Accepted for publication in *Nature Human Behavior* (January 27, 2021)

Integrated vaccination and physical distancing interventions to prevent future COVID-19 waves in Chinese cities

Bo Huang[11,2,3†*], Jionghua Wang[1†], Jixuan Cai[4†], Shiqi Yao[1], Paul Kay Sheung Chan[5,6*], Tony Hong-wing Tam[3], Ying-Yi Hong[7], Corrine W. Ruktanonchai[8,9], Alessandra Carioli[8], Jessica R. Floyd[8], Nick W. Ruktanonchai[8,9], Weizhong Yang[10], Zhongjie Li[11], Andrew J. Tatem[8*], Shengjie Lai[8,10,12†]

The COVID-19 pandemic has posed significant challenges to the formulation of preventive interventions, especially given that the effects of physical distancing and upcoming vaccines on reducing susceptible social contacts and eventually halting transmission are still unclear. Using anonymized mobile geolocation data in China, we devise a mobility-associated social contact index to quantify the impact of both physical distancing and vaccination measures in a unified way. On this basis, our epidemiological model reveals that vaccination combined with physical distancing can contain resurgences without relying on stay-at-home restrictions, whereas a gradual vaccination process alone cannot achieve this. Further, for cities with medium-population density, vaccination can shorten the duration of physical distancing by 36%-78%, whereas for cities with high-population density, infection numbers can well be controlled through moderate physical distancing. These findings improve our understanding of the joint effects of vaccination and physical distancing with respect to a city's population density and social contact patterns.

[1]Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong SAR. [2]Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Hong Kong SAR. [3]Department of Sociology and Center for Population Research, The Chinese University of Hong Kong, Hong Kong SAR. [4]Tencent Inc., Shenzhen, China. [5]Department of Microbiology, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR. [6]Stanley Ho Centre for Emerging Infectious Diseases, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR. [7]Department of Management, The Chinese University of Hong Kong, Hong Kong SAR. [8]WorldPop, School of Geography and Environmental Science, University of Southampton, England, UK. [9]Population Health Sciences, Virginia Tech, Blacksburg, VA 24061, USA. [10]School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China. [11]Division of Infectious Diseases, Chinese Center for Disease Control and Prevention, Beijing, China. [12]School of Public Health, Fudan University, Shanghai, China. [*]E-mail: bohuang@cuhk.edu.hk; paulkschan@cuhk.edu.hk; A.J.Tatem@soton.ac.uk.

Since the coronavirus disease (COVID-19) was first identified in December 2019 in Wuhan, China[1], the entire world has been adversely affected by the ensuing pandemic[2–4]. As infections decreased during the summer (2020) months, many countries relaxed their lockdown and physical distancing measures in the course of reopening their economies and societies. Due to the increasing mobility and social contact rates, accompanied with the large numbers of susceptible people in the population, countries worldwide have been experiencing COVID-19 resurgences[5,6].

In the absence of an effective vaccine, physical distancing interventions (e.g., closure of schools and workplaces) are critical to contain the resurgences, even though such interventions have caused significant disruptions to societies and economies[7,8]. As of January 2021, ten SARS-CoV-2 vaccines have received either full approval or limited approval for emergence use[9]. Policymakers are also looking forward to the arrival of other vaccines in the coming months so that physical distancing restrictions can be alleviated, as vaccination can reduce a portion of susceptible contacts that may result in transmission[10,11]. However, the extent of potential alleviation is still unclear, especially given that the supply of vaccines will likely not be sufficient to achieve herd immunity in the immediate future. Therefore, more comprehensive interventions including both physical distancing and vaccine implementation strategies should be developed to avoid COVID-19 resurgences. To this end, there is an urgent need to understand the interrelationships among mobility, social contacts, physical distancing, vaccination, and virus transmission for tailoring and adjusting preventive interventions.

Human movement and contact rates play fundamental roles in shaping the transmission patterns of infectious diseases[12,13]. Their impact on COVID-19 inter-city spread has been extensively investigated using anonymized mobile phone data[14–16], and previous studies have attempted to assess the effects of travel and physical distancing measures on the first wave of the COVID-19 pandemic[17–22]. In addition, mobility data have been recognized to effectively reflect the resumed human activities after lifting lockdown measures[23,24], and have also been used as a proxy for measuring the effectiveness of interventions, such as stay-at-home orders, to mitigate or contain the transmission of COVID-19[25,26]. However, most studies have primarily used existing publicly available datasets[27–31] to derive and provide coarse information on population mobility to measure changes in inter-city travel flow, check-in intensity, or trip length under interventions. More refined data with anonymized geolocation information have been under-utilized to directly inform social contact rates[32,33]. The use of only mobility data would also not be able to facilitate assessing the impact of vaccination measures, which requires deriving information on safe social contacts in modeling transmission dynamics.

Because mobility data cannot directly inform physical distancing and reductions in contact rates, studies examining the effect of lockdown policies on transmission have generally assumed that when people reduce their mobility, they proportionally reduce their contacts[23,24]. However, in reality, similar levels of social contact have been observed at both high and low levels of mobility in various studies[32,33], and using mobility or social contact data alone may not be sufficient for precisely measuring physical distancing. The paired relationship between reductions/restorations in mobility and social contacts is more suited to serve this purpose. Thus, it is crucial to understand the interaction between mobility and social contact rates over time under COVID-19 interventions.

Future studies on epidemic spread should be able to consider the strong interaction between physical distancing and mobility to reduce contact rates with or without vaccination in case of resurgences. Ideally,

investigation of the reduction in contact rates would involve creating a singular index that encompasses the combined impact of reduced mobility and physical distancing on contact rates so as to directly inform the estimates of contact levels among populations over time. This index should also be able to account for the impact of vaccination on reducing susceptible contact rates with minimal adaptations. In this paper, we propose a social contact index (SCI) to represent *the daily average potential social contact* (or spatiotemporal co-presence) per person, which was derived from an anonymized mobile geolocation dataset. The index associates a series of mobility levels (e.g., 10% to 100% with increments of 10%) with their corresponding social contact rates. Based on this association, an empirical relationship was established using allometric equations[34,35] to compute the contact rates under a given level of physical distancing and population density for future scenario simulations. As a proxy for daily social contact rates per person, this index was incorporated into a modified susceptible-exposed-infectious-recovered (SEIR) model to quantify the impact of physical distancing interventions in light of the COVID-19 outbreak across China. The risks of COVID-19 resurgence under a no-vaccination scenario were then assessed under varying mobility, physical distancing, and population density scenarios. Subsequently, the effect of administering vaccines to relax physical distancing interventions and reduce unsafe social contacts was assessed in preventing a resurgence of infections and curtailing the pandemic.

Results

Deriving a social contact metric in populations. The proposed SCI used to measure the level of contacts per person in a city was created using the *total number of potential contact events* (or the total social contacts index; TSCI), which was directly determined using a large, near-real-time anonymized mobile device positioning dataset from Tencent[30] that covers more than 70% of the population in mainland China[31]. In this dataset, the potential contact events are detected in the context of the spatiotemporal co-presence of people[36–38] or when their mobile devices request positioning services within a specified space-time bin (i.e., 250 meters and 10 minutes in our case) (Extended Data Fig. 1 A).

The SCI is an average per person TSCI. It is paired with mobility in a non-linear form (see Methods), but is also influenced by physical distancing (Extended Data Fig. 1 B) and population density. Here, *mobility* refers to the number of trips people make outside their homes. During the COVID-19 outbreak, people might have been ordered to, encouraged to, or volunteered to quarantine and work from home to reduce social contacts, thereby leading to a decrease in the SCI. However, after the lockdown measures were lifted, mobility was gradually restored. *Physical distancing* measures are applied to keep people who are outside their homes away from crowded places by closing parts of facilities or limiting the maximum number of people gathering in facilities through, for example, staggered entry and exit. Generally, high mobility leads to a high SCI, as does a high population density. However, strong physical distancing leads to a low SCI. Vaccination can protect people by greatly enhancing immunity and thus reduce the SCI of COVID-19 transmission between susceptible population and infectors. The population density of a city may change owing to its population inflow and outflow, which in turn affects the SCI and disease transmission (see Methods).

Taking Wuhan as an example, an overview of the overall social contact rates, as reflected by the TSCI of the entire city, is provided, together with the changes in the TSCI before, during, and after the lockdown

(January 23–April 7, 2020) due to the implementation of physical distancing measures of varying intensities (Figure 1 A–C). The TSCI was categorized into five types based on the areas of interest[30] where contact occurred, namely residential communities, workplaces, schools, shopping/recreation facilities, and other facilities. Before the lockdown (normal period) in December 2019, more than 80% of the contact occurred in the first four types of places. Figure 1 D shows the changes in the TSCI of Wuhan and those of four other major cities in China (Beijing, Shanghai, Guangzhou, and Shenzhen) against their normal patterns (the average levels in December 2019). After the implementation of nationwide interventions since January 23, 2020, two days before the start of Chinese New Year (CNY) (January 25, 2020), the TSCI of the five cities decreased drastically, and especially that of Wuhan dropped to only 2% of its normal level (i.e., 1.016) on January 27, i.e., the day 4 after imposing the lockdown measures. On April 8, the lockdown was lifted in Wuhan and the TSCI slowly recovered reaching 50% on May 31, while the TSCI values of other Chinese cities nearly returned to the respective normal levels (Figure 1 D).

As revealed by the variations in the SCI values shown in Figure 1, physical distancing restrictions of varying intensities were imposed over different periods in Wuhan, namely pre-lockdown (e.g., December 2019)[39,40], lockdown (from January 23 to April 7, 2020)[39,40], shortly after lockdown lifting (SALDL) (from April 8 to May 20, 2020), and longer after lockdown lifting (LALDL) (from May 20 to May 30, 2020, after nucleic acid testing of all citizens)[41]. Various interventions were implemented together, which made it difficult to quantify their effects on the contact rate. To solve this problem, we derived four explicit sets of physical distancing interventions using the mobility dataset, namely "no," "mild," "moderate," and "strong" (see Methods and Supplementary Table 1). Each set of these interventions at a certain intensity was found to cause similar impacts on the SCI values as those imposed together during the same period in Wuhan. This outcome allowed us to inform transmission dynamics under explicit sets of physical distancing measures.

The relationship between mobility and the SCI was modeled as an allometric growth curve[34,35], but it varied over the three levels of population density and the four intensities of physical distancing (or the four explicit sets of physical distancing measures) (Figure 2). The SCI values for fitting each curve under a series of randomly sampled mobility levels were determined using the mobility dataset during the normal period (i.e., December 2019 in our case) because there was no significant population migration or physical distancing intervention. The modeled SCI values were validated against the actual SCI values extracted from the original dataset under each mobility level (see Methods). Thus, a mobility-SCI coupled metric was formulated in the form of these equations to determine the SCI. This metric can also be adapted to evaluate the effect of vaccination on the SCI. As the people with immunity to SARS-CoV-2 cannot infect or be infected by others, their contacts would not impact the transmission process. This group of people can therefore be treated smartly in a way similar to mobility reduction. Consequently, the empirical relationships were used in scenario-based simulations to derive the resulting contact rates in accordance with mobility restoration (representing the levels of reopening economies) and mobility reduction, physical distancing, and/or vaccination interventions under certain level of population density.

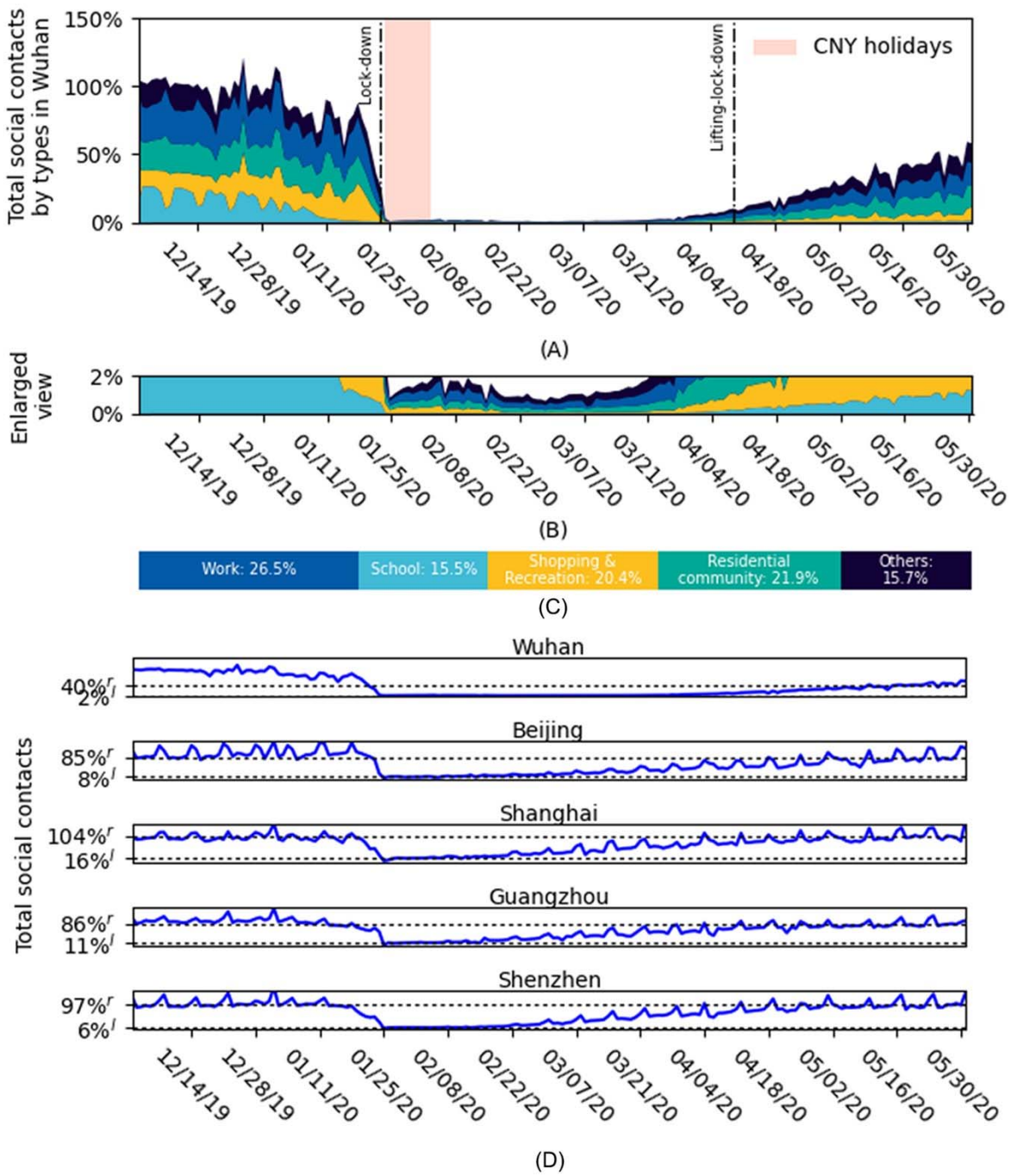

Figure 1. Change in total social contacts index (TSCI) in Wuhan and four other major cities in China (Beijing, Shanghai, Guangzhou, and Shenzhen). **(A)** Change of TSCI in Wuhan from December 2019 through May 2020 in the form of a percentage of the average pre-lockdown level in December 2019 (100%). **(B)** The enlarged view of (A) on the part of TSCI between 0% and 2%. **(C)** The proportions of TSCI that occurred in different colored categories of places in December 2019. Different colors in (A) and (B) represent the same as those in (C). **(D)** The change of TSCI in the five cities. "r" denotes the pre-lockdown TSCI that the city restored to post-lockdown and "l" the pre-lockdown TSCI that the city decreased to

during the lockdown. The two vertical lines in (A) denoting the lockdown and lockdown-lifting dates are only applicable to Wuhan as other cities announced the two dates differently. The Chinese New Year (CNY) holiday period is highlighted by a colored area.

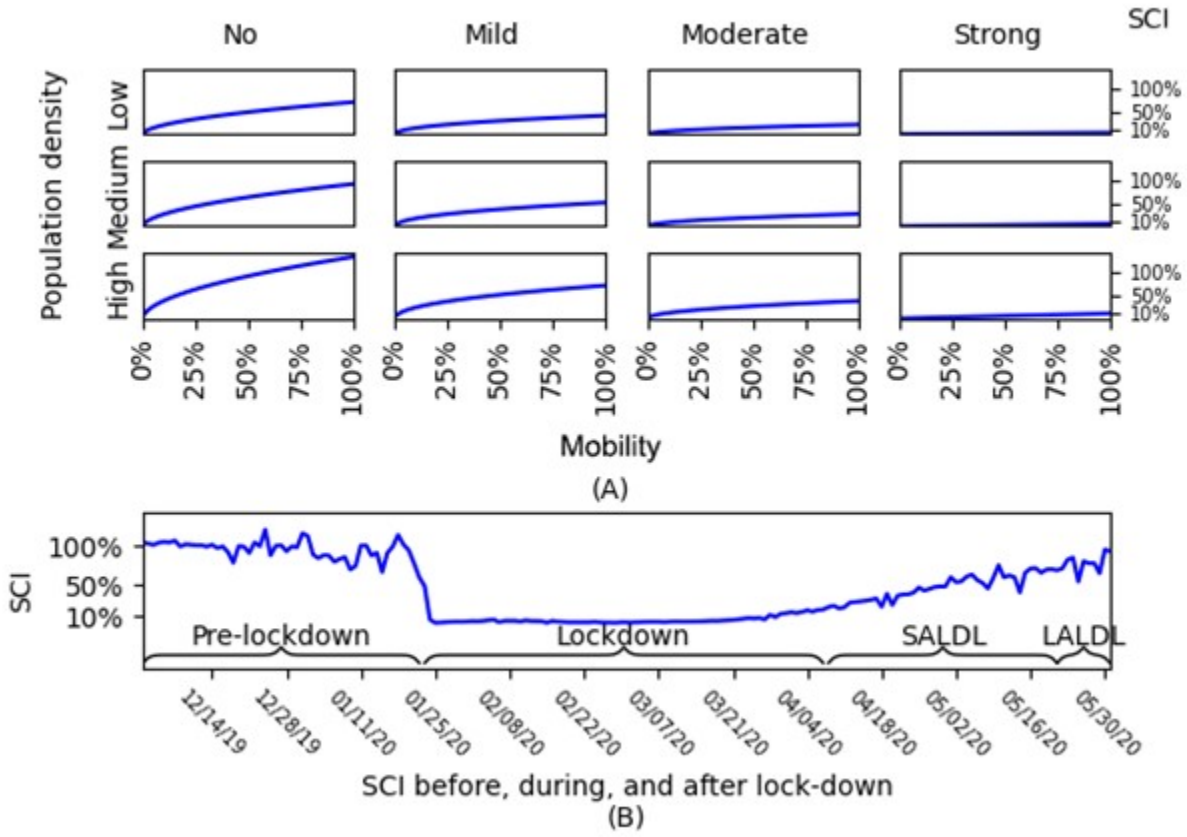

Figure 2. Change in the mobility - social contact index (SCI) relationship in Wuhan under different levels of population density and physical distancing. (A) Each chart shows the changing trend of SCI (in the form a percentage of the averaged pre-lockdown SCI) over the mobility level increasing from 0% (no population movement), to 25%, 50%, and finally 100% (no stay-at-home order) under a certain population density (low, medium, or high) for a specific intensity of physical distancing (no, mild, moderate, or strong). The four columns of charts from left to right are denoted with varying intensities of physical distancing imposed during different periods in Wuhan, i.e., pre-lockdown, LALDL, SALDL, and lockdown, respectively. **(B)** The daily change of SCI (in the form of a percentage of the averaged pre-lockdown SCI) in Wuhan from December 2019 through May 2020.

Reconstruction of COVID-19 transmissions using a mobility and contact-based SEIR (MC-SEIR) model. A classical SEIR model was modified to accommodate both intra-city and inter-city mobility and social contacts (forming a mobility and contact-based SEIR, MC-SEIR) to rebuild the transmission process of COVID-19 in Wuhan from December 2, 2019 through March 31, 2020. Because the social contact rate directly affects the transmission process, to more precisely assess the impact of physical distancing

measures on transmission, a dynamic daily transmission rate derived from SCI was leveraged to replace the fixed transmission rate in the conventional SEIR model.

Specifically, the mobility and SCI data were utilized to estimate the instantaneous effective reproduction number (R_t) via a generalized linear model. The transmission dynamic was calibrated using a Bayesian optimization method[42] with the reported case data of Wuhan (see Methods). The model predicted daily new cases over the period from December 2019 through March 2020 with relatively high accuracy (R^2 = 0.95, Figure 3). Under a non-linearity assumption, the Kendall's tau correlation between SCI and R_t was examined, with the coefficient of 0.59 (95% CI: 0.47 to 0.72; p <0.001; n = 61). If using the mobility data alone, the correlation coefficient decreased substantially to 0.41 (95% CI: 0.30 to 0.54; p <0.001; n = 61). This result suggested that SCI was more strongly rank-correlated with R_t than mobility.

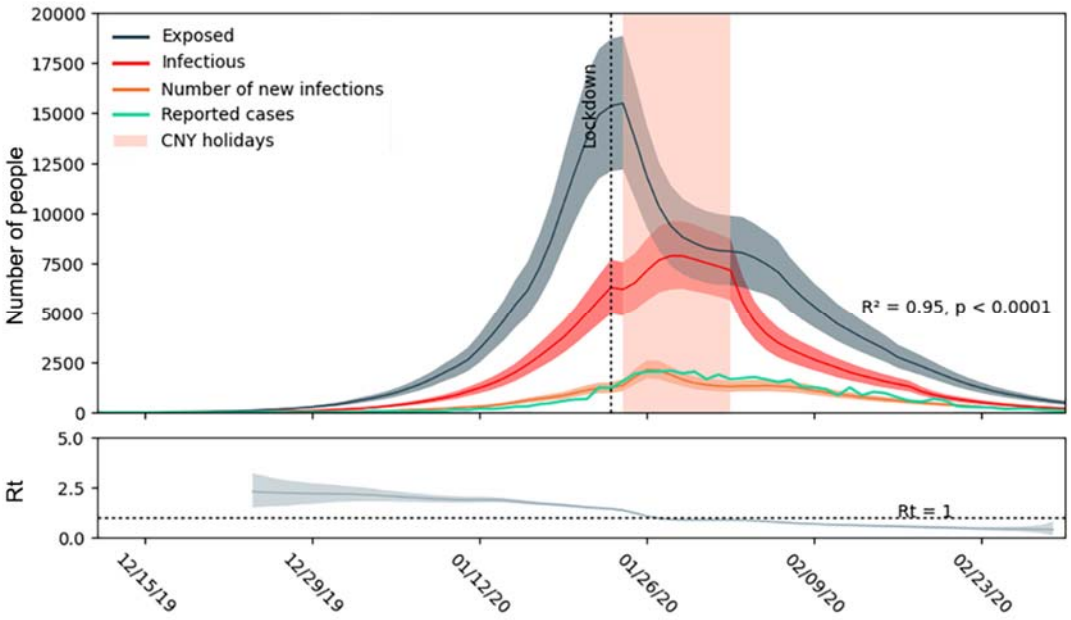

Figure 3. Fitted curves and R_t as predicted by the mobility and contact-based SEIR model. From top to bottom, the three curves in the upper chart represent the estimation of the daily exposed, infectious, and number of new infections from December 2019 through March 2020. The number of new infections examined against the daily reported cases yields the R^2 of 0.95 (n = 92) at the statistical significance level of 0.001. The corresponding daily R_t over the same period is displayed in the lower panel. The Chinese New Year (CNY) holiday period is highlighted by a colored area.

Effect of physical distancing interventions on future resurgences without vaccination. The effectiveness of interventions in preventing a COVID-19 resurgence was assessed by factoring in mobility, physical distancing, and population density under a no-vaccination scenario. The median duration required to contain a resurgence was estimated for each scenario, as shown in Figure 4. Specifically, physical distancing measures were applied under a certain level of mobility and a selected population density scenario when the new cases exceeded 10 per day. The measures were lifted after no new cases were

registered for 14 days. The scenarios and corresponding simulation results are expected to be useful in designing preventive interventions against COVID-19 for other cities worldwide with similar population density variations (see Methods).

Physical distancing is deemed necessary for cities with a low (30% of the pre-lockdown population density in Wuhan), medium (50% of the pre-lockdown population density in Wuhan), or high (100% of the pre-lockdown population density in Wuhan; 1282 people/km2) population density to curb resurgences of coronavirus infections (Figure 4). Moderate and strong physical distancing measures could help a city with a low population density to relax interventions within three months if mobility reduction (e.g., reduced to 50% pre-lockdown level) was simultaneously applied. However, for a city with a high population density, a combination of mobility reduction and physical distancing measures would be needed to contain the resurgence within nine months. We found that the duration of intervention implementation [259 days; inter-quartile range (IQR): 162-345] for a city with high population density would almost triple that for a city with a population density of 50% or lower (86 days; IQR: 71-104) if both moderate physical distancing interventions plus 50% pre-lockdown mobility reduction were adopted. Thus, to shorten the duration for a city with high population density, the physical distancing measures of strong intensity would be suggested.

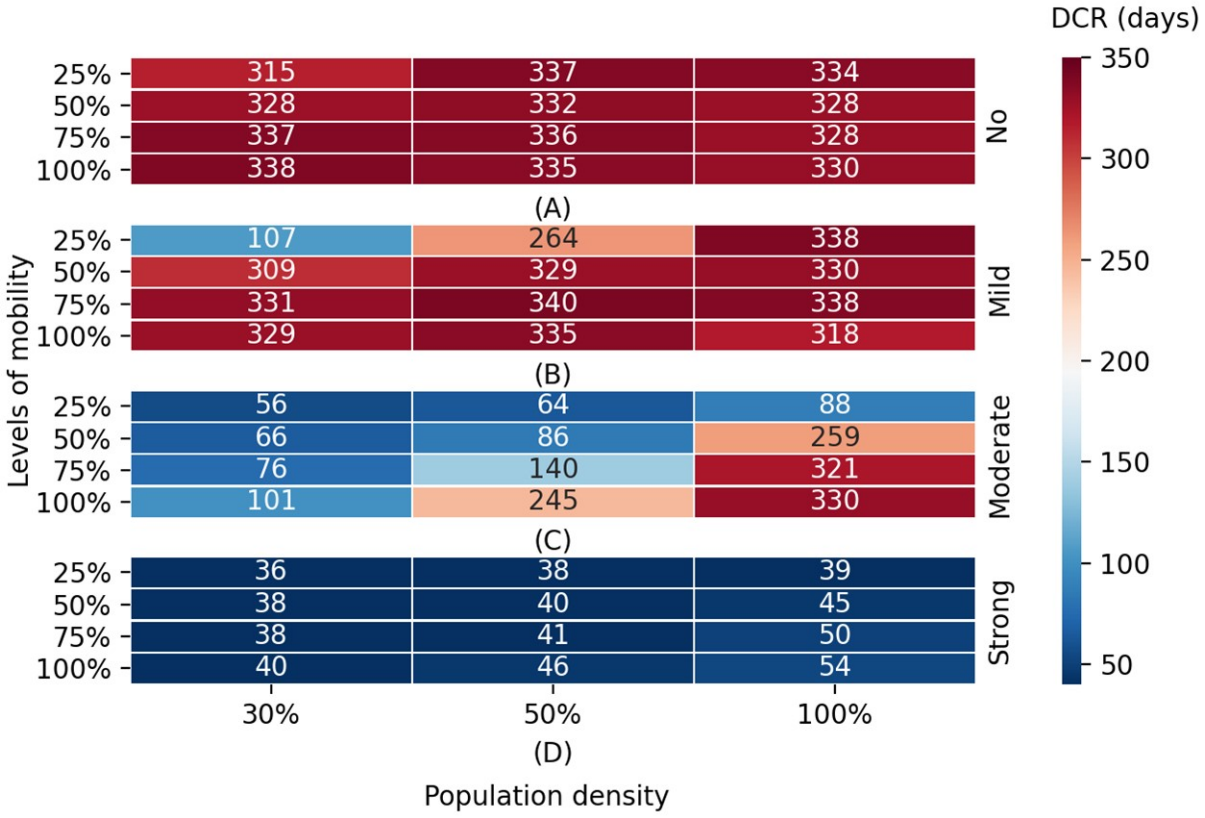

Figure 4. Estimated effects of control measures on containing a resurgence of infections under different population densities. (A)-(D) show, for four intensities of physical distancing (no, mild, moderate, and strong), respectively, the median duration required to contain a resurgence (DCR, unit: days) in Wuhan with four levels of pre-lockdown mobility (25%, 50%, 75%, and 100%) under different population densities. The value in each cell denotes the DCR with respect to its associated level of mobility, intensity of physical

distancing, and population density. The population densities are 30%, 50%, and 100%, respectively, of that in Wuhan pre-lockdown.

Joint effects of vaccination and physical distancing in avoiding resurgences. Vaccines against SARS-CoV-2 have become available since late 2020, and it is imperative to assess the potential combined effects of vaccination and physical distancing, especially there will be a period when only limited supplies of vaccines are available or only a limited proportion of people (lower than the theoretical herd immunity threshold) are vaccinated. Thus, we designed a set of scenarios where physical distancing measures were leveraged to end the epidemic for a city with an increasing proportion of vaccinated population. In total, 64.2% of the population (i.e., the herd immunity threshold derived from the median of a set of reported R_0 values[43]) would be vaccinated within one year. The effectiveness of vaccines (the seroprotection rate) was set as 75% (neutral scenario) (see Methods). According to the simulation results (Table 1), the combination of physical distancing and vaccination was predicted to further reduce the number of infected cases compared with vaccination alone. The reduction effects were predicted to be more significant for cities with a high population density, such as Wuhan. Specifically, 97.72%, 99.99%, and 99.99% of the infections were predicted to be avoided under mild, moderate, and strong physical distancing intensities, respectively. Strong and moderate physical distancing together with vaccination were predicted to suppress the infections to low levels, i.e., 213 (95% CI: 122–347) and 1800 (95% CI: 991–2821) cases in one year, respectively, and thus were recommended to be adopted. Meanwhile, the total duration of physical distancing to end the resurgences decreased gradually from 350 days (95% CI: 338–354; mild) to 234 days (95% CI: 166–395; moderate) and then 43 days (95% CI: 33–64; strong). The results implied that strong but short physical distancing would be a better solution for curtailing resurgences in terms of the case number and intervention duration.

Compared with the no-vaccination scenario, vaccination combined with physical distancing was predicted to contain the resurgence without relying on mobility reduction, whereas a gradual vaccination process alone could not achieve this. Specifically, for cities with low population density, physical distancing would no longer be required. For cities with medium population density, vaccination could shorten the duration of physical distancing measures required to end the resurgence by 36%–78% and limit the number of infected cases to 298 (95% CI: 130–438) and 776 (95% CI: 368–1064) under moderate and mild physical distancing, respectively. For cities with high population density, vaccination enabled strong physical distancing to be replaced by moderate physical distancing. Consequently, it is of value to apply joint physical distancing and vaccination interventions while approaching herd immunity, especially in large cities with high population densities, such as Wuhan.

To evaluate uncertainties, the above simulations were replicated under pessimistic and optimistic scenarios with the effectiveness set as 50% and 100%, respectively. Without loss of generality, the population density was set as the 100% pre-lockdown population density of Wuhan for the optimistic, neutral, and pessimistic scenarios. The results showed that the uncertainty (IQR of daily new cases) would be high when applying limited physical distancing measures. In the pessimistic scenario, the peak of daily new infections was 1.25 to 8.08 times that in the neutral scenario when only no or mild physical distancing measures were applied, whereas the peak of daily new cases was similar (1.08 and 0.97 times) when

moderate or strong physical distancing was imposed (Figure 5). This further confirmed that the joint implementation of physical distancing and vaccination can reduce the uncertainty in ending the epidemic.

To test the generalizability of our proposed SCI models under varying population densities, six cities in China (low-density: Zhuzhou and Qiqihar; medium-density: Hefei and Hangzhou; and high-density: Beijing and Chengdu) were examined with each density set forming a group under one population density scenario. The four sets of physical distancing measures (e.g., "the closure of schools and 20% reduction in contact in all other categories" for mild intensity) were directly applied to these cities, but an SCI-mobility curve was derived for each city (see Methods). According to the results (Supplementary Table 4), all of the low-density scenarios required no physical distancing when vaccination was applied. The medium-density cities (Hefei and Hangzhou) had lower than 1500 cases in one year even with mild physical distancing. Thus, we could further infer that the mild, moderate, and strong physical distancing measures would all be acceptable for medium-density cities according to the number of potential cases in one year. This result is similar to what was confirmed in the 50% population density scenario in Wuhan (Supplementary Table 4). In all three high-density scenarios, the no and mild physical distancing measures failed to stop the resurgence and reduce the high number of infections. Thus, the moderate and strong physical distancing measures would be acceptable for high-density cities when vaccines became available. Strong physical distancing measures should be applied first because this intensity would likely end the need for interventions within two months. The results in all six cities showed similar results to those in the corresponding population density scenarios in Wuhan. This evidence suggests that our proposed mobility-SCI model and the explicit sets of physical distancing measures could be used to inform the combined effects of interventions in other cities with similar population densities.

Table 1. Simulated joint effects of vaccination and physical distancing measures

Population density	Physical distancing	Infected cases in one year (95% CI)	Reduction rate of cases[a]	Duration under physical distancing
30% population density of Wuhan	No	47 (9-189)	-	-
	Mild	54 (11-196)	-	0 (0-0)
	Moderate	49 (11-194)	-	0 (0-0)
	Strong	59 (16-218)	-	0 (0-0)
50% population density of Wuhan	No	6894 (1302-11565)	-	-
	Mild	776 (368-1064)	88.74%	163 (90-242)
	Moderate	298 (130-438)	95.68%	55 (35-80)
	Strong	205 (94-347)	97.02%	29 (0-45)
100% population density of Wuhan	No	17.72% (16.26%-18.74%)[b]	-	-
	Mild	0.40% (0.22%-0.60%)[b]	97.72%	350 (338-354)
	Moderate	1800 (991-2821)	99.99%	234 (166-295)
	Strong	213 (122-347)	99.99%	43 (33-64)

[a] The reduction rate of cases refers to the percentage of cases that could be reduced if a physical distancing intensity (listed on the left) were applied compared with a no physical distancing scenario. The no physical distancing scenario and the scenarios with a very limited number of cases that do not

require physical distancing are marked as "-."

[b] The number of cases is presented in the form of a percentage of potentially exposed population (approximately 102.25 million in Wuhan) given the large numbers; see Methods for more details.

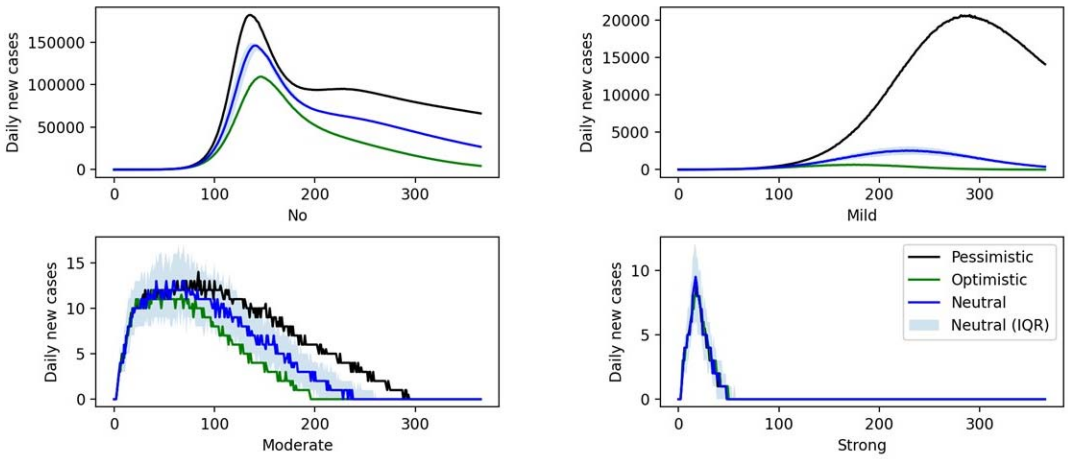

Figure 5. Joint effects of vaccination and physical distancing. The daily new cases are simulated under the scenarios of varying vaccine effectivenesses (optimistic, pessimistic, and neutral) and physical distancing intensities (no, mild, moderate, and strong). The *y*-axis represents the daily new cases and the *x*-axis the number of days since the start of the simulation. All the curves share the same graphical legend in the chart of strong physical distancing. For the neutral scenario, the 25% and 75% quantiles of daily new cases are also displayed.

Discussion

Our study assessed the effects of physical distancing interventions on the resurgence of COVID-19 with increased mobility and social contacts following the lifting of lockdown measures under both vaccination and no-vaccination scenarios. The effectiveness and duration of physical distancing interventions in containing future resurgences highly depends on the intensity of measures, population density, and the availability of vaccines across space and time. Large cities with a high population density, such as Wuhan (1282 people/km^2), are more vulnerable to resurgence when reopening the economy and society. On the one hand, the naturally high social contact rates in these cities would require more effort (longer and stronger physical distancing and stronger mobility reduction) than that necessary for less dense cities. On the other hand, containing a resurgence in the absence of an effective vaccine is a costly process, especially for cities with high population density; it would require two months for the containment even with strong physical distancing measures. Thus, it is important to take precautionary measures against future resurgences in the upcoming several months before effective vaccines become widely available. Socioeconomic activities can be safely restored to normal levels if vaccinated population reach the herd immunity threshold (64.2%) as derived from the R_0 value (2.79). Before achieving herd immunity, however, if there are cases imported into regions where outbreaks have been fully contained, the cities with high

population density would still require moderate or strong physical distancing measures, whereas the cities with low population density could curtail the transmission solely by vaccination.

Effective vaccination with a high coverage in population can greatly ease the reliance on physical distancing interventions in terms of their implementation intensity and duration. Specifically, stay-at-home orders (mobility and contact reductions used in our simulation) would no longer be needed even only with a gradual vaccination process (till reaching 64.2% of the population in one year). However, only the cities with low population density can fully interrupt the transmission without implementing any physical distancing measures. For cities with medium or high population density, physical distancing would still be required to lower the infection number. Meanwhile, the joint implementation of vaccination and physical distancing can limit the uncertainty of transmissions caused by the ineffectiveness of vaccines or short-term immunities. This result suggests strong but short physical distancing interventions might be more effective than mild but long ones when the long-term effectiveness of vaccines has yet to be confirmed.

Although this study showed that non-pharmaceutical interventions are likely to substantially reduce COVID-19 transmission, it is difficult for the public to adhere to travel and physical distancing measures for a long period[6,44,45], and subsequent waves of resurgence may emerge after relaxing interventions and resuming normal levels and patterns of travel[6,23,24] before achieving herd immunity through vaccination. Our research provides a framework and set of outputs that can be used across a wide range of settings, including (i) more precise estimates of COVID-19 outbreaks and the efficacy of interventions under both vaccination and no-vaccination scenarios; (ii) identification of the most effective combinations of physical distancing and vaccination interventions and their intensities for preventing or suppressing resurgences; and (iii) supporting disease control strategy design through improved understanding of interventions and their effects across space and time in regions with different population densities.

Our findings should be considered in the context of several assumptions and data limitations. First, we did not derive individual-level mobility and social contacts to estimate COVID-19 transmission owing to the data availability and privacy issues. If demographic attributes of individuals are available in the future[32,33,46], then we could extend the methodology and analyses to detect potential social disparities in the vulnerability to COVID-19 and assess potential heterogeneities in the efficacy of intervention. Second, the mobile phone-derived data in China for parameterizing travel and physical distancing interventions in our SCI model might not be representative of populations in other regions or countries owing to variations in user coverage, population dynamics across regions, and human behavior patterns. However, the data used here cover over 70% of the population in China[31], and thus represent the Chinese population reasonably well. Third, the accuracy of our model relies on the accuracy of epidemiological parameters derived from reported case data, the quality of which might be constrained by case definitions, the capacity for diagnosis and surveillance, and other factors varying across countries, regions, and time[46–48]. Fourth, other factors and interventions, such as hand washing and wearing facemasks, may also contribute to mitigating COVID-19 spread across space and time[46,49,50], but our simulations did not specify their contributions to transmission. Fifth, the exemplified intervention measures and vaccination settings in our simulation scenarios might not be complete, especially given the ongoing and rapidly changing challenges of vaccine supply and logistics as well as the potential impact of new virus variants on the vaccine effectiveness[11,51]. Although the heterogeneity in human behavior and interventions across regions or countries may limit the generalizability of our model and findings, this study provides an evidence-based assessment of the joint effect of physical

distancing and vaccination interventions on COVID-19 in other cities worldwide with similar levels of population density, settings, and human behavior patterns. Additionally, our methods integrate near-real time mobility and social contact data as well as vaccination scenarios, thereby suggesting that our approaches can be adapted to address emergent needs given the rapid changes in the COVID-19 transmission dynamics post-lockdown.

Methods

Our research complies with relevant ethical regulations. No statistical methods were used to predetermine the sample size in fitting the relationship between contact rates and mobility. However, our sample was large enough as the original data cover more than 70% of the population in China (e.g., 19,728 spatiotemporal bins in Wuhan were used as samples). The individual characteristics (e.g., sex and age) and the exact number of smartphone users were not available as aggregated data were used. We were also aware that users opted in to provide their location information. The aggregated data included all available users. In other words, there was no randomization in collecting the data of social contact events.

Case incidence data. The daily numbers of COVID-19 cases by date of illness onset in Wuhan as of April 17, 2020, as obtained from the national information reporting system for notifiable infectious diseases in China, were used to further evaluate the performance of the baseline model. There was an abnormal increase in the number of cases in Wuhan on February 1, 2020, based on the date of illness onset. We interpolated the number on February 1 as the mean number of cases reported on January 31 and February 2 in the epicurve.

Estimation of population migration and associated population density variations. Population migration played an important role in our assessment of the population in the study area (i.e., Wuhan) and subsequently significantly impacted the transmission dynamics. Population migration between cities was detected, and the corresponding data were acquired from Tencent's mobile device dataset. For a given day (t), the relative values of the population inflow ($I_{r,t}$) and outflow ($O_{r,t}$) were calculated by detecting the number of move-in and move-out mobile device users. These values were used to estimate the actual inflow (I_t) and outflow (O_t) together with the relative value of the population (U_t) (number of mobile device users after deduplication) and the ambient population (P_t) (the average total population in a particular location, e.g., a grid cell or a city, throughout a period, e.g., one hour or one day) on day t (i.e., daily ambient population). Finally, the P_t was updated based on the estimated inflow and outflow on that day, as follows:

$$I_t = I_{r,t} * \frac{P_t}{U_t}, \quad (1)$$

$$O_t = O_{r,t} * \frac{P_t}{U_t} \quad (2)$$

$$P_t = P_{t-1} + I_t - O_t \quad (3)$$

The population density of a city is influenced by population migration, so it varies over time. Population density may directly affect the base contact rate between individuals. In our simulations, the P_t of Wuhan at the beginning of the study period was obtained from the government statistical data [i.e., 11.21 million (Wuhan's permanent population); December 1, 2019][52] and updated on a daily basis using the migration data. Consequently, the population density also changed daily during the study period, but it was assumed to be homogeneous within the city.

In addition, the cases that were registered in the city, regardless of being members of permanent population or members of migration population, would be included in the total number of cases in a simulation scenario. Thus, to better evaluate the risk of COVID-19 to the population under this statistical caliber[30], we used an equivalent population base named as "potentially exposed population", which was calculated by summing the city's permanent population and its annualized inflow population[53]. The annualized inflow population refers to the weighted sum of daily inflow population over one year. For each daily inflow, its weight is the proportion of the number of days between the inflow date and the end of the year in a whole year. For example, the weight of the inflow population on the first day is 364/365.

Social contact measurement. The numbers of total potential contact events (i.e., TSCI) and outside-home trips (mobility) were derived from Tencent's mobile device geolocation dataset (Extended Data Fig. 2). By using the government statistical data (2020)[54] and considering the population coverage of this dataset, the time-varying gridded population was then obtained and aggregated to estimate the population stock number in any specified space-time bin according to the user coverage rate. Each bin was indexed with a spatial coordinate *l* and a temporal coordinate *j* and also labeled with a land use type *k* (e.g., residential community, work, school, shopping/recreation, or others). TSCI denotes the overall scale of potential social contact (or contact events) occurring in a city within a given time period (e.g., 1 day). Each contact event refers to a one-time co-presence within a space-time bin (Extended Data Fig. 1). The number of such events can be calculated as follows:

$$TSCI_t = \sum G_{l,j,k}\left(G_{l,j,k} - 1\right) \tag{4}$$

$$G_{l,j,k} = \frac{U_{l,j,k}}{C} \tag{5}$$

where $G_{l,j,k}$ denotes the ambient population in bin (l, j, k), $U_{l,j,k}$ the number of mobile device users in the same bin, and C the conversion ratio, which is a constant.

On a per-person level, the SCI on day t can subsequently be calculated (Extended Data Fig. 2) as follows:

$$SCI_t = \frac{TSCI_t}{P_t}, \tag{6}$$

where P_t is the total ambient population of the study area on day *t* (aggregated from all the bins over all data collection time intervals).

The SCI computed using mobile geolocation data can help one to understand the transmission process with a higher spatiotemporal resolution by providing the number of potential social contacts on a per-person basis. To explore the non-linear relationship between SCI and mobility, 10% to 100% (incremented by 10%) of mobile device users were randomly selected, and their corresponding contact events were detected in the

form of spatiotemporal co-presence (Extended Data Fig. 1). Thus, a series of mobility-SCI pairs was formed, based on which an empirical relationship was built to estimate SCI values in scenario-based simulations, given an assumed restoration or reduction in mobility levels.

Quantifying the relationship between mobility and social contacts. To quantify the association between mobility and the SCI, we used an allometric curve[34,35], as follows:

$$SCI_t = \alpha \times (h \times P_t/A)^{\beta} \quad (7)$$

where A is the size of the study area and $\frac{P_t}{A}$is the population density. h denotes a ratio for quantifying the mobility level of the entire population (i.e., P_t), which can also reflect the level of mobility reduction/restoration. For instance, a stay-at-home order followed by approximately 50% of the population can be simulated by setting h as 0.5. Thus, the mobility is calculated as $h \times P_t$ and the mobility per unit area is calculated as $h \times \frac{P_t}{A}$. α and β together define a power law equation representing the influence of physical distancing measures. Higher values of α and β indicate weaker physical distancing, thereby implying more social contact under a given population density and mobility level. In addition, an inflow of population can cause an increase in P_t and an increase in population density and SCI_t.

To fit this model, the social contact data of Wuhan for one week during the pre-lockdown period (December 1 to December 7, 2019) were extracted and used to generate a set of mobility-SCI observations. The social contacts were categorized into five types according to the places where they occurred, namely the residential community (C_r), work (C_w), school (C_{sch}), shopping/recreation (C_s), and others (C_o).

$$TSCI = C_r + C_w + C_{sch} + C_s + C_o \quad (8)$$

The values of each of these types of contact under a given mobility level (from 1% to 100%) were also recorded and denoted as $C_{k,i}$, where $k \in \{r, w, sch, s, o\}$ and $i \in \{1\%, 2\%, 3\%, \ldots, 100\%\}$. The mobility level was obtained by randomly sampling the overall mobile device users (represented by i) and calculating the contact events belonging to the given category k. The contact value ($SCI_{t,i}$) with respect to a given mobility level i and a physical distancing intervention was then derived (Extended Data Fig. 2), where $p_{k,i}$, denoting a ratio between 0 and 1, was used to quantify the contacts of a given category k under the mobility level i on day t, as follows:

$$SCI_{t,i} = (\sum p_{k,i} \times C_{k,i,t})/P_t \quad (9)$$

A set of $SCI_{t,i}$ and i values were generated under a given physical distancing intervention specified by α and β and a population density (P_t/A). These observations were used to quantify the relationship between a specified mobility level and an SCI value for a given population density. Therefore, the α and β values for a given physical distancing intervention could be obtained as follows:

$$SCI_{t,i} = \alpha \times (i \times h \times P_t/\mathrm{A})^{\beta} \quad (10)$$

The above process was conducted four times to determine the parameters α and β under varying intensities of physical distancing interventions, namely no, mild, moderate, and strong. Using the mobility data of Wuhan from December 2019 through May 2020, we were able to derive the social contact rates at different places and infer their opening and closure during the periods in which physical distancing

interventions of varying intensities were implemented. As shown in Supplementary Table 1, the four intensities were consistent with those imposed in the following periods: pre-lockdown, LALDL, SALDL, and lockdown. Each set of measures was as follows. No physical distancing refers to the normal status pre-lockdown; mild physical distancing includes the closure of schools and a 20% reduction in contact in all other categories; moderate physical distancing includes the closure of schools, an 80% reduction in contact in shopping/recreation, and a 50% reduction in contact in all other categories; and strong physical distancing includes only 50% of the pre-lockdown-level contact in residential communities and the closure of all other non-essential facilities.

The above considerations indicate that the four intensities of physical distancing can be implemented explicitly because each intensity is linked to a set of control measures (e.g., closure of schools and 20% closure of other services), which can then inform the changes in the SCI. These changes then indicate the potential variations in disease transmission.

To validate the proposed mobility-SCI relationships under the four intensities of physical distancing or the four sets of explicit physical distancing measures (Figure 2), the relationships were examined against the observed mobility-SCI values in Wuhan over the corresponding periods of pre-lockdown, LALDL, SALDL, and lockdown under the same population density (low, medium, or high). The low population density was set to approximately 30% of the population of Wuhan, medium population density to 50% of the population, and high population density to 100% of the population. Under each physical distancing intensity, the modeled mobility-SCI relationships for the three population densities followed the same allometric function[34,35]. Thus, only four relationships required validation (Extended Data Fig. 3).

The results indicated that the modeled SCI values under each mobility level (10% to 100% with increments of 10%), that is, for each relationship, were significantly correlated with the observed SCI values (Pearson correlation coefficients ranging from 0.95 to 0.97 for all of the relationships, each with $p < 0.001$; mean absolute error ranging from 0.03 to 1.42; n = 10). Therefore, the proposed mobility-SCI metric was considered useful for measuring the actual intensities of physical distancing that existed over the aforementioned periods for different population densities. Accordingly, this metric was used to quantify the intensities of physical distancing in our simulations.

SEIR computation. For the purpose of this computation, the population (N) was subdivided into five groups, namely susceptible (S), exposed (E), infectious (I), recovered/removed (R), and vaccinated (V). During each time step, five sub-steps were performed sequentially, as follows:

$$N = S + E + I + R + V \tag{11}$$

$$\frac{dS}{dt} = S - E_I \frac{SI}{N} + \xi_R R + \xi_V V \tag{12}$$

$$\frac{dE}{dt} = E_I \frac{SI}{N} - E_c \tag{13}$$

$$\frac{dI}{dt} = E_c - r_t I \tag{14}$$

$$\frac{dR}{dt} = r_t I - \xi_R R \tag{15}$$

$$\frac{dV}{dt} = V_{e,t} - \xi_V V \quad (16)$$

First, the disease transmission coefficient E_I was calculated based on a transmission process following a Poisson $(\lambda = R_s)$ distribution[55]. The theoretical number of newly exposed people $E_I * I$ was then multiplied by the percentage of susceptible people $(\frac{S}{N})$ in the city. Here, the SCI-adjusted transmission rate (R_s) was obtained from the basic reproduction rate (R_0) (2.2; 95% CI: 1.4–3.9), divided by the average number of days (t_g) (5.8; 95% CI: 4.3–7.5) between the onset and first medical visit and isolation, and weighted using the level of social contact (S_c) determined using the mobility data. S_c was computed using a generalized linear model consisting of the parameters β_0, β_1, and β_2. The observed SCI values were derived from the TSCI divided by the total ambient population of the city when fitting the models. In the simulation processes, the SCI values were derived using our proposed physical distancing and mobility reduction measures. We assumed that the mean incubation period for exposed people was 4 days (IQR: 2–7 days)[11]. A set of onset dates (i.e., current time point plus a stochastic incubation period) was stochastically generated and recorded together with the onset dates of the previous exposed population, as follows:

$$R_s = S_c * \frac{R_0}{t_g} \quad (17)$$

$$S_c = \beta_0 * SCI^{\beta_1} + \beta_2 \quad (18)$$

Second, the exposed people were considered infectious if their onset dates (specified in the previous sub-step) were equal to the current time step (i.e., day *t*). The number of such people is denoted by E_c. Typically, a direct estimation based on the total number of existing exposed people and the conversion rate (σ) derived from the delay in symptom onset distribution were applied. For instance, 20% of the exposed population on day $t-1$ would be converted to the infected population if σ is equal to 0.2. However, such an approach can cause a premature conversion of the exposed population to the infectious population, e.g., the abrupt peak in mass social contact before the CNY holidays might have resulted in an immediate increase in the exposed population but a delayed increase in the infectious population. In contrast, a zero-delay peak of new infectious people would appear on the next day if only the number of people exposed in the current timestep were considered instead of their potential onset days.

Third, the infected people were later removed/recovered at an average rate of r_t (recovery/removal rate) where $t \in \{1, 2, 3, 4, 5\}$, which represents the five periods identified in Wuhan[39,40]. A dynamic r_t was applied in light of the significant changes in the diagnosis and isolation strategies implemented during December 2019 to March 2020. r_t was modeled as an optimizable parameter and was determined using a Bayesian optimization method[42]. Everyone in the recovered/removed group lost immunity at a rate of ξ_R every day during the period.

Fourth, when fitting against the real situation in Wuhan, the initial model did not include the vaccination group and vaccination process (the vaccinated population was 0). However, they were included in the SEIR model in simulating the combined effects of vaccination and physical distancing. On day *t*, part of the vaccinated population $(V_{e,t})$ was immunized, and everyone in the vaccinated group lost immunity at a rate of ξ_V every day.

Finally, the exported and imported population were processed to update the total population in the city using the migration data. The exported population followed the same fractions of the susceptible, exposed,

infectious, and recovered/removed population in the current timestep in the city, while the imported population was considered to join the susceptible population. This indicated a limitation that the exposed and infectious populations could only be input at the start timestep of simulations.

Optimization of parameters for SEIR modeling. We developed the MC-SEIR model by modifying the classical SEIR model with mobility and social contact data to reconstruct the transmission dynamics of COVID-19 in Wuhan between December 2019 and March 2020. The model was calibrated within a Bayesian optimization framework by using a tree-structured Parzen estimator (TPE)[42], in which the relationship between social contact and transmission rate as specified by the generalized linear model, the removal rate r_t, and the initial cases were optimizable. These parameters were estimated by minimizing the squared error between the model-estimated daily new cases and the actual case report data. To deal with the uncertainty in the stochastic SEIR model, each parameter set was evaluated 150 times. The mean squared errors were finally used as the object function values by the TPE.

We also tested the non-linear correlation between the SCI (4 days ahead) and R_t by using the case report data of Wuhan for the period of December 1, 2019 through March 31, 2020. A 4-day (median value of the incubation period) time lag was applied to examine the lagged correlation between R_t and SCI.

Simulation for containing resurgences through physical distancing without vaccination. The resurgence of COVID-19 is highly possible, and there is a high risk of resurgence in the near future. Therefore, it is imperative to devise appropriate physical distancing interventions that can help to effectively contain potential resurgences. We simulated resurgence under different intervention strategies and levels of mobility and evaluated the effectiveness of the strategies using the median duration required to contain the resurgence. A physical distancing intervention was commenced under a certain level of mobility when the number of daily new cases exceeded a threshold (i.e., 10 people in the simulation). Thus, the SCI could be controlled under the proposed intervention, along with the level of mobility, to decelerate the transmission process. The SCI value was acquired from the SCI curve defined earlier with respect to the level of mobility, intensity of physical distancing, and population density.

Ideally, the number of new cases would decrease owing to significant reductions in the TSCI and SCI. The interventions were lifted when there were no new cases for 14 consecutive days. Otherwise, the measures were continuously implemented for 1 year (the remainder of the total simulation period). In other words, an estimated duration of more than 300 days indicated that the conducted interventions could not contain the resurgence effectively.

Simulation of the joint effects of vaccination and physical distancing. To understand the joint effects of vaccination and physical distancing in the cities with varying population densities, a set of scenarios differentiated by vaccination, physical distancing, and population density were designed. During the simulation period (i.e., 365 days), the same number of people would be vaccinated every day (approximately 0.18% of the total population). By the end of the simulation period, 64.2% of the population would be vaccinated. The proportion of the vaccinated population ($1-1/R_0$) was derived from R_0 (2.79), i.e., the median R_0 value as reported in a set of previous studies[43] (Supplementary Table 2).

The vaccinated population was assumed to inject two doses, and gradually obtained immunity to SARS-CoV-2. Specifically, six types of COVID-19 vaccines[56–61] that had finished phase II trials (Supplementary Table 3) were reviewed. The probability of inducing an immune response (probability of seroconversion) was recorded on different observation days (e.g., 75%; day 14). These records were later grouped by date and used to calculate the quantiles (i.e., 25%, 50%, and 75%) of seroconversion on each date (e.g., 14, 28, or 42 days since the first shot) (Extended Data Fig. 4). In our simulations, the median value of seroconversion was used. Under pessimistic, neutral, and optimistic scenarios, 50%, 75%, and 100% of the population that had experienced seroconversion would acquire immunity, respectively. The population with immunity could not infect or be infected by other people.

The long-term immunity loss was also considered. The population can obtain immunity by either recovery from infection or vaccination, but the achieved immunity will fade at different speeds. For the recovered group, we assumed that their immunity would follow a similar decreasing curve such as that for severe acute respiratory syndrome (SARS), which is also caused by coronavirus, because there was no available systematic review or report on the immunity fading rate of COVID-19 recovered patients.

Specifically, we assumed that 6.12% of the recovered population would lose their immunity to COVID-19 in the first year after their recovery[62]. For the vaccinated group, we assumed that their immunity would fade at a higher speed. Thus, a unique fading curve was adopted to simulate the immunity fading of the vaccinated population. However, there were also no data on the long-term effects of COVID-19 vaccines as well as other coronaviruses (e.g., SARS and Middle East respiratory syndrome-related coronavirus). Therefore, the fading trend of influenza vaccination was used instead. We assumed that 53.05% (95% CI: 45.79%–60.29%)[63] of the vaccinated population would lose their immunity in the first year.

Simultaneously, physical distancing measures (with mild, moderate, or strong intensity) would commence when daily new cases exceeded a threshold (i.e., 10), which would later be lifted if there were 14 consecutive days with no new cases. It was considered that vaccination and mobility reduction could achieve a similar effect, that is, exposure reduction in population that may potentially cause infection. However, compared with vaccination measures, travel restrictions have serious adverse socioeconomic effects. Moreover, travel restrictions may be difficult to enforce in some countries. Therefore, the physical distance measures that we adopted excluded mobility reduction in this scenario; that is, the mobility was 100% in the simulations.

The scenarios were simulated under three population densities [i.e., 30% (low), 50% (medium), and 100% (high) of the population density of Wuhan] and four physical distancing intensities. Each of the 12 scenarios was run 200 times to evaluate the uncertainty. In the simulations, contacts caused by people belonging to the vaccinated and removed/recovered groups were removed from the TSCI because they could not infect or be infected by other people. Finally, the cumulated infected population and physical distancing durations were reported with 95% CIs in Table 1.

Application of the joint vaccination and physical distancing interventions to other cities. To examine the extensibility of our proposed SCI models grouped by population density, we replicated the simulations of the combined effects of vaccination and physical distancing measures in six other cities, namely Zhuzhou and Qiqihar (low density), Hefei and Hangzhou (medium density), and Beijing and Chengdu (high density), and compared them with Wuhan's population density scenarios. For each city, the mobility-SCI

relationships were first extracted under no, mild, moderate, and strong physical distancing intensities. In this process, the set of measures under a physical distancing intensity in Wuhan remained unchanged to examine if the measures can be feasibly applied to a city with a similar population density. The extracted SCI curves were later used to derive the SCI value of a given city under various physical distancing and vaccination scenarios. The simulations, which were the same as those for Wuhan, were then replicated, in which all of the conditions remained unchanged and the major differences in cities were in their different SCI-mobility curves. Finally, the estimated number of cases (in a 1-year period) and duration of physical distancing were reported as indicators for the assessment.

Data availability

Datasets of clinical and laboratory data of COVID-19 are available at the following GitHub repository: https://github.com/gubb673/MC-SEIR. Population dynamics data obtained from Tencent for this study are not publicly available due to stringent licensing agreements, but the information on the process of requesting access to the data that supports the findings of this study is available from the corresponding author.

Code availability

Code for the model simulations is available at the following GitHub repository: https://github.com/gubb673/MC-SEIR.

Ethics declarations

The collection and analysis of COVID-19 case data were determined by the National Health Commission of China to investigate and control the outbreak. Ethical clearance for collecting and using secondary data in this study was granted by the institutional review board of the University of Southampton (No. 61865). All data included in models were supplied and analyzed in an anonymous and aggregated format, without access to personal identifying information.

References

1. Li, Q. *et al.* Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* **382**, 1199–1207 (2020).
2. World Health Organization. *Coronavirus disease (COVID-19) – World Health Organization.* https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (2020).
3. Guan, D. *et al.* Global supply-chain effects of COVID-19 control measures. *Nat Hum Behav* **4**, 577–587 (2020).
4. de Souza, W. M. *et al.* Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil. *Nat Hum Behav* **4**, 856–865 (2020).
5. Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H. & Lipsitch, M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **368**, 860–868 (2020).
6. López, L. & Rodó, X. The end of social confinement and COVID-19 re-emergence risk. *Nat Hum Behav* **4**, 746–755 (2020).
7. Lai, S. *et al.* Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* **585,** 410–413 (2020) doi:10.1038/s41586-020-2293-x.
8. Bonaccorsi, G. *et al.* Economic and social consequences of human mobility restrictions under COVID-19. *Proc. Natl. Acad. Sci. USA.* **117**, 15530–15535 (2020).
9. Zimmwe, C., Corum, J. & Wee, S.-L. *Coronavirus Vaccine Tracker. The New York Times* https://www.nytimes.com/interactive/2020/science/coronavirus-vaccine-tracker.html (2020) (accessed December 2, 2020).
10. Jeyanathan, M. *et al.* Immunological considerations for COVID-19 vaccine strategies. *Nat Rev Immunol* **20**, 615–632 (2020).
11. Wang, W. *et al.* Global, regional, and national estimates of target population sizes for covid-19 vaccination: descriptive study. *BMJ* **371**, m4704 (2020).
12. Wesolowski, A. *et al.* Quantifying the impact of human mobility on malaria. *Science* **338**, 267–270 (2012).
13. Lai, S. *et al.* Exploring the use of mobile phone data for national migration statistics. *Palgrave Commun* **5**, (2019).
14. Buckee, C. O. *et al.* Aggregated mobility data could help fight COVID-19. *Science* **368**, 145–146 (2020).
15. Oliver, N. *et al.* Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Sci Adv* **6**, eabc0764 (2020).
16. Budd, J. *et al.* Digital technologies in the public-health response to COVID-19. *Nat Med* **26**, 1183–1192 (2020).
17. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet* **395**, 689–697 (2020).
18. Kraemer, M. U. G. *et al.* The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).

19. Tian, H. *et al.* An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* **368**, 638–642 (2020).
20. Jia, J. S. *et al.* Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* **582**, 389–394 (2020).
21. Hsiang, S. *et al.* The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature* **584**, 262–267 (2020).
22. Ruktanonchai, N. W. *et al.* Assessing the impact of coordinated COVID-19 exit strategies across Europe. *Science* (2020) doi:10.1126/science.abc5096.
23. Leung, K., Wu, J. T., Liu, D. & Leung, G. M. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. *Lancet* **395**, 1382–1393 (2020).
24. Nouvellet, P. *et al. Report 26: Reduction in mobility and COVID-19 transmission*. http://spiral.imperial.ac.uk/handle/10044/1/79643 (2020) doi:10.25561/79643 (accessed November 5, 2020).
25. Gao, S. *et al.* Mobile phone location data reveal the effect and geographic variation of social distancing on the spread of the COVID-19 epidemic. Preprint at *arXiv* http://arxiv.org/abs/2004.11430 (2020).
26. Badr, H. S. *et al.* Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases* (2020) doi:10.1016/S1473-3099(20)30553-3 (accessed November 5, 2020).
27. Baidu. Baidu Migration. https://qianxi.baidu.com/ (accessed November 5, 2020).
28. Google. *COVID-19 Community Mobility Reports*. https://www.google.com/covid19/mobility (accessed November 5, 2020).
29. Facebook. Our Work on COVID-19 - Facebook Data for Good. https://dataforgood.fb.com/docs/covid19/ (accessed November 5, 2020).
30. Tencent. Tencent Big Data Platform. https://heat.qq.com/ (accessed November 5, 2020).
31. Iqbal, M. *WeChat Revenue and Usage Statistics (2020)*. https://www.businessofapps.com/data/wechat-statistics/ (2020) (accessed November 5, 2020).
32. Zhang, J. *et al.* Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* **368**, 1481–1486 (2020).
33. Prem, K. *et al.* The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health* **5**, e261–e270 (2020).
34. Hu, H., Nigmatulina, K. & Eckhoff, P. The scaling of contact rates with population density for the infectious disease models. *Math. Biosci.* **244**, 125–134 (2013).
35. Chowell, G., Hyman, J. M., Eubank, S. & Castillo-Chavez, C. Scaling laws for the movement of people between locations in a large city. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **68**, 066102 (2003).
36. Hägerstrand, T. What about people in Regional Science? *Papers of the Regional Science Association* **24**, 6–21 (1970).
37. Kwan, M. GIS methods in time-geographic research: geocomputation and geovisualization of human activity patterns. *Geografiska Annaler: Series B, Human Geography* **86**, 267–280 (2004).
38. Yin, L. & Shaw, S.-L. Exploring space–time paths in physical and social closeness spaces: a space–time GIS approach. *International Journal of Geographical Information Science* **29**, 742–761 (2015).

39. Pan, A. *et al.* Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. *JAMA* **323**, 1915–1923 (2020).
40. Hao, X. *et al.* Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. *Nature* **584**, 420-424 (2020).
41. Cao, S. *et al.* Post-lockdown SARS-CoV-2 nucleic acid screening in nearly ten million residents of Wuhan, China. *Nat. Commun.* **11**, 5917 (2020).
42. Bergstra, J. S., Bardenet, R., Bengio, Y. & Kégl, B. Algorithms for Hyper-Parameter Optimization. in *Advances in Neural Information Processing Systems* 2546–2554 (2011).
43. Liu, Y., Gayle, A. A., Wilder-Smith, A. & Rocklöv, J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J Travel Med* **27**, (2020).
44. Flaxman, S. *et al.* Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* (2020) doi:10.1038/s41586-020-2405-7.
45. Walker, P. G. T. *et al.* The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science* (2020) doi:10.1126/science.abc0035.
46. Allen, W. E. *et al.* Population-scale longitudinal mapping of COVID-19 symptoms, behaviour and testing. *Nat Hum Behav* **4**, 972–982 (2020).
47. Tsang, T. K. *et al.* Effect of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China: a modelling study. *Lancet Public Health* **5**, e289–e296 (2020).
48. Li, Z. *et al.* Active case finding with case management: the key to tackling the COVID-19 pandemic. *Lancet* **396**, 63–70 (2020).
49. Reiner, R. C. *et al.* Modeling COVID-19 scenarios for the United States. *Nat. Med.* 1–12 (2020) doi:10.1038/s41591-020-1132-9.
50. Chu, D. K. *et al.* Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *Lancet* **395**, 1973–1987 (2020).
51. Wise, J. Covid-19: New coronavirus variant is identified in UK. *BMJ* **371**, m4857 (2020) doi: https://doi/org/10.1136/bmj.m4857 (accessed December 19, 2020).
52. Wuhan City Bureau of Statistics, Wuhan's national economic, social development statistical yearbook of 2019 (accessed September 14, 2020).
53. Tencent. Tencent Migration Platform. https://heat.qq.com/qianxi.php (accessed October 18, 2020).
54. National Bureau of Statistics of China, China statistical yearbook, http://www.stats.gov.cn/tjsj/tjgb/ndtjgb/ (accessed September 14, 2020).
55. Kirkeby, C., Halasa, T., Gussmann, M., Toft, N. & Græsbøll, K. Methods for estimating disease transmission rates: Evaluating the precision of Poisson regression and two novel methods. *Sci Rep* **7**, (2017).
56. Logunov, D. Y. *et al.* Safety and immunogenicity of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine in two formulations: two open, non-randomised phase 1/2 studies from Russia. *The Lancet* **396**, 887–897 (2020).

57. Zhu, F.-C. *et al.* Safety, tolerability, and immunogenicity of a recombinant adenovirus type-5 vectored COVID-19 vaccine: a dose-escalation, open-label, non-randomised, first-in-human trial. *The Lancet* **395**, 1845–1854 (2020).
58. Folegatti, P. M. *et al.* Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *The Lancet* **396**, 467–478 (2020).
59. Zhu, F.-C. *et al.* Immunogenicity and safety of a recombinant adenovirus type-5-vectored COVID-19 vaccine in healthy adults aged 18 years or older: a randomised, double-blind, placebo-controlled, phase 2 trial. *The Lancet* **396**, 479–488 (2020).
60. Zhang, Y.-J. *et al. Immunogenicity and Safety of a SARS-CoV-2 Inactivated Vaccine in Healthy Adults Aged 18-59 years: Report of the Randomized, Double-blind, and Placebo-controlled Phase 2 Clinical Trial*. http://medrxiv.org/lookup/doi/10.1101/2020.07.31.20161216 (2020) doi:10.1101/2020.07.31.20161216.
61. Xia, S. *et al.* Effect of an Inactivated Vaccine Against SARS-CoV-2 on Safety and Immunogenicity Outcomes: Interim Analysis of 2 Randomized Clinical Trials. *JAMA* **324**, 951 (2020).
62. Lin, Q., Zhu, L., Ni, Z., Meng, H. & You, L. Duration of serum neutralizing antibodies for SARS-CoV-2: Lessons from SARS-CoV infection. *J Microbiol Immunol Infect* **53**, 821–822 (2020).
63. Song, J. Y. *et al.* Long-term immunogenicity of influenza vaccine among the elderly: Risk factors for poor immune response and persistence. *Vaccine* **28**, 3929–3935 (2010).

Acknowledgements

BH was supported by the Hong Kong Research Grants Council (CRF C4139-20G and AoE/E-603/18), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA19090108), and the National Key R&D Program of China (2019YFC1510400 and 2017YFB0503605). AJT was supported by the Bill & Melinda Gates Foundation (INV-002697, OPP1106427, OPP1032350, OPP1134076, and OPP1094793), the Clinton Health Access Initiative, the UK Department for International Development (DFID) and the Wellcome Trust (106866/Z/15/Z and 204613/Z/16/Z), and the EU H2020 (MOOD 874850). SL was supported by the Bill & Melinda Gates Foundation (INV-024911 and OPP1134076), the National Natural Science Foundation of China (81773498), the National Science and Technology Major Project of China (2016ZX10004222-009), and the EU H2020 (MOOD 874850). NR was supported by the Bill & Melinda Gates Foundation (OPP1170969) and the Vodafone Institute. The funders had no role in study design, data collection and analysis, and decision to publish or preparation of the manuscript.

Author contributions

BH, JXC, JHW, and SL designed the research. JHW and BH built the models and ran simulations. BH, JHW, JXC, and SL carried out analyses. JXC and SQY provided technical support and helped with data curation. BH, JHW, and SL wrote the manuscript. CWR, AC, JRF, NWR, ZL, WY, AJT, PC, HYY, and TT interpreted the findings, commented on, and revised drafts of the manuscript.

Competing interests

The authors declare no competing interests.