-19 surveillance can achieve similar performance to that applied to all populations.

## DISCUSSION

Using the data from SARS-CoV-2 flare-up outbreaks, this study evaluated the effects of PLT and

CT on curbing COVID-19 resurgences. It showed that different combinations of PLT and CT lead to dramatically different scenarios of control. Considering the cost of PLT, there is both an economic and public health benefit to launching testing as early as possible and shortening testing intervals.

Testing capacity may be a challenge for some cities. However, a certain level of testing should be guaranteed given the dramatically enlarged capacity of polymerase chain reaction (PCR) tests and the new rapid testing methods made available since the start of the pandemic (13). Although the sensitivity of rapid antigen testing is lower than PCR tests, recent studies found that test sensitivity is secondary to frequency and turnaround time for COVID-19 screening (14). In regions with constrained resources, optimal pooled testing strategies may be employed (15).

The Omicron variant poses a great challenge for COVID-19 control. This study found that Omicron outbreaks could be controlled using multiple rounds of

PLT. If new variants with higher transmissibility emerge, more rounds of PLT and a higher success fraction of CT would be needed to contain the outbreak. The PLT strategy can also be optimized among different age groups if the average tests per individual is used as a benchmark. Considering the importation of COVID-19 cases, this study also evaluated time needed to detect COVID-19 cases under different routine testing rates and different targeted testing populations for surveillance. The results indicate that testing the 21–60 age group for COVID-19 can achieve similar performance to that of testing all populations.

There are a few limitations to this study. In this model, no spatial heterogeneity is assumed. In reality, populations living in residential areas with high-rise housing would be priority tested multiple times. Further, individuals in high-risk regions would be tested first. Therefore, the necessary rounds of PLT would be smaller than the prediction in this research. In addition, the model with the constancy of population size is formulated because no travel in and out is assumed. This assumption thus implies that there are no imported COVID-19 cases after the outbreak was detected. In reality, Omicron outbreak is harder to control compared to the B.1.1 variant. However, the results for these two variants based on our analysis cannot be compared directly because the parameters for them are quite different. Next, due to the insufficient surveillance of SARS-CoV-2, the reproduction number, proportion, and clinical and immunological profiles of asymptomatic infections are still not clear (16–17). More studies about the immunity profile (induced by primary infection and vaccines) and the protection of different vaccines against different lineages are needed to calibrate such calculations in the future. Finally, the age distribution of imported COVID-19 cases is assumed based on the age structure of Beijing. However, the deviation of imported COVID-19 cases from this assumption would have influences on the effectiveness of routine testing — especially for the different target testing populations.

In summary, our modeling analysis provides insights to local governments on what is necessary to control future COVID-19 resurgences in regards to population-level testing and contact tracing. Further investigation is required to understand whether the outcomes of frequent population-level testing can be replicated outside the context of China.

**Conflicts of interest:** No conflicts of interest.

**Acknowledgements:** We would like to acknowledge the CDC staff and local health workers in China who helped us collect data and who will continue to work to contain COVID-19 in China.

**Funding:** Provided by the Beijing Science and Technology Planning Project (Z221100007922019, Z201100005420010); Scientific and Technological Innovation 2030 — Major Project of New Generation Artificial Intelligence (2021ZD0111201); National Natural Science Foundation of China (82073616, 82204160); National Key Research and Development Program of China (2022YFC2303803, 2021YFC0863400); Beijing Advanced Innovation Program for Land Surface Science; Research on Key Technologies of Plague Prevention and Control in Inner Mongolia Autonomous Region (2021ZD0006); Fundamental Research Funds for the Central Universities.

doi: 10.46234/ccdcw2023.016

# Corresponding author: Huaiyu Tian, tianhuaiyu@gmail.com.

<sup>1</sup> State Key Laboratory of Remote Sensing Science, Center for Global Change and Public Health, College of Global Change and Earth System Science, Beijing Normal University, Beijing, China; <sup>2</sup> School of National Safety and Emergency Management, Beijing Normal University, Beijing, China; <sup>3</sup> Beijing Center for Disease Prevention and Control, Beijing, China.

✉ Joint first authors.

Submitted: November 05, 2022; Accepted: December 07, 2022

## REFERENCES

1. Fisman DN, Tuite AR. Asymptomatic infection is the pandemic's dark matter. *Proc Natl Acad Sci USA* 2021;118(38):e2114054118. <http://dx.doi.org/10.1073/pnas.2114054118>.
2. Moghadas SM, Fitzpatrick MC, Sah P, Pandey A, Shoukat A, Singer BH, et al. The implications of silent transmission for the control of COVID-19 outbreaks. *Proc Natl Acad Sci USA* 2020;117(30):17513 – 5. <http://dx.doi.org/10.1073/pnas.2008373117>.
3. Viana R, Moyo S, Amoako DG, Tegally H, Scheepers C, Althaus CL, et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* 2022;603(7902):679 – 86. <http://dx.doi.org/10.1038/s41586-022-04411-y>.
4. Mena GE, Martinez PP, Mahmud AS, Marquet PA, Buckee CO, Santillana M. Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile. *Science* 2021;372(6545):eabg5298. <http://dx.doi.org/10.1126/science.abg5298>.
5. Du ZW, Pandey A, Bai Y, Fitzpatrick MC, Chinazzi M, Piontti APY, et al. Comparative cost-effectiveness of SARS-CoV-2 testing strategies in the USA: a modelling study. *Lancet Public Health* 2021;6(3):e184 – 91. [http://dx.doi.org/10.1016/S2468-2667\(21\)00002-5](http://dx.doi.org/10.1016/S2468-2667(21)00002-5).
6. 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](http://dx.doi.org/10.1016/S1473-3099(20)30457-6).
7. Hellewell J, Russell TW, The SAFER Investigators and Field Study Team, The Crick COVID-19 Consortium, CMMID COVID-19 Working Group, Beale R, et al. Estimating the effectiveness of routine asymptomatic PCR testing at different frequencies for the detection of

- SARS-CoV-2 infections. *BMC Med* 2021;19(1):106. <http://dx.doi.org/10.1186/s12916-021-01982-x>.
8. Aleta A, Martín-Corral D, Piontti APY, Ajelli M, Litvinova M, Chinazzi M, et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat Hum Behav* 2020;4(9):964 – 71. <http://dx.doi.org/10.1038/s41562-020-0931-9>.
  9. Wells CR, Townsend JP, Pandey A, Moghadas SM, Krieger G, Singer B, et al. Optimal COVID-19 quarantine and testing strategies. *Nat Commun* 2021;12(1):356. <http://dx.doi.org/10.1038/s41467-020-20742-8>.
  10. Quilty BJ, Clifford S, Hellewell J, Russell TW, Kucharski AJ, Flasche S, et al. Quarantine and testing strategies in contact tracing for SARS-CoV-2: a modelling study. *Lancet Public Health* 2021;6(3):e175 – 83. [http://dx.doi.org/10.1016/S2468-2667\(20\)30308-X](http://dx.doi.org/10.1016/S2468-2667(20)30308-X).
  11. Elliott J, Whitaker M, Bodinier B, Riley S, Ward H, Cooke G, et al. Predictive symptoms for COVID-19 in the community: REACT-1 study of over 1 million people. *PLoS Med*. 2021 Sep 28;18(9):e1003777. <http://dx.doi.org/10.1371/journal.pmed.1003777>.
  12. Pavelka M, Van-Zandvoort K, Abbott S, Sherratt K, Majdan M, CMMID COVID-19 Working Group, et al. The impact of population-wide rapid antigen testing on SARS-CoV-2 prevalence in Slovakia. *Science* 2021;372(6542):635 – 41. <http://dx.doi.org/10.1126/science.abf9648>.
  13. Wells CR, Pandey A, Moghadas SM, Singer BH, Krieger G, Heron RJL, et al. Comparative analyses of eighteen rapid antigen tests and RT-PCR for COVID-19 quarantine and surveillance-based isolation. *Commun Med* 2022;2(1):84. <http://dx.doi.org/10.1038/s43856-022-00147-y>.
  14. Larremore DB, Wilder B, Lester E, Shehata S, Burke JM, Hay JA, et al. Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening. *Sci Adv* 2021;7(1):eabd5393. <http://dx.doi.org/10.1126/sciadv.abd5393>.
  15. Cleary B, Hay JA, Blumenstiel B, Harden M, Cipicchio M, Bezney J, et al. Using viral load and epidemic dynamics to optimize pooled testing in resource-constrained settings. *Sci Transl Med* 2021;13(589):eabf1568. <http://dx.doi.org/10.1126/scitranslmed.abf1568>.
  16. Long QX, Tang XJ, Shi QL, Li Q, Deng HJ, Yuan J, et al. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat Med* 2020;26(8):1200 – 4. <http://dx.doi.org/10.1038/s41591-020-0965-6>.
  17. Subramanian R, He QX, Pascual M. Quantifying asymptomatic infection and transmission of COVID-19 in New York City using observed cases, serology, and testing capacity. *Proc Natl Acad Sci USA* 2021;118(9):e2019716118. <http://dx.doi.org/10.1073/pnas.2019716118>.

## SUPPLEMENTARY MATERIAL

This section will introduce the models applied to the coronavirus disease 2019 (COVID-19) data in Tonghua City and Beijing Municipality, China.

### Population-Level Testing and Contact Tracing Model Without Age-Structure

**Basic notations and set-up:** We developed a transmission-dynamic model that incorporated the asymptomatic and symptomatic cases. Specifically, we considered susceptible ( $S$ ), exposed ( $E$ ), pre-symptomatic ( $P$ ), infectious asymptomatic ( $A$ ), infectious symptomatic ( $I$ ), and recovered ( $R$ ) individuals (see the model illustration in Figure 1). Control measures [population-level polymerase chain reaction (PCR) testing and contact tracing] were implemented, and the infected individuals were identified through either population-level testing (PLT) ( $T$ ) or contact tracing (CT) ( $C$ ). Therefore, they would be quarantined and exit the transmission chain. Healthy individuals may also be traced and quarantined ( $Q$ ). Although the contacts detected by CT were large, they account for quite a small proportion of the overall population given the millions of people in Tonghua. The proportion of susceptible persons didn't change a lot. 0 daily infections for 2 successive days was used as an index of a controlled flare-up. In this study, 1 round of PLT is defined as everyone tested 1 time in a population. The rounds of PLT needed to control resurgences was quantified.

**Population-level testing and contact tracing in the model:** Let  $\tau$  (the proportion of population tested per day) represent the testing rate for a city. To simplify the model, the spatial heterogeneity of testing was not included. Further, the tracing and quarantining of secondary contacts were not modeled for simplicity. Due to the detailed epidemiological investigations available to learn from, onwards infections could be identified relatively effectively from CT and PLT even in individuals without overt symptoms. Once an infection was identified, the contacts would be in different compartments at the time of tracing. Therefore, traced individuals are removed through different compartments (see the model illustration in Figure 1). To model the contact tracing in a detailed way, the contact tracing rate, the contact tracing precision (i.e. the proportion of traced contacts who were infected), and the probability that a contact traced through an infection from compartment  $i$  had progressed to compartment  $j$  at the time of tracing was formulated.

Considering that the sensitivity of PCR testing depends on the disease's progress, the PCR test sensitivities for different compartments were included. Specifically,  $\pi_E, \pi_A, \pi_P, \pi_I$  represent the sensitivity of PCR tests for the individuals in compartments  $E, A, P$  and  $I$ , respectively. This study then modeled the CT like those in the study of Davin Lunz et al. ( $I$ ) with extension. The contact tracing rate  $\alpha_i$  for compartment  $i, i \in \{E, A, P, I\}$  is given by the testing rate  $\tau$ , the sensitivity of PCR test, the fraction of contacts that were successfully traced  $\kappa$ , the contact number per day  $M$ , and the pre-defined CT time window ( $L$  days), which is  $\alpha_i = \pi_i \tau \kappa L M$ .  $\kappa$  represents the strictness and capacity of CT in a city.  $L$  depends on the specific infectious disease. The contact tracing precision  $\theta_i$  for the primary cases from compartment  $i$  is defined as the proportion of traced contacts through compartment  $i$  that were infected. It is related to the average transmission rate and the proportion of susceptible persons in the population, which is  $\theta_i = S \hat{\beta}_i / (N \times M)$ .  $N$  is the population size. For compartments  $A$  and  $P$ ,  $\hat{\beta}_A = \beta_A$  and  $\hat{\beta}_P = \beta_P$ . For compartment  $I$ , the average transmission rate is  $\hat{\beta}_I = \frac{\beta_P(\gamma_P + \tau\pi_P)^{-1} + \beta_I(\gamma_I + \tau\pi_I)^{-1}}{(\gamma_P + \tau\pi_P)^{-1} + (\gamma_I + \tau\pi_I)^{-1}}$ .  $\beta_i$  is the transmission rate

for an individual in compartment  $i$ . Note that the individuals in  $E$  are not infectious and the contact precision for compartment  $E$  is 0. The contacts who were traced through infections in compartment  $i$  are removed at the rate of  $\alpha_i \theta_i i$  and removed from compartment  $j$  with the proportion of  $p_{ij}$ , where  $\sum_{j \in \{E, A, P, I\}} p_{ij} = 1$ . Note that  $p_{ij}$  depends on dynamic of COVID-19. Note that the contacts traced through the individuals in compartment  $i$  may be either not infected, or were infected by someone else rather than the identified cases. Therefore, they are removed at the rate  $\alpha_i(1 - \theta_i)i$ . The full set of equations representing the transmission is given by

$$S(t+1) = S(t) - \Lambda + \omega R(t) + qQ(t) - \mu \frac{S(t)}{N}$$

$$E(t+1) = E(t) + \Lambda - (\tau\pi_E + (1 - \tau\pi_E)\gamma_E)E(t) - \alpha_I \theta_I I(t) p_{IE} - \alpha_P \theta_P P(t) p_{PE} - \alpha_A \theta_A A(t) p_{AE} - \mu \frac{E(t)}{N}$$

$$\begin{aligned}
A(t+1) &= A(t) + p_a(1 - \tau\pi_E)\gamma_E E(t) - (\tau\pi_A + \gamma_A(1 - \tau\pi_A))A(t) - \alpha_I\theta_I I(t)p_{IA} - \alpha_P\theta_P P(t)p_{PA} - \alpha_A\theta_A A(t)p_{AA} - \mu\frac{A(t)}{N} \\
P(t+1) &= P(t) + (1 - p_a)(1 - \tau\pi_E)\gamma_E E(t) - (\tau\pi_P + (1 - \tau\pi_P)\gamma_P)P(t) - \alpha_I\theta_I I(t)p_{IP} - \alpha_P\theta_P P(t)p_{PP} - \alpha_A\theta_A A(t)p_{AP} - \mu\frac{P(t)}{N} \\
I(t+1) &= I(t) + \gamma_P(1 - \tau\pi_P)P(t) - (\tau\pi_I + (1 - \tau\pi_I)\gamma_I)I(t) - \alpha_I\theta_I I(t)p_{II} - \alpha_P\theta_P P(t)p_{PI} - \alpha_A\theta_A A(t)p_{AI} - \mu\frac{I(t)}{N} \\
R(t+1) &= R(t) + \gamma_A(1 - \tau\pi_A)A(t) + \gamma_I(1 - \tau\pi_I)I(t) - \omega R(t) \\
Q(t+1) &= Q(t) - qQ(t) + \mu\frac{S(t)}{N} \\
C(t+1) &= C(t) + \alpha_I\theta_I I(t) + \alpha_P\theta_P P(t) + \alpha_A\theta_A A(t) + \mu\frac{E(t) + A(t) + P(t) + I(t)}{N} \\
T(t+1) &= T(t) + \tau(\pi_E E(t) + \pi_A A(t) + \pi_P P(t) + \pi_I I(t))
\end{aligned}$$

where

$$\begin{aligned}
\mu &= \alpha_I(1 - \theta_I)I(t) + \alpha_P(1 - \theta_P)P(t) + \alpha_A(1 - \theta_A)A(t) + \alpha_E E(t) \\
\Lambda &= \frac{S(t)}{N} \{ \beta_A A(t) + \beta_P P(t) + \beta_I I(t) \} \\
\sum_{j \in \{E, A, P, I\}} p_{ij} &= 1, i \in \{A, P, I\} \\
N &= S(t) + E(t) + A(t) + P(t) + I(t) + R(t) + Q(t) + C(t) + T(t)
\end{aligned}$$

In this study,  $p_a$  is the proportion of asymptomatic cases.  $1/\gamma_E$  is the latent period.  $1/\gamma_P$  is the pre-symptomatic period for symptomatic cases.  $1/\gamma_A$  and  $1/\gamma_I$  are the time to recover for asymptomatic and symptomatic cases, respectively.  $\beta_A$ ,  $\beta_P$ ,  $\beta_I$  are the transmission rates for asymptomatic, pre-symptomatic, and symptomatic cases, respectively. This model takes  $\beta_A$  for infectious asymptomatic individuals to be  $\lambda_1\beta_I$  and  $\beta_P$  for pre-symptomatic individuals to be  $\lambda_2\beta_I$  similar to the setting of (2). Due to other detailed epidemiological investigations, onwards infections could be identified relatively effectively from CT even in individuals without overt symptoms. In the model, traced individuals are tracked through different compartments. Similar to CT, the infections can also be identified through different compartments by population-level PCR tests, regardless of presence or absence of symptoms. For the individuals in  $R$ , they will enter the state of  $S$  due to the decay of antibodies at a rate of  $\omega$ . It is set to be 0 unless otherwise stated. The quantities  $\alpha_i$ ,  $\theta_i$  depends on the disease dynamic,  $\beta_I$  and  $\kappa$ . For  $p_{ij}$ , it also depends on the disease dynamic and the CT delay ( $t_0$ ) and can be derived from the model. Please refer to the following sections for more details about  $p_{ij}$ . According to the next generation matrix,  $R_0 = p_a\frac{\beta_A}{\gamma_A} + (1 - p_a)\frac{\beta_P}{\gamma_P} + (1 - p_a)\frac{\beta_I}{\gamma_I}$ .  $\beta_n$  is used to represent the reduced percentage of transmission rates due to other NPIs (for example, wearing face masks and following social distancing guidelines). Therefore, the actual transmission rate for symptomatic cases would be  $\beta_I(1 - \beta_n)$ . The unknown parameters for this model are the reduced percentage of transmission rate ( $\beta_n$ ), the fraction of contacts that were successfully traced ( $\kappa$ ), and the initial values for  $A$ ,  $P$ ,  $I$ , and  $E$ .

**The model fitting:** This study has two sets of observations: the daily new infections identified from CT and the daily new infections identified from PLT. It models the number of daily new infections identified by CT and the number of new infections identified by PLT as a random variable following Poisson distribution with expectation  $\lambda_t^C$  and  $\lambda_t^T$ , respectively. Specifically,

$$\begin{aligned}
\lambda_t^C &= \alpha_I\theta_I I(t) + \alpha_P\theta_P P(t) + \alpha_A\theta_A A(t) + \mu\frac{E(t) + A(t) + P(t) + I(t)}{N} \\
\lambda_t^T &= \tau \{ \pi_E E(t) + \pi_A A(t) + \pi_P P(t) + \pi_I I(t) \}
\end{aligned}$$

$\alpha_I\theta_I I(t) + \alpha_P\theta_P P(t) + \alpha_A\theta_A A(t)$  represents the infected contacts traced through infections in compartments  $A$ ,  $P$  and  $I$ .  $\mu\frac{E(t) + A(t) + P(t) + I(t)}{N}$  represents the traced contacts who were infected by someone else rather than the identified cases. Therefore,  $\lambda_t^C$  is the mean of daily infections identified by CT.  $\tau$  is the proportion of population tested per day and  $\lambda_t^T$  is the mean of daily infections identified by PLT. This study fitted the model to 2 sets of observations with a 3-day rolling mean. Model fitting was performed using the Metropolis–Hastings Markov chain Monte Carlo (MCMC) algorithm with the MATLAB (version R2020a) toolbox DRAM (Delayed Rejection Adaptive Metropolis). 100,000 iterations were set for burn-in. After that, another 100,000 iterations were performed.

**The calculation of  $p_{ij}^g$ :** This section calculates the probability that a contact who was traced through an infection in compartment  $i$  has progressed to compartment  $j$  at the time of tracing,  $i \in \{A, P, I\}$ ,  $j \in \{E, A, P, I\}$ . The calculation process is similar to that in (I), but expanded to  $A, P$  and  $I$  compartments from different age groups. Note that the age group  $g$  should be omitted in the model without age structure. The full descriptions with age structure are given here. For readability, the details are described here. The transition probability is introduced

$$P_{B, S_2|A, S_1} = \mathbb{P}(\text{individual in B at } t = S_2 \mid \text{individual in A at } t = S_1),$$

by assuming the individual progresses along a continuous-time Markov chain following the disease's progress. With time-homogeneity,  $P_{B, S_2|A, S_1} = P_{B, S_2-S_1|A, 0} =: P_{B|A}(S_2 - S_1)$ . Defining the time  $t=0$  as the time of obtaining the positive PCR tests report for the tested case of age group  $g$  in compartment  $i \in \{A, P, I\}$ , this study calculates the probability ( $p_{ij}^g$ ) that a contact traced through this case is in compartment  $j \in \{E, A, P, I\}$  at  $t = t_0 \geq 0$ .  $t_0$  represents the contact tracing delay. It is set to be 0 unless otherwise stated. Let  $\mathbb{Q}$  be the probability density of an individual infecting a contact (given the individual tests positive at time  $t=0$ ).

$$\begin{aligned} v_{ij}^g &= \mathbb{P}(\text{a contact traced through } I \text{ in } j \text{ at } t = 0) \\ &= \int_0^{tL} P_{j, t_0|E, -s} \mathbb{Q}(\text{infecting the contact at } t = -s) ds \\ &\propto \int_0^{tL} P_{j|E}(t_0 + s) \sum_{K \in \{P, I\}} \frac{K^g(t = -s)}{I^g(t = -s) + I^g(t = -s)} P_{K, -s|I, 0} ds \\ &= \int_0^{tL} P_{j|E}(t_0 + s) \sum_{K \in \{P, I\}} \frac{K^g(t = -s)}{I^g(t = -s) + I^g(t = -s)} P_{I, 0|K, -s} \frac{\mathbb{P}(\text{individual in } K^g \text{ at } t = -s)}{\mathbb{P}(\text{individual in } I^g \text{ at } t = 0)} ds \\ &\approx \int_0^{tL} P_{j|E}(t_0 + s) \sum_{K \in \{P, I\}} \frac{K^g(t = -s)}{I^g(t = -s) + I^g(t = -s)} P_{I|K}(s) \frac{K^g(t = -s)}{I^g(t = 0)} ds \\ &\approx \sum_{s=1}^{tL} P_{j|E}(t_0 + s) \sum_{K \in \{P, I\}} \frac{K^g(t = -s)}{I^g(t = -s) + I^g(t = -s)} P_{I|K}(s) \frac{K^g(t = -s)}{I^g(t = 0)} \end{aligned}$$

Similarly, there is

$$\begin{aligned} v_{Pj}^g &= \mathbb{P}(\text{a contact traced through } P \text{ in } j \text{ at } t = 0) \\ &= \int_0^{tL} P_{j, t_0|E, -s} \mathbb{Q}(\text{infecting the contact at } t = -s) ds \\ &\propto \int_0^{tL} P_{j|E}(t_0 + s) P_{P, -s|P, 0} ds \\ &= \int_0^{tL} P_{j|E}(t_0 + s) P_{I|P}(s) \frac{\mathbb{P}(\text{individual in } I^g \text{ at } t = -s)}{\mathbb{P}(\text{individual in } I^g \text{ at } t = 0)} ds \\ &\approx \sum_{s=1}^{tL} P_{j|E}(t_0 + s) P_{I|P}(s) \frac{I^g(t = -s)}{I^g(t = 0)} \end{aligned}$$

and

$$\begin{aligned} v_{Aj}^g &= \mathbb{P}(\text{a contact traced through } A \text{ in } j \text{ at } t = 0) \\ &= \int_0^{tL} P_{j, t_0|E, -s} \mathbb{Q}(\text{infecting the contact at } t = -s) ds \\ &\propto \int_0^{tL} P_{j|E}(t_0 + s) P_{A, -s|A, 0} ds \\ &= \int_0^{tL} P_{j|E}(t_0 + s) P_{A|A}(s) \frac{\mathbb{P}(\text{individual in } A^g \text{ at } t = -s)}{\mathbb{P}(\text{individual in } A^g \text{ at } t = 0)} ds \\ &\approx \sum_{s=1}^{tL} P_{j|E}(t_0 + s) P_{A|A}(s) \frac{A^g(t = -s)}{A^g(t = 0)} \end{aligned}$$

Above equations are continue-time. To be compatible with the model,  $v_{ij}^g$  will be calculated in discrete time way with step 1. At last, by normalizing,  $p_{ij}^g = \frac{v_{ij}^g}{\sum_{j \in \{E,A,P,I\}} v_{ij}^g}$ ,  $i \in \{A, P, I\}$ .  $p_{ij}^g$  depends on the disease dynamic, testing rate ( $\tau$ ) and contact tracing delay ( $t_0$ ). Here we assume  $t_0 = 0$ .  $tL$  represents the entire period of COVID-19 and is set to 25 days.

**Transition probability  $P_{B|A}$ :** In this section, we will calculate the transition probability. The derivation is same with that in (I). For readability, we write down the details. The transition probability is defined as

$$P_{B|A}(s) = \mathbb{P}(\text{individual in } B \text{ at } t = s \mid \text{individual in } A \text{ at } t = 0).$$

For the simplification, the transition rate from compartment  $K$  by  $\gamma_K^*$  is defined as  $\gamma_K^* = \gamma_K + \tau$ . Note that  $\tau$  may be changing during the flare-up in reality. However, to facilitate the computation, we used the average of  $\tau$ . Starting and finishing in the same compartment is

$$P_{K|K}(s) = 1 - \mathbb{P}(\text{leave } K \text{ by } t = s) = 1 - \gamma_K^* \int_0^s e^{-\gamma_K^* r} dr = e^{-\gamma_K^* s}.$$

For the transition from  $J$  to  $K$ , the fraction of individuals who leave  $J$  that reach  $K$  as  $q_{J \rightarrow K}$ . These are given by

$$q_{E \rightarrow A} = \frac{p_A \gamma_E}{\gamma_E^*}, q_{E \rightarrow P} = \frac{(1 - p_A) \gamma_E}{\gamma_E^*}, q_{P \rightarrow I} = \frac{\gamma_P}{\gamma_P^*}, q_{I \rightarrow R} = \frac{\gamma_I}{\gamma_I^*}, q_{A \rightarrow R} = \frac{\gamma_A}{\gamma_A^*}$$

Each transition requires a new integration and a reduction by the fraction of arrivals. So, we have

$$\begin{aligned} P_{A|E}(s) &= q_{E \rightarrow A} \int_0^s \mathbb{Q}(\text{leave } E \text{ at } t = r) (1 - \mathbb{P}(\text{leave } A \text{ by } t = s \mid \text{enter } A \text{ at } t = r)) dr \\ &= q_{E \rightarrow A} \int_0^s \gamma_E^* e^{-\gamma_E^* r} e^{-\gamma_A^*(s-r)} dr = q_{E \rightarrow A} \gamma_E^* \frac{e^{-\gamma_E^* s} - e^{-\gamma_A^* s}}{\gamma_A^* - \gamma_E^*} \end{aligned}$$

Similarly,

$$\begin{aligned} P_{P|E}(s) &= q_{E \rightarrow P} \gamma_E^* \frac{e^{-\gamma_E^* s} - e^{-\gamma_P^* s}}{\gamma_P^* - \gamma_E^*} \\ P_{I|P}(s) &= q_{P \rightarrow I} \gamma_P^* \frac{e^{-\gamma_P^* s} - e^{-\gamma_I^* s}}{\gamma_I^* - \gamma_P^*} \\ P_{I|E}(s) &= q_{E \rightarrow P} q_{P \rightarrow I} \frac{\gamma_E^* \gamma_P^*}{\gamma_P^* - \gamma_I^*} \left( \frac{e^{-\gamma_E^* s} - e^{-\gamma_P^* s}}{\gamma_E^* - \gamma_P^*} - \frac{e^{-\gamma_E^* s} - e^{-\gamma_I^* s}}{\gamma_E^* - \gamma_I^*} \right) \end{aligned}$$

### Population-Level Testing and Contact Tracing Model with Age-Structure

In this section, we extended above model and introduced the age-stratified population-level testing and contract tracing model. Specifically, we considered susceptible ( $S_b$ ), exposed ( $E_b$ ), pre-symptomatic ( $P_b$ ), infectious asymptomatic ( $A_b$ ), infectious symptomatic ( $I_b$ ), recovered ( $R_b$ ) individuals for age group  $b$ ,  $b=1, \dots, G$ . The infected individuals in age group  $b$  would be identified through population-level testing ( $T_b$ ) and contact tracing ( $C_b$ ). Healthy individuals may also be traced and quarantined ( $Q_b$ ). Next, we described how the contact tracing rate, the contact tracing precision and the probability that a contact traced through compartment  $i$  has progressed to compartment  $j$  at the time of tracing was formulated in details for age group  $b$ .

The contact tracing rate  $\alpha_b$  is given by the testing rate  $\tau$  (the proportion of population tested per day), the fraction of contacts that can be successfully traced  $\kappa$ , the contact number per day  $M_{bg}$  with other age group  $g$ , and the pre-defined contact tracing time window ( $L$  days), which is  $\alpha_b = \tau \kappa L \left( \sum_{g=1}^G M_{bg} i_g \right)$ .  $\kappa$  represents the strictness and capacity of contact tracing in a city. The contact tracing precision  $\theta_{bg}^i$  for the primary cases from compartment  $i$  in age group  $g$  contributing to age group  $b$  is defined as the proportion of traced contacts through compartment  $i$  of age group  $g$  were infected. It is related to the average transmission rate in age group  $g$  contributing to age group  $b$  and the proportion of susceptible in the population of age group  $b$ , which is  $\theta_{bg}^i = S_b \widehat{\beta}_{bg}^i / (N_b \times M_{bg})$ .  $N_b$  is the population size for age group  $b$ . For compartment  $A$  and  $P$ ,  $\widehat{\beta}_{bg}^A = \phi_b \beta_{A} M_{bg}$  and  $\widehat{\beta}_{bg}^P = \phi_b \beta_{P} M_{bg}$ . For compartment  $I$ ,

the average transmission rate is  $\widehat{\beta}_{bg}^I = \phi_b \frac{\beta_P(\gamma_P + \pi_P\tau)^{-1} + \beta_I(\gamma_I + \pi_I\tau)^{-1}}{(\gamma_P + \pi_P\tau)^{-1} + (\gamma_I + \pi_I\tau)^{-1}} M_{bg}$ .  $\phi_b$  is the susceptibility to infection for age group  $b$  and  $\beta_i$  is the probability of getting infected for each effective contact with individual in compartment  $i$ . Note that the individuals in  $E_b$  are not infectious and the contact precision for compartment  $E_b$  is 0. The contacts traced through compartment  $i$  are removed from age group  $b$  at the rate of  $\pi_i\tau\kappa L \left\{ \sum_{g=1}^G M_{bg} i_g \theta_{bg}^i \right\}$  and removed from compartment  $j$  of age group  $b$  with the proportion of  $p_{ij}^g$ , where  $\sum_{j \in \{E, A, P, I\}} p_{ij}^g = 1$ . Note that  $p_{ij}^g$  depends on dynamic of COVID-19 in age group  $g$ . The full set of equations representing the transmission is given by

$$\begin{aligned}
 S_b(t+1) &= S_b(t) - \Lambda_b + \omega R_b(t) + qQ_b(t) - \frac{\mu \cdot S_b S_b(t)}{N_b(t)} \\
 E_b(t+1) &= E_b(t) + \Lambda_b - (\tau\pi_E + (1 - \tau\pi_E)\gamma_E)E_b(t) - \frac{\mu \cdot S_b E_b(t)}{N_b(t)} \\
 &\quad - \tau\pi_I\kappa L \left\{ \sum_{g=1}^G M_{bg} I_g \theta_{bg}^I p_{IE}^g \right\} - \tau\pi_P\kappa L \left\{ \sum_{g=1}^G M_{bg} P_g \theta_{bg}^P p_{PE}^g \right\} - \tau\pi_A\kappa L \left\{ \sum_{g=1}^G M_{bg} A_g \theta_{bg}^A p_{AE}^g \right\} \\
 A_b(t+1) &= A_b(t) + p_a^b (1 - \tau\pi_E)\gamma_E E_b(t) - (\tau\pi_A + (1 - \tau\pi_A)\gamma_A)A_b(t) - \frac{\mu \cdot S_b A_b(t)}{N_b(t)} \\
 &\quad - \tau\pi_I\kappa L \left\{ \sum_{g=1}^G M_{bg} I_g \theta_{bg}^I p_{IA}^g \right\} - \tau\pi_P\kappa L \left\{ \sum_{g=1}^G M_{bg} P_g \theta_{bg}^P p_{PA}^g \right\} - \tau\pi_A\kappa L \left\{ \sum_{g=1}^G M_{bg} A_g \theta_{bg}^A p_{AA}^g \right\} \\
 P_b(t+1) &= P_b(t) + (1 - p_a^b)(1 - \tau\pi_E)\gamma_E E_b(t) - (\tau\pi_P + (1 - \tau\pi_P)\gamma_P)P_b(t) - \frac{\mu \cdot S_b P_b(t)}{N_b(t)} \\
 &\quad - \tau\pi_I\kappa L \left\{ \sum_{g=1}^G M_{bg} I_g \theta_{bg}^I p_{IP}^g \right\} - \tau\pi_P\kappa L \left\{ \sum_{g=1}^G M_{bg} P_g \theta_{bg}^P p_{PP}^g \right\} - \tau\pi_A\kappa L \left\{ \sum_{g=1}^G M_{bg} A_g \theta_{bg}^A p_{AP}^g \right\} \\
 I_b(t+1) &= I_b(t) + (1 - \tau\pi_P)\gamma_P P_b(t) - (\tau\pi_I + (1 - \tau\pi_I)\gamma_I)I_b(t) - \frac{\mu \cdot S_b I_b(t)}{N_b(t)} \\
 &\quad - \tau\pi_I\kappa L \left\{ \sum_{g=1}^G M_{bg} I_g \theta_{bg}^I p_{II}^g \right\} - \tau\pi_P\kappa L \left\{ \sum_{g=1}^G M_{bg} P_g \theta_{bg}^P p_{PI}^g \right\} - \tau\pi_A\kappa L \left\{ \sum_{g=1}^G M_{bg} A_g \theta_{bg}^A p_{AI}^g \right\} \\
 R_b(t+1) &= R_b(t) + (1 - \tau\pi_A)\gamma_A A_b(t) + (1 - \tau\pi_I)\gamma_I I_b(t) - \omega R_b(t) \\
 Q_b(t+1) &= Q_b(t) - qQ_b(t) + \mu \cdot S_b \\
 C_b(t+1) &= C_b(t) + \tau\pi_I\kappa L \left\{ \sum_{g=1}^G M_{bg} I_g \theta_{bg}^I \right\} + \tau\pi_P\kappa L \left\{ \sum_{g=1}^G M_{bg} P_g \theta_{bg}^P \right\} + \tau\pi_A\kappa L \left\{ \sum_{g=1}^G M_{bg} A_g \theta_{bg}^A \right\} \\
 T_b(t+1) &= T_b(t) + \tau(\pi_E E_b(t) + \pi_A A_b(t) + \pi_P P_b(t) + \pi_I I_b(t))
 \end{aligned}$$

where

$$\begin{aligned}
 \mu \cdot S_b &= \pi_I\tau\kappa L \left\{ \sum_{g=1}^G M_{bg} I_g (1 - \theta_{bg}^I) \right\} + \pi_P\tau\kappa L \left\{ \sum_{g=1}^G M_{bg} P_g (1 - \theta_{bg}^P) \right\} + \tau\pi_A\kappa L \left\{ \sum_{g=1}^G M_{bg} A_g (1 - \theta_{bg}^A) \right\} + \pi_E\tau\kappa L \left\{ \sum_{g=1}^G M_{bg} E_g \right\} \\
 \Lambda_b &= \phi_b \frac{S_b(t)}{N_b} \sum_{g=1}^G M_{bg} (\beta_A A_g(t) + \beta_P P_g(t) + \beta_I I_g(t)) \\
 \widehat{\beta}_{bg}^I &= \phi_b \frac{\beta_P(\gamma_P + \pi_P\tau)^{-1} + \beta_I(\gamma_I + \pi_I\tau)^{-1}}{(\gamma_P + \pi_P\tau)^{-1} + (\gamma_I + \pi_I\tau)^{-1}} M_{bg} \\
 \sum_{j \in \{E, A, P, I\}} p_{ij}^g &= 1, i \in \{A, P, I\} \\
 N_b &= S_b(t) + E_b(t) + A_b(t) + P_b(t) + I_b(t) + R_b(t) + Q_b(t) + C_b(t) + T_b(t)
 \end{aligned}$$

In our analysis,  $p_a^b$  is the proportion of asymptomatic cases for age group  $b$ . The quantities  $\alpha_b$ ,  $\theta_{bg}^i$  depend on the disease dynamic,  $\beta_I$  and  $\kappa$ . For  $p_{ij}^g$ , it also depends on the disease dynamic and the contact tracing delay ( $t_0$ ) and can be derived from the model.

Similar to previous model, we modeled the number of daily new infections as a random variable following Poisson distribution with expectation  $\lambda_i^C$  and  $\lambda_i^T$  for contact tracing and population-level testing, respectively. Specifically,

$$\lambda_t^C = \sum_{b=1}^G \tau \pi_I \kappa L \left\{ \sum_{g=1}^G M_{bg} I_g \theta_{bg}^I \right\} + \sum_{b=1}^G \tau \pi_P \kappa L \left\{ \sum_{g=1}^G M_{bg} P_g \theta_{bg}^P \right\} + \sum_{b=1}^G \tau \pi_A \kappa L \left\{ \sum_{g=1}^G M_{bg} A_g \theta_{bg}^A \right\} + \sum_{b=1}^G \frac{\mu_{sb} \{E_b(t) + A_b(t) + P(t) + I_b(t)\}}{N_b(t)}$$

$$\lambda_t^T = \sum_{b=1}^G \tau (\pi_E E_b(t) + \pi_A A_b(t) + \pi_P P_b(t) + \pi_I I_b(t)).$$

## Modeling the Probability of Detecting the First Case Under Routine Testing

We estimated the probability of detecting the first case  $Z$  for each day under routine testing since one SARS-CoV-2 infection was imported. Assuming the first detected case is found on the day  $t$ , it means that no infections have been detected in the past  $t-1$  days.  $Z = P_t \prod_{k=1}^{t-1} (1 - P_k)$ . Hence, we first formulated the probability for detecting at least one case  $P_t$  for day  $t$ .

Assuming the total number of cases tested for day  $t$  is  $B_t$ , We considered the number of cases tested positive as a binomial distribution with parameter  $B_t$  and  $p_t$ .  $p_t$  represents the probability of success for each trial. Therefore, the probability for detecting at least one case for day  $t$  is  $P_t = 1 - (1 - p_t)^{B_t}$ . The success probability  $p_t$  of having a positive PCR test for each tested case is a function of the testing sensitivity of PCR tests and the dynamics of outbreak. To simulate total number of cases tested  $B_t$ , we also developed a transmission-dynamic model with age structure, which incorporated susceptible ( $S$ ), exposed ( $E$ ), pre-symptomatic ( $P$ ), infectious asymptomatic ( $A$ ), infectious symptomatic ( $I$ ), recovered ( $R$ ) compartment. It is important to note that no control measures were implemented to cut the transmission chain because of no reported cases. The full set of equations representing the transmission is given by

$$\begin{aligned} S_b(t+1) &= S_b(t) - \Lambda_b + \omega R_b(t) \\ E_b(t+1) &= E_b(t) + \Lambda_b - \gamma_E E(t) \\ A_b(t+1) &= A_b(t) + p_a^b \gamma_E E_b(t) - \gamma_A A_b(t) \\ P_b(t+1) &= P_b(t) + (1 - p_a^b) \gamma_E E_b(t) - \gamma_P P_b(t) \\ I_b(t+1) &= I_b(t) + \gamma_P P_b(t) - \gamma_I I_b(t) \\ R_b(t+1) &= R_b(t) + \gamma_A A_b(t) + \gamma_I I_b(t) - \omega R_b(t) \end{aligned}$$

where

$$\begin{aligned} \Lambda_b &= \varphi_b \frac{S_b(t)}{N_b} \sum_{g=1}^G M_{bg} (\beta_A A_b(t) + \beta_P P_g(t) + \beta_I I_g(t)) \\ N_b &= S_b(t) + E_b(t) + A_b(t) + P_b(t) + I_b(t) + R_b(t) \end{aligned}$$

Considering that each tested case may be in any state of  $E$ ,  $A$ ,  $P$ ,  $I$ , and the sensitivity of PCR testing in each status is different, we estimated daily average positive probability  $p_t$  weighted by the proportion of population for each status for day  $t$ . Specifically, we have

$$\begin{aligned} B_t &= \sum_{b=1}^G \tau_b \{E_b(t) + A_b(t) + P_b(t) + I_b(t)\} \\ p_t &= \frac{1}{B_t} \sum_{b=1}^G \{\pi_E \tau_b E_b(t) + \pi_A \tau_b A_b(t) + \pi_P \tau_b P_b(t) + \pi_I \tau_b I_b(t)\} \\ P_t &= 1 - (1 - p_t)^{B_t} \end{aligned}$$

where  $\tau_b$  is the routine testing rate for age group  $b$ . We considered that the first imported infection is at the exposed ( $E$ ) status and distributed among the age groups according to the age proportion of Beijing.

## REFERENCES

- Lunz D, Batt G, Ruess J. To quarantine, or not to quarantine: A theoretical framework for disease control via contact tracing. *Epidemics* 2021;34:100428. <http://dx.doi.org/10.1016/j.epidem.2020.100428>.
- Aleta A, Martín-Corral D, Piontti APY, Ajelli M, Litvinova M, Chinazzi M, et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat Hum Behav* 2020;4(9):964 - 71. <http://dx.doi.org/10.1038/s41562-020-0931-9>.

## Methods and Applications

# Knowledge Graph: Applications in Tracing the Source of Large-Scale Outbreak — Beijing Municipality, China, 2020–2021

Ying Shen<sup>1</sup>; Yonghong Liu<sup>1</sup>; Xiaokang Jiao<sup>2</sup>; Yuxin Cai<sup>1</sup>; Xiang Xu<sup>1</sup>; Hui Yao<sup>1</sup>; Xiaoli Wang<sup>1,3,\*</sup>

## ABSTRACT

**Introduction:** Tracing transmission paths and identifying infection sources have been effective in curbing the spread of coronavirus disease 2019 (COVID-19). However, when facing a large-scale outbreak, this is extremely time-consuming and labor-intensive, and resources for infection source tracing become limited. In this study, we aimed to use knowledge graph (KG) technology to automatically infer transmission paths and infection sources.

**Methods:** We constructed a KG model to automatically extract epidemiological information and contact relationships from case reports. We then used an inference engine to identify transmission paths and infection sources. To test the model's performance, we used data from two COVID-19 outbreaks in Beijing.

**Results:** The KG model performed well for both outbreaks. In the first outbreak, 20 infection relationships were identified manually, while 42 relationships were determined using the KG model. In the second outbreak, 32 relationships were identified manually and 31 relationships were determined using the KG model. All discrepancies and omissions were reasonable.

**Discussion:** The KG model is a promising tool for predicting and controlling future COVID-19 epidemic waves and other infectious disease pandemics. By automatically inferring the source of infection, limited resources can be used efficiently to detect potential risks, allowing for rapid outbreak control.

Knowledge graphs (KGs) have been widely used in the construction of knowledge bases for search engines since their inception by Google. During the coronavirus disease 2019 (COVID-19) pandemic, KGs have played an important role in areas such as the construction of COVID-19-related knowledge bases (1–2), bibliometrics, drug information management,

drug repurposing (3–4), auxiliary diagnosis and treatment, and knowledge surveys. However, their application has been limited in exploring infection paths among cases (5–7) and identifying infection sources.

Tracing transmission paths can help to promptly identify the source of infection, detect high-risk areas that may otherwise be overlooked, and facilitate the identification of key populations, important sites with high infection risk, and possible superspreaders, thus allowing for timely actions to cut off the transmission chain and effectively contain the spread of an outbreak. However, in the face of the current COVID-19 pandemic and possible future pandemics with a huge number of infected cases, it is extremely time-consuming and labor-intensive to conduct epidemiologic investigation, identification, and management of close contacts, thus further limiting the resources allocated to tracing transmission paths and identifying infection sources. It is difficult to manually extract key information and trace infection paths among cases from the vast amount of unstructured textual data in case reports. Therefore, the use of information technology is important to quickly extract demographic and epidemiologic information, infer transmission paths and infection sources, identify key populations and key sites of high risk, and prevent further transmission at the community level.

To improve the effectiveness of epidemiological investigation and facilitate tracing of an infection source, we used natural language processing (NLP) and KG technologies to automatically extract structured data from case reports, determine the infection relationships among cases, trace the sources of infection, and construct a directed KG to identify infection sources using parameters including relationship intensity and transmission intensity.

## METHODS

### Study Design

Epidemiological data for COVID-19 cases in two

clusters involving severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) wild-type strain and Alpha variant were obtained from the Beijing Center for Disease Prevention and Control (Beijing CDC). Transmission chains were determined for each cluster by health professionals at the Beijing CDC, which served as the gold standard in the KG model.

### KG Model Framework

Using NLP technology, case reports were automatically structured and data were extracted, including sociodemographic characteristics, epidemiologic information, and case relationships. Epidemiologic information included the time of exposure, time of onset, time of first positive nucleic acid test, and time of diagnosis. Case relationships included both clear contacts between cases (e.g., living in the same household, dining together, contacts during medical visits, working or studying in the same room, and traveling in the same vehicle) and unclear contacts, such as being in the same location at the same time. Unclear contacts served as a supplement to clear contacts.

The possibility (intensity) of each edge in the network was calculated using the extracted

information. Edges with the highest intensity were preserved in the KG model, which was then presented with three components: 1) name of the infected individual (or individuals); 2) transmission paths (relationships among cases); and 3) intensity of transmission. A preliminary directed KG was constructed using the above data, and the source of infection was identified through pruning and reconstruction of the directed KG (Figure 1).

### Preliminary Construction of Directed Knowledge Graph

The directed KG was constructed in five steps: 1) inferring the viral shedding period of infectors; 2) inferring the infection time of infectees; 3) calculating the transmissibility in each case; 4) tracing the transmission paths between cases; and 5) constructing the directed KG. The details of each step are outlined below.

The viral shedding period of infectors was inferred. We assumed that cases are most contagious two days prior to and five days after symptom onset (8). Therefore, given onset time  $t$ , the most highly contagious period would be  $[t-2, t+5]$ . The time of the last positive nucleic acid test was denoted as NAT\_end.

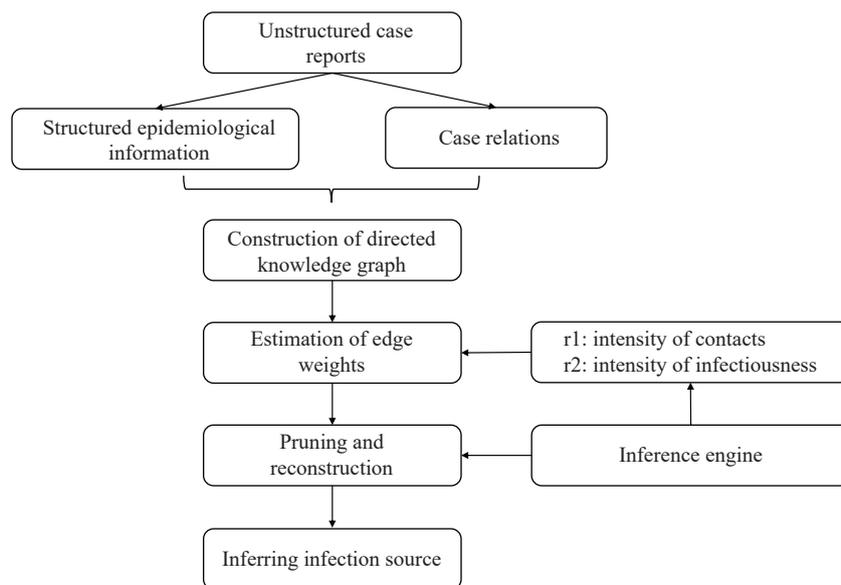


FIGURE 1. Flow chart of knowledge graph construction.

Note: This figure depicts the process of knowledge graph construction. Epidemiological information and case relationships were first retrieved from unstructured case reports. This information included cases' sociodemographic characteristics, time of exposure, time of onset, time of first positive nucleic acid test, time of diagnosis, and symptoms. Case relations included clear contacts such as sharing the same household, dining together, contacts during medical visits, working or studying in the same room, and traveling in the same vehicle, as well as unclear contacts such as appearing in the same location at the same time. Edge weights were then inferred based on the intensity of contacts and infectiousness. Finally, pruning was conducted according to the edge weights and inferred infection source.

If  $t+5 < \text{NAT\_end}$ , the case would still be infectious to a certain extent during the  $[t+5, \text{NAT\_end}]$  time interval.

The infection time of each confirmed case was inferred by tracing the transmission paths among cases. The infection time of each infectee was set as (date of onset – incubation period), which was approximately 3 to 14 days before symptom onset. For asymptomatic cases, the infection time was defined as 3 to 14 days before the first positive nucleic acid test result.

The transmissibility of each case was calculated, assuming that asymptomatic cases are less infectious than symptomatic ones.

Transmission paths among cases were traced and all possible relationships between a case and all other cases were calculated. As a result, multiple relationships may be found between two cases. Unclear contacts were assumed to have a lower probability of causing infection than clear contacts.

A preliminary directed KG was constructed. Based on the transmission paths discovered in the previous steps, a directed KG was constructed with cases as nodes and relationships as edges in the form of  $A \rightarrow B$ . The establishment of edges took time factors into full consideration, i.e., a case with an earlier onset was more likely to be a spreader and the infection time of infectees had to fall within the viral shedding period of the infectors.

## Pruning and Reconstruction of Directed Knowledge Graph

Given that there may be multiple relationships between two cases in the directed KG, pruning of edges based on their respective weight  $w$  was required for construction of the transmission paths:

$$w = r1 \times r2 \times rp$$

with  $r1$  representing the coefficient for viral shedding,  $r2$  the coefficient for the likelihood of case relationships,  $rp$  the individual characteristics, and  $w$  the weight of the edges. The edge with the highest  $w$  was taken as the most likely relationship between the two cases.

The value of  $r1$  was determined primarily by the time point of exposure to infectors, i.e., how infectious was the infector when the infectee was exposed? Because concrete time points of exposure and viral shedding were difficult to determine, the model classified case infectiousness using three scales where the infectiousness 5 days after symptom onset  $b >$  the

infectiousness 2 days prior to symptom onset  $a >$  other time window  $c$ . As a result,

$$R(t\_start, t\_end) = Rule\left(\sum_{t=t\_start}^{t=t\_end} \frac{sum(Patient(t))}{len(Infector(t))}, tol\right)$$

In which,  $patient(t)$  represents the whole set of infectees and  $infector(t)$  represents the whole set of infectors who have caused secondary cases in  $t$  days. Function  $Rule$  is the pre-defined infectiousness scale. Here, we defined  $b=1.3$  and  $a=2c$ . When the calculated relationship between  $a$ ,  $b$ , and  $c$  was larger than  $tol$  [for instance, if  $b > (1.3 + tol) \times a$  or  $b < (2 - tol) \times c$ ], the model would constrain the iteration of  $a$ ,  $b$ , and  $c$  until the constraint condition was met.

The value of  $r2$  represents the intensity of the relationships extracted from case reports. The model classifies the intensity of relationships into different categories based on the frequency and duration of contacts. For instance, the transmission likelihood could be assumed to be as follows: living together  $>$  working together  $>$  dining together  $>$  traveling in the same vehicle  $>$  living in the same community. The value of  $r2$  was further standardized between 0 and 1.

Compared with symptomatic cases, asymptomatic cases may have a longer incubation period but lower infectiousness. On the basis of the literature, we assumed that the infectiousness of asymptomatic cases was 30% ( $rp$ ) of the infectiousness of symptomatic cases (9).

By taking the aforementioned steps, preliminary pruning of the directed KG was performed. However, in practice, direct transmission and indirect transmission may coexist between cases. For instance, transmission paths of  $A \rightarrow B \rightarrow C$  and  $A \rightarrow C$  might both be reasonable. In such a situation, the model would consider both edge weights and case onset dates to determine whether an edge should be pruned or not. Specifically, edges with lower weights were pruned first; if two edges had the same weight, the model would further compare the time points at which the infectee was exposed to the infector. If the time points were the same, no edges would be pruned and all transmission paths would be retained.

The process described above can essentially guarantee that there is only one relationship (edge) between two cases (nodes). However, some nodes may be left on their own. In such cases, the model would select one edge with the highest  $w$  value from those pruned to reconstruct a relationship between cases, with the aim of linking as many cases as possible.

## Identifying the Source of Infection

Depth-first search (DFS) was conducted starting from a random node  $s$  and running through all other nodes, generating a list  $L$  of all traversed nodes. With DFS performed on all nodes, the starting node  $S$  with the longest  $L$  was identified as the source of infection. That is, the case with the most comprehensive transmission paths was selected as the infection source.

## RESULTS

In the knowledge graph generated by the KG model, COVID-19 cases were represented by entities, the transmission paths (relationships) between cases were

represented by edges, and infectors' onset time and viral shedding duration comprised entity attributes. With this structure, a complete knowledge graph with transmission information was generated automatically. To compare the transmission paths between the KG model and professional determination more directly, illustrations were drawn manually in the same style (Figures 2 and 3).

A cluster of COVID-19 cases caused by the wild-type strain of SARS-CoV-2 occurred in Shunyi District, Beijing in 2020, involving a total of 42 cases. Through epidemiological investigation and professional judgment, 20 infection relationships were identified manually. The KG model identified 42

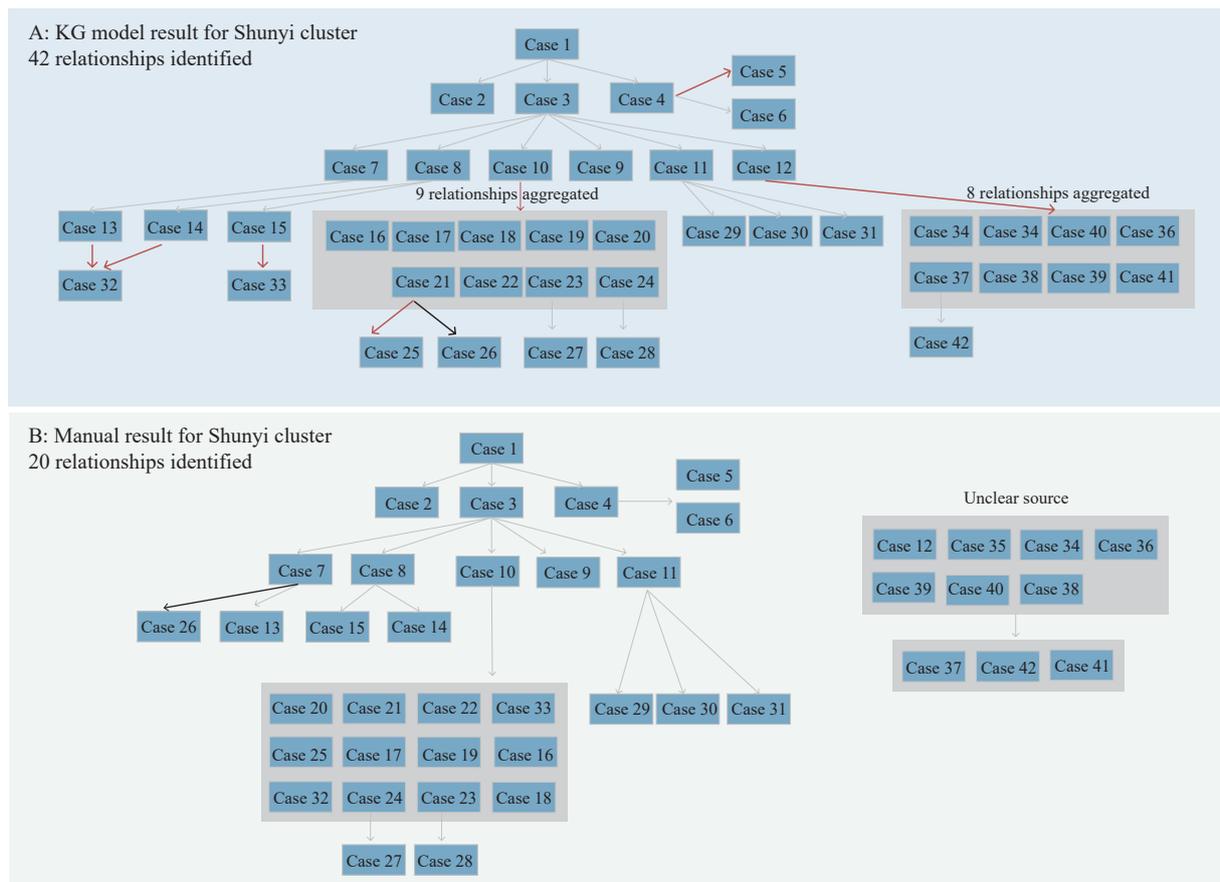


FIGURE 2. Transmission paths for Shunyi cluster. (A) A total of 42 relationships were identified in the knowledge graph (KG) model for the Shunyi cluster. (B) A total of 20 relationships were identified by public health professionals for the Shunyi cluster. Sources for Cases 12, 34–36, and 30–40 were unclear and were presented separately.

Note: Red arrows represented additional relationships identified by the KG model; black arrows represented different relationships between the KG model and manual determination; gray arrows represented the same relationships. For illustration purposes, 9 relationships from Case 10 to Case 16–24 were aggregated and presented in a gray square; 8 relationships from Case 12 to Case 34–41 were aggregated and presented in a gray square. The edge weights for Case 13 → Case 32 and Case 14 → Case 32 were the same thus both were kept. Sources for Cases 12, 34–36, and 30–40 were unclear and were presented separately. For illustration purposes, 9 relationships from Case 10 to Case 16–24 were aggregated and presented in a gray square; 8 relationships from Case 12 to Case 34–41 were aggregated and presented in a gray square. The edge weights for Case 13 → Case 32 and Case 14 → Case 32 were the same thus both were kept.

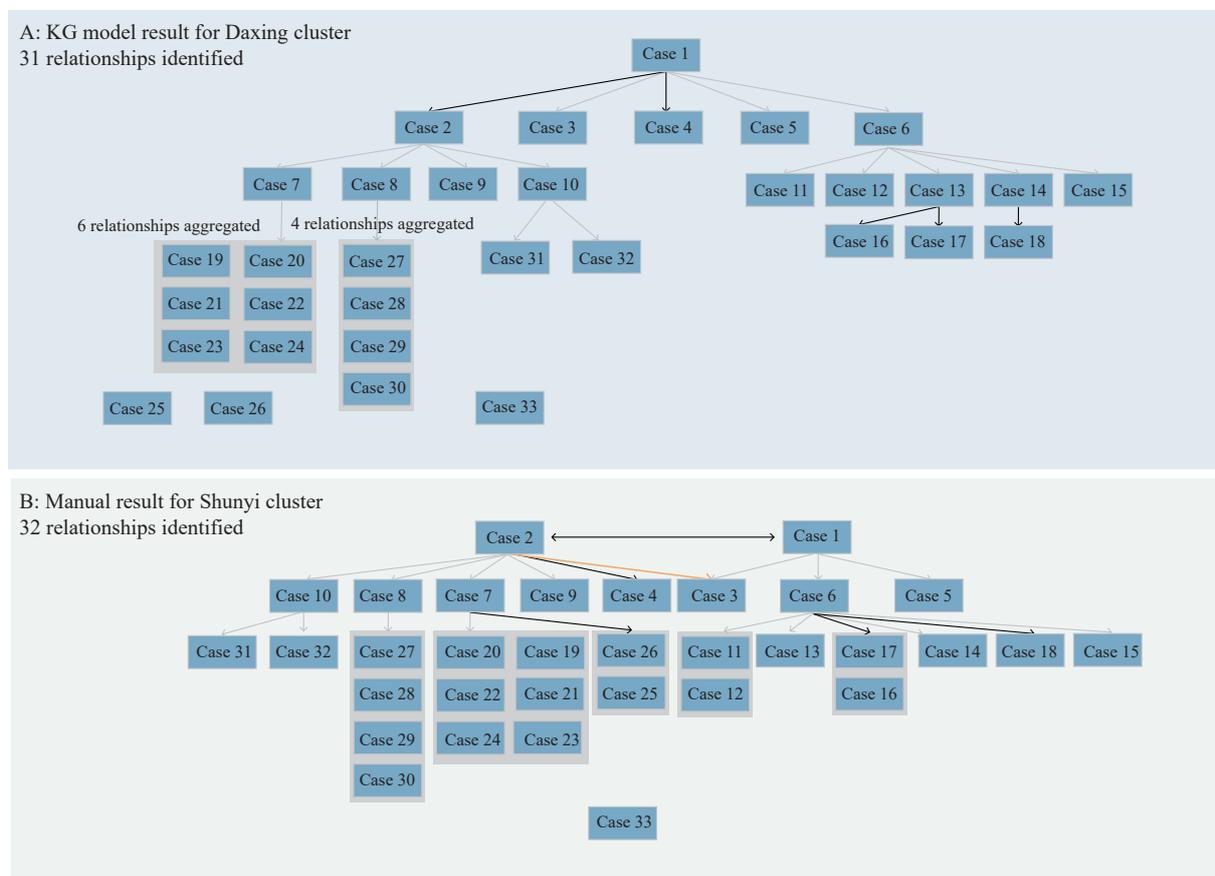


FIGURE 3. Transmission paths for Daxing cluster. (A) A total of 31 relationships were identified in the KG model for the Daxing cluster. (B) A total of 32 relationships were identified by public health professionals for the Daxing cluster.

Note: Red arrows indicated additional relationships identified by the KG model; black arrows indicated different relationships between the KG model and manual determination; gray arrows indicated the same relationships; and orange arrows indicated relationships omitted by the KG model. For illustration purposes, 9 relationships from Cases 7 to 21–24 were aggregated, and four relationships from Case 8 to Cases 27–30 were aggregated. The infection source for Case 33 was unclear. For illustration, 4 relationships from Cases 8 to 27–30 were aggregated; 8 relationships from Cases 7 to 19–23 and 25–26 were aggregated; and 4 relationships from Cases 6 to 11–12 and 16–17 were aggregated. The infection source for Case 33 was unclear, while Cases 1 and 2 were both possible sources.

Abbreviation: KG=knowledge graph.

relationships, and an additional 22 relationships were found to be possible after one-by-one deduction, which could serve as an important supplement in the judgment of the epidemic's development. One relationship was different between the KG model and manual determination, but both were found to be possible (Figure 2).

An outbreak of the SARS-CoV-2 Alpha variant occurred in Daxing District, Beijing in 2021, involving 33 cases in total. Manual tracing identified 32 relationships, while the KG model identified 31. Of these, 24 edges were the same as those traced manually, 7 were different, and 1 edge was omitted. In manual deduction, both infection paths were deemed possible; thus, both were preserved, while the KG model only kept the most likely infection path and pruned the

other. Among the seven edges, five were related to the identification of index cases in households. The model suggested that, compared to family member B, family member A had a greater likelihood of introducing the virus into the household. The other two edges were identified after a detailed analysis of case reports (Figure 3).

## DISCUSSION

The KG model described herein was able to automatically extract data from unstructured text in epidemiologic case reports and sort out complex infection relationships. A directed KG that depicted the identified case relationships and infection sources was successfully constructed through a detailed

pruning and reconstruction process. We tested the KG model using two actual COVID-19 outbreaks that occurred in Beijing, China, and the model was proven effective in targeting the infection source.

Using the KG model to deduce transmission pathways, “Case Zero” can be quickly identified, allowing the government to direct limited resources and determine the possible infection source (6). Furthermore, the KG model can be used to identify key transmission sites and key spreaders, which can then inform the detection of populations at higher risk, improve the efficiency of case screening, and help contain the spread of an outbreak in a timely manner. Additionally, a focused investigation could be organized for lonely nodes in the KG (i.e., cases whose transmission paths were not clear) to identify hidden infection sources in a timely fashion. This could help to quickly review the overall epidemic prevention and control direction and address potential issues rapidly, thus avoiding worsening of a current outbreak and preventing future outbreaks. Although the prevention and control strategy for COVID-19 has changed substantially, the KG technology presented in this paper could still enrich the current toolbox of public health countermeasures and offer insights for future epidemics caused by other emerging or existing infectious diseases.

This study has some limitations. First, the KG model is a tool for analyzing infection sources, and its performance is largely affected by the completeness of epidemiological case reports. To be used in future epidemics, essential information from case reports must be clarified in advance. Second, this model was tested in small outbreaks, with good performance; however, the model requires further validation in larger outbreaks.

**Conflicts of interest:** No conflicts of interest.

**Acknowledgments:** All health professionals fighting against COVID-19, especially our colleagues in Beijing Center for Disease Prevention and Control.

**Funding:** Supported by National Key Research and Development Program of China (2021ZD0114102),

Science Program of Beijing City (Z221100007922019), and Beijing Natural Science Foundation (7202073).

doi: 10.46234/ccdcw2023.017

# Corresponding author: Xiaoli Wang, wangxiaoli198215@163.com.

<sup>1</sup> Beijing Office of Global Health, Beijing Center for Disease Prevention and Control, Beijing, China; <sup>2</sup> Yidu Cloud Technology Co Ltd, Beijing, China; <sup>3</sup> School of Public Health, Capital Medical University, Beijing, China.

Submitted: December 13, 2022; Accepted: January 18, 2023

## REFERENCES

- Chen CM, Ross KE, Gavali S, Cowart JE, Wu CH. COVID-19 knowledge graph from semantic integration of biomedical literature and databases. *Bioinformatics* 2021;37(23):4597 – 8. <http://dx.doi.org/10.1093/bioinformatics/btab694>.
- Domingo-Fernández D, Baksi S, Schultz B, Gadiya Y, Karki R, Raschka T, et al. COVID-19 knowledge graph: a computable, multi-modal, cause-and-effect knowledge model of COVID-19 pathophysiology. *Bioinformatics* 2021;37(9):1332 – 4. <http://dx.doi.org/10.1093/bioinformatics/btaa834>.
- Hsieh K, Wang YY, Chen LY, Zhao ZM, Savitz S, Jiang XQ, et al. Drug repurposing for COVID-19 using graph neural network and harmonizing multiple evidence. *Sci Rep* 2021;11(1):23179. <http://dx.doi.org/10.1038/s41598-021-02353-5>.
- Al-Saleem J, Granet R, Ramakrishnan S, Ciancetta NA, Saveson C, Gessner C, et al. Knowledge graph-based approaches to drug repurposing for COVID-19. *J Chem Inf Model* 2021;61(8):4058 – 67. <http://dx.doi.org/10.1021/acs.jcim.1c00642>.
- Jiang BC, You X, Li K, Li TT, Zhou XJ, Tan LH. Interactive analysis of epidemic situations based on a spatiotemporal information knowledge graph of COVID-19. *IEEE Access* 2022;10:46782 – 95. <http://dx.doi.org/10.1109/ACCESS.2020.3033997>.
- Wang J, Wang K, Li J, Jiang J, Wang Y, Mei J, et al. Accelerating epidemiological investigation analysis by using NLP and knowledge reasoning: a case study on COVID-19. *AMIA Annu Symp Proc* 2020;2020:1258-67. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8075493/>.
- Chen LM, Liu D, Yang JK, Jiang MY, Liu SQ, Wang Y. Construction and application of COVID-19 infectors activity information knowledge graph. *Comput Biol Med* 2022;148:105908. <http://dx.doi.org/10.1016/j.combiomed.2022.105908>.
- Hakki S, Zhou J, Jonnerby J, Singanayagam A, Barnett JL, Madon KJ, et al. Onset and window of SARS-CoV-2 infectiousness and temporal correlation with symptom onset: a prospective, longitudinal, community cohort study. *Lancet Respir Med* 2022;10(11):1061 – 73. [http://dx.doi.org/10.1016/S2213-2600\(22\)00226-0](http://dx.doi.org/10.1016/S2213-2600(22)00226-0).
- Johansson MA, Quandelacy TM, Kada S, Prasad PV, Steele M, Brooks JT, et al. SARS-CoV-2 transmission from people without COVID-19 symptoms. *JAMA Netw Open* 2021;4(1):e2035057. <http://dx.doi.org/10.1001/jamanetworkopen.2020.35057>.

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 © 2023 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).



*Vol. 5 No. 4 Jan. 27, 2023*

---

**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](mailto:weekly@chinacdc.cn)

**CSSN**

ISSN 2096-7071

CN 10-1629/R1