# Article

# Effect of non-pharmaceutical interventions to contain COVID-19 in China

Shengjie Lai[1,2,10 ✉], Nick W. Ruktanonchai[1,3,10 ✉], Liangcai Zhou[4], Olivia Prosper[5], Wei Luo[6,7], Jessica R. Floyd[1], Amy Wesolowski[8], Mauricio Santillana[6,7], Chi Zhang[9], Xiangjun Du[9], Hongjie Yu[2] & Andrew J. Tatem[1 ✉]

On 11 March 2020, the World Health Organization (WHO) declared coronavirus disease 2019 (COVID-19) a pandemic[1]. The strategies based on non-pharmaceutical interventions that were used to contain the outbreak in China appear to be effective[2], but quantitative research is still needed to assess the efficacy of non-pharmaceutical interventions and their timings[3]. Here, using epidemiological data on COVID-19 and anonymized data on human movement[4,5], we develop a modelling framework that uses daily travel networks to simulate different outbreak and intervention scenarios across China. We estimate that there were a total of 114,325 cases of COVID-19 (interquartile range 76,776–164,576) in mainland China as of 29 February 2020. Without non-pharmaceutical interventions, we predict that the number of cases would have been 67-fold higher (interquartile range 44–94-fold) by 29 February 2020, and we find that the effectiveness of different interventions varied. We estimate that early detection and isolation of cases prevented more infections than did travel restrictions and contact reductions, but that a combination of non-pharmaceutical interventions achieved the strongest and most rapid effect. According to our model, the lifting of travel restrictions from 17 February 2020 does not lead to an increase in cases across China if social distancing interventions can be maintained, even at a limited level of an on average 25% reduction in contact between individuals that continues until late April. These findings improve our understanding of the effects of non-pharmaceutical interventions on COVID-19, and will inform response efforts across the world.

As of 30 March 2020, the outbreak of COVID-19, which is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has resulted in 693,282 confirmed cases and 33,106 deaths across the world[6]. As the disease has only recently emerged, effective pharmaceutical interventions are not expected to be available for months[7], and health-care resources will be limited for treating all cases. Non-pharmaceutical interventions (NPIs) are therefore essential components of the public health response to COVID-19 outbreaks[6,8–10]. These include the isolation of individuals who are ill, contact tracing, quarantine of exposed individuals, travel restrictions, school and workplace closures, cancellation of mass gatherings, and hand-washing, among others[8–10]. Such measures aim to reduce the transmission of the virus by delaying the timing and reducing the size of the peak of the epidemic, thus buying time for preparations to be made in the healthcare system and creating the potential for vaccines and drugs to be used at a later stage[8].

Three major groups of NPIs have been implemented to contain the spread and reduce the size of the outbreak of COVID-19 across China[11]. First, intercity travel restrictions were used to prevent further seeding of the virus during the Chinese New Year holiday period. A cordon sanitaire of Wuhan and surrounding cities in Hubei province was put in place on 23 January 2020, two days before the Chinese New Year, which started on 25 January 2020. After this date, travel restrictions were also put in place in other provinces across the country. Second, the early identification and isolation of cases was prioritized, including improving the screening, identification, diagnosis, isolation, reporting and contact tracing of people who were suspected or confirmed to have the disease[11]. Local governments across China encouraged and supported the routine screening and quarantine of travellers from Hubei province in an attempt to detect COVID-19 infections as early as possible. The average interval from the onset of symptoms to laboratory confirmation dropped from 12 days in the early stages of the outbreak to 3 days in early February, indicating that these efforts improved detection and diagnosis[3,12]. Third, contact restrictions and social distancing measures, together with personal preventive actions such as hand-washing, were implemented to reduce the risk of exposure at the community level. As part of these social distancing policies, the Chinese government encouraged people to stay at home as much as possible, cancelled or postponed large public events and mass gatherings, and closed libraries, museums and workplaces[13,14]. School holidays were also extended, with the end date of the Chinese New Year holiday period changed from 30 January 2020 to 10 March 2020 for Hubei province, and to 9 February 2020 for many other provinces[15,16].

[1]WorldPop, School of Geography and Environmental Science, University of Southampton, Southampton, UK. [2]School of Public Health, Fudan University, Key Laboratory of Public Health Safety, Ministry of Education, Shanghai, China. [3]Population Health Sciences, Virginia Tech, Blacksburg, VA, USA. [4]Wuhan Center for Disease Control and Prevention, Wuhan, China. [5]Department of Mathematics, University of Tennessee, Knoxville, TN, USA. [6]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, USA. [7]Department of Pediatrics, Harvard Medical School, Boston, MA, USA. [8]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [9]School of Public Health (Shenzhen), Sun Yat-sen University, Shenzhen, China. [10]These authors contributed equally: Shengjie Lai, Nick W. Ruktanonchai. ✉e-mail: shengjie.lai@soton.ac.uk; nr1e14@soton.ac.uk; a.j.tatem@soton.ac.uk

The implementation of these NPIs coincided with a rapid decline in the number of new cases across China, albeit at high economic and social costs[3,12]. Previous studies have examined the effects of the lockdown of Wuhan[17,18], travel restrictions[19], airport screening[20], isolation of cases and contact tracing on the containment of the disease[21]. However, a comprehensive and quantitative comparison of the effectiveness of different NPIs, and the time at which they were implemented, for containing the outbreak of COVID-19 in China is lacking. On the basis of epidemiological data on COVID-19 and historical and near-real-time anonymized data on human movement, we developed a stochastic susceptible–exposed–infectious–removed (SEIR) modelling framework based on travel networks to simulate the spread of COVID-19 across 340 prefecture-level cities in mainland China. Within each city, we estimated the numbers of susceptible, exposed, infectious, and recovered/removed ('removed' refers to the individuals who were isolated to prevent further transmission, and deceased individuals) people per day from 1 December 2019. Using this modelling framework, we conducted before-and-after comparable analyses to quantify the relative effect of the three major groups of NPIs—that is, the restriction of intercity population movement, the identification and isolation of cases, and the reduction of travel and contact within cities to increase social distance—in China. We also assessed the risk of COVID-19 transmission since the lifting of travel restrictions on 17 February 2020.

## Reconstructing the spread of COVID-19

The epidemiological parameters that were estimated for the early stage of the outbreak in Wuhan were initially used to parameterize the epidemic before interventions were widely implemented[5]. The three major groups of NPIs outlined above were derived and measured using data on population movement between and within cities (obtained from smartphone users of Baidu location-based services[4]) and data on the delay between the onset of illness and the reporting of cases across the country. Population travel and contact patterns changed substantially after the implementation of interventions, and the timeliness of case reporting also improved (Fig. 1, Supplementary Tables 1, 2). These indicators were then incorporated into the model (see Methods).

We estimated that there were a total of 114,325 cases of COVID-19 (interquartile range (IQR) 76,776–164,576) in mainland China as of 29 February 2020, 85% of which were in Hubei province (Extended Data Table 1). The outbreak increased exponentially before Chinese New Year, but the peaks of epidemics across the country quickly appeared around the time of Chinese New Year after the implementation of NPIs. The estimated epidemics and peaks were consistent with patterns of reported data by onset date, with strong correlations between daily estimates and reported data across time and regions (Extended Data Fig. 1). The overall correlation between the estimated number of cases and the reported number by province, as of 29 February 2020, was also significant ($P < 0.001$, $R^2 = 0.86$), with a high sensitivity (91%, 280/308) and specificity (69%, 22/32) in predicting cities with or without cases of COVID-19 (Extended Data Fig. 1a, b).

## Quantifying the effect of different NPIs

Without NPIs, our model predicted the number of cases of COVID-19 to increase rapidly across China, with a 51-fold (IQR 33–71) increase in Wuhan, a 92-fold (58–133) increase in other cities in Hubei province and a 125-fold (77–180) increase in other provinces by 29 February 2020. However, the apparent effectiveness of different interventions varied (Fig. 2). The lockdown of Wuhan might not have prevented the seeding of the virus from the city, as the travel ban was put in place at the latter stages of population movement out of the city before Chinese New Year[22] (Fig. 1b). Nevertheless, if intercity travel restrictions had not been implemented, cities and provinces outside of Wuhan would have received more cases from Wuhan, and the affected geographical range
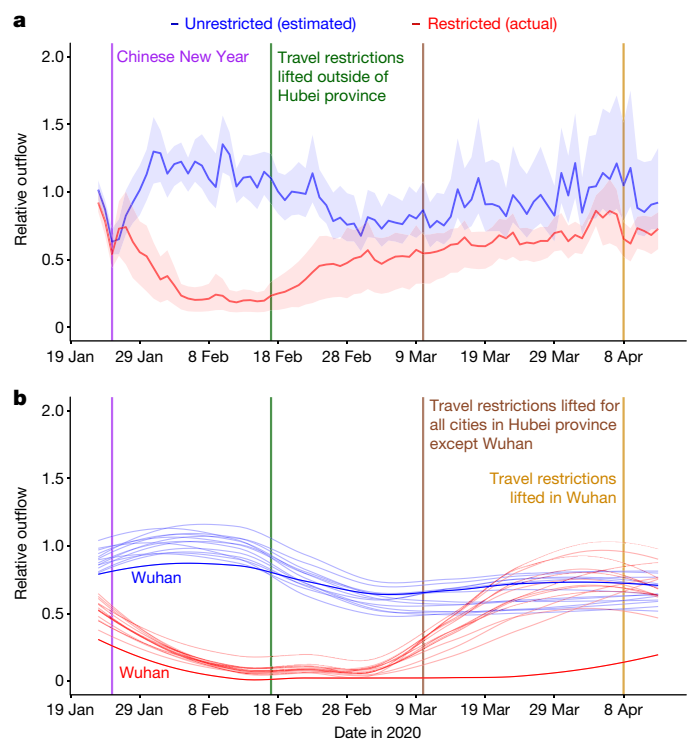


**Fig. 1 | Relative daily volume of outbound travellers from cities across mainland China between 23 January and 13 April 2020. a**, Relative outbound flows of travellers for all cities at prefecture level ($n = 340$) in mainland China, presented as the median (solid line) and IQR (shading). **b**, Relative outbound flows of travellers for cities in Hubei province. Wuhan is highlighted using darker colours. Each red line represents the outflow for each city in 2020, standardized by the mean of daily outflows for each city from 20 to 22 January 2020. Each blue line represents an estimate of the normal outflow by city under the scenario of no travel restrictions (on the basis of travel in previous years). The lines in **b** were smoothed using locally estimated scatterplot smoothing (LOESS) regression.

would have expanded to the remote western areas of China (Extended Data Fig. 2c). In general, we estimated that the early detection and isolation of cases quickly and substantially prevented more infections than did the introduction of contact reduction and social distancing measures across the country (5-fold versus 2.6-fold). However, without the contact reduction intervention, in the longer term the epidemics would have increased exponentially across regions (Fig. 2c, f). Therefore, combined NPIs would bring about the strongest and most rapid effect on containment of the COVID-19 outbreak, with an interval of about one week between the introduction of NPIs and the peak of the epidemic (Extended Data Table 1).

## Timing of interventions

Our model suggests that, theoretically, if interventions in China had been implemented one week, two weeks or three weeks earlier than they actually were, the number of cases of COVID-19 could have been reduced by 66% (IQR 50–82%), 86% (81–90%) or 95% (93–97%), respectively (Fig. 3a). The geographical range of affected areas would also shrink from 308 cities to 192, 130 or 61 cities, respectively (Extended Data Fig. 3). However, if NPIs had been introduced one week, two weeks or three weeks later than they were, the number of cases might have increased by 3-fold (IQR 2–4), 7-fold (5–10) or 18-fold (11–26), respectively (Fig. 3b).
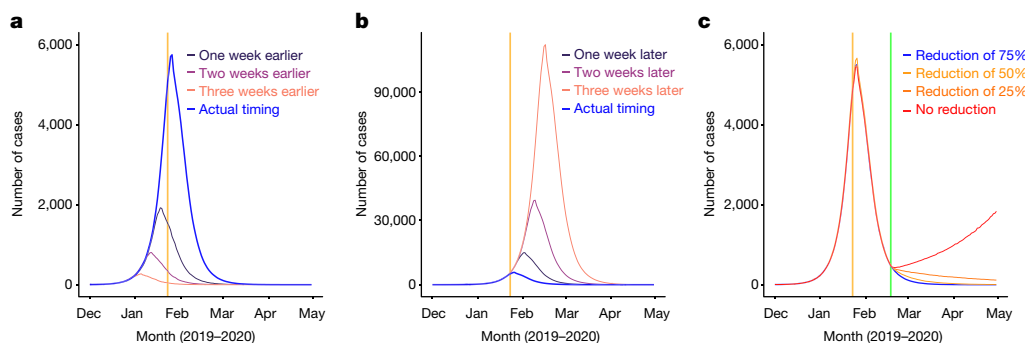
## Lifting of travel restrictions

Under the interventions that were implemented from 17 February 2020—that is, the lifting of travel restrictions—the epidemics outside
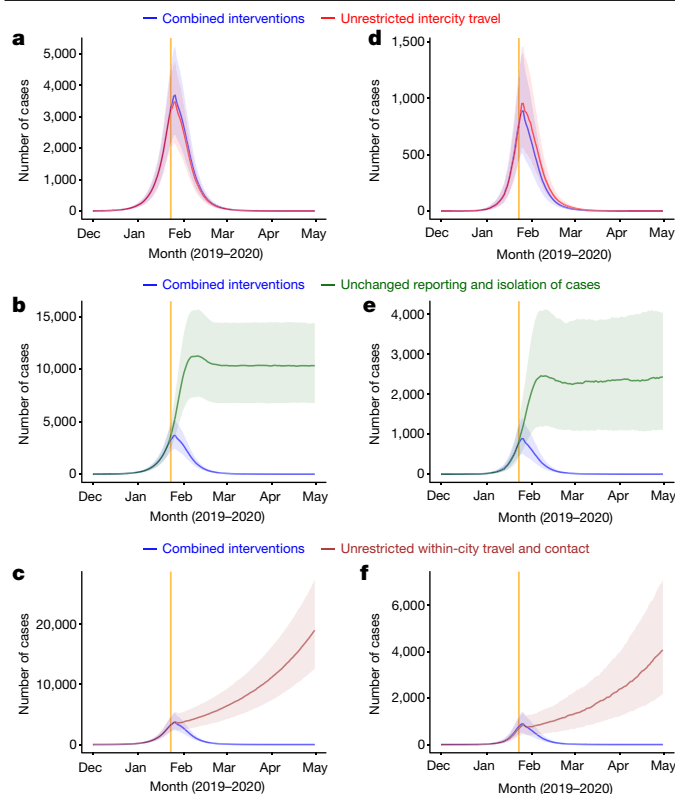
**Fig. 2 | Estimated epidemic curves of the COVID-19 outbreak under various scenarios with or without NPIs by region. a–c**, Estimates for the city of Wuhan. **d–f**, Estimates for cities outside of Hubei province in mainland China. The blue lines represent estimated transmission under combined NPIs, and the other coloured lines represent the scenario without one type of intervention. Data are presented as the median (solid line) and IQR (shading) of estimates (1,000 simulations). The orange vertical lines indicate the date on which the lockdown of Wuhan began (23 January 2020).

of Hubei province probably reached a low level (fewer than 10 cases per day, excluding imported cases from other countries) in early March, whereas Hubei province might need another four weeks to reach the same level as other provinces. However, if population contact resumed to normal levels, the lifting of travel restrictions might cause case numbers to rise again (Fig. 3c). Accordingly, our simulations suggest that maintaining social distancing even to a limited extent (for example, a 25% reduction in contact between individuals on average) through to late April would help to ensure control of COVID-19 in epicentres such as Wuhan.

Our estimates were sensitive to the basic reproduction number ($R_0$); under a higher $R_0$ value, the peaks of epidemics were higher and later, and more time was needed to contain the outbreak (Extended Data Fig. 3). Sensitivity analyses also suggested that our model could have robustly measured relative changes in the efficacy of interventions under different epidemiological parameters and transmission scenarios (Extended Data Figs. 4–9).

## Discussion

Our findings show that combined NPIs substantially reduced the transmission of COVID-19 across China. Earlier implementation of NPIs could have notably reduced the magnitude and geographical range of the outbreak, but—equally—a delayed response would have led to a larger outbreak. China's aggressive, multifaceted response is likely to have prevented a far worse situation, which would have accelerated the spread of the virus globally. The evidence from China provides information that will be of use in efforts to contain the spread of COVID-19 and mitigate the effects of the disease in other regions around the world[3,12].

Our results suggest three key points. First, they support and validate the idea that population movement and close contact has a major role in the spread of COVID-19 within and beyond China[22,23]. As the lockdown of Wuhan happened at the latter stages of population movement before Chinese New Year, travel restrictions did not halt the seeding of the virus from Wuhan, but did prevent cases being exported from Wuhan to a wider area. Second, the importance and effects of the three types of NPIs differed. Compared with travel restrictions, improved detection and isolation of cases, as well as social distancing, probably had a greater effect on the containment of the outbreak. The social distancing intervention reduced contact between people who travelled from the epicentre of the outbreak and other individuals. This is likely to have been especially helpful in curbing the spread of an emerging pathogen to the wider community, and to have reduced the risk of spread from asymptomatic or mild infections[8]. Third, given that travel and work have begun to resume in China, the country should consider at least the partial continuation of NPIs to ensure that the COVID-19 outbreak is sustainably controlled for the first wave of this outbreak. For example, the early identification and isolation of cases should be maintained—which might also help to prevent and delay the arrival of a second wave, considering the increasing numbers of cases that are imported from other countries and the presence of asymptomatic or subclinical infections in China[24].

The analyses presented here provide a comprehensive quantitative assessment of the effect of NPIs on the transmission of COVID-19. The model framework accounts for daily interactions of populations and interventions between and within cities, as well as the inherent



**Fig. 3 | Estimates of the COVID-19 outbreak under various scenarios of intervention timing and lifting of travel restrictions across China. a**, Estimated epidemic curves for interventions implemented earlier than their actual timing. **b**, Estimated epidemic curves for interventions implemented later than their actual timing. **c**, Estimated spread of COVID-19 for different rates of population contact after the lifting of intercity travel restrictions. The orange vertical lines indicate the date on which the lockdown of Wuhan began (23 January 2020), and the green line shows the date on which travel restrictions were lifted (17 February 2020).

statistical uncertainty that is associated with a paucity of epidemiological parameters before and after the implementation of interventions. The network-based SEIR model is methodologically robust and is built on the basic SEIR models that have been used previously to predict the transmission of COVID-19 in its early stages[23]. Considering the delays that exist in the reporting of cases, our approach can be used to enable a rapid, ongoing estimation of the effectiveness of various NPIs in different countries, and to aid decision-making relating to the control of outbreaks of COVID-19.

Our study has several limitations. First, our simulations were based on parameters that were estimated for symptomatic cases identified in the early stage of the outbreak in Wuhan, and might not account for asymptomatic and mild infections; we may therefore have underestimated the total number of infections. Second, our findings could be confounded by other factors that changed during the outbreak. Although we have shown that the apparent fall in the incidence of COVID-19 after Chinese New Year (25 January 2020) in China is likely to be attributed to the interventions taken, we cannot rule out the possibility that the decrease was partially attributable to other unknown seasonal factors—for example, temperature and absolute humidity[25,26]. Third, if the epidemiological parameters of COVID-19 transmission in other cities across China differed from the estimates[5]—which were based on the data in the early stage of the outbreak, when no NPIs were in place in Wuhan—then our estimates of the effectiveness of interventions in reducing the transmission of COVID-19 could be biased. Fourth, there are probably biases in population coverage, given that our model relies on data from mobile phone and Baidu users. Although a high percentage (from 46.9% in 2013 to 55.3% in 2018) of the population of China owns smartphones[27] (https://en.wikipedia.org/wiki/List_of_countries_by_smartphone_penetration), the group of mobile-phone users does not include specific subgroups of the population, particularly children. Therefore, our data on population movement may provide an incomplete picture, and differences between the characteristics of smartphone owners and non-owners may also bias our estimates. In addition, the magnitude and patterns of population movements could change year by year—although previous studies have suggested that travel patterns are consistent in their seasonality across years in China and other countries[22]. Finally, we only examined three main groups of NPIs, and other interventions might also have contributed to the containment of the outbreak. For example, owing to the sources of data that were available, we did not assess the effect of personal hygiene and protective equipment on containing the spread of COVID-19. Other sources of data and further investigations are needed to measure and evaluate the efficacy of each intervention.

COVID-19 has placed a substantial burden on health systems and society across many countries. From a public health standpoint, our results highlight that countries should consider proactively planning NPIs and relevant strategies for containment and mitigation, as the earlier implementation of NPIs could have led to substantial reductions in the size of the outbreak in China. Our results also provide guidance for countries as to the likely effectiveness of different NPIs at different stages of an outbreak. Suspected and confirmed cases of the disease should be identified, diagnosed, isolated and reported as early as possible to control the source of infection, and the implementation of cordon sanitaires or travel restrictions for areas that are heavily affected might prevent the virus spreading to wider regions. Reducing contact and increasing social distance between individuals, together with improved personal hygiene, can help to protect vulnerable populations and mitigate the spread of COVID-19 at the community level, and these interventions should be promoted throughout the outbreak to avoid resurgence. Our findings suggest that—as advocated by WHO—strategies that involve the early implementation of integrated NPIs should be prepared, deployed and adjusted to maximize the benefits of these interventions and minimize the health, social and economic effects of COVID-19 around the world[3].

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2293-x.

1.  World Health Organization. *WHO Director-General's opening remarks at the media briefing on COVID-19 – 11 March 2020* https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 (WHO, 2020).
2.  Epidemiology Working Group for NCIP Epidemic Response, Chinese Center for Disease Control and Prevention. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China [in Chinese with English abstract]. *Zhonghua Liu Xing Bing Xue Za Zhi* **41**, 145–151 (2020).
3.  World Health Organization. *Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)* https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf (WHO, 2020).
4.  *Baidu Migration* [in Chinese] https://qianxi.baidu.com/ (2020).
5.  Li, Q. et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* **382**, 1199–1207 (2020).
6.  World Health Organization. *Coronavirus disease (COVID-19) pandemic* https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (WHO, 2020).
7.  Heymann, D. L. & Shindo, N. COVID-19: what is next for public health? *Lancet* **395**, 542–545 (2020).
8.  Fong, M. W. et al. Nonpharmaceutical measures for pandemic influenza in nonhealthcare settings—social distancing measures. *Emerg. Infect. Dis.* **26**, 976–984 (2020).
9.  Ryu, S. et al. Nonpharmaceutical measures for pandemic influenza in nonhealthcare settings—international travel-related measures. *Emerg. Infect. Dis.* **26**, 961–966 (2020).
10.  Xiao, J. et al. Nonpharmaceutical measures for pandemic influenza in nonhealthcare settings—personal protective and environmental measures. *Emerg. Infect. Dis.* **26**, 967–975 (2020).
11.  Chen, W. et al. Early containment strategies and core measures for prevention and control of novel coronavirus pneumonia in China [in Chinese with English abstract]. *Zhonghua Yu Fang Yi Xue Za Zhi* **54**, 239–244 (2020).
12.  World Health Organization. *Press Conference of WHO-China Joint Mission on COVID-19* https://www.who.int/docs/default-source/coronaviruse/transcripts/joint-mission-press-conference-script-english-final.pdf?sfvrsn=51c90b9e_2 (2020).
13.  The State Council of the People's Republic of China. *The announcement from Wuhan's headquarter on the novel coronavirus prevention and control* [in Chinese] http://www.gov.cn/xinwen/2020–01/23/content_5471751.htm (2020).
14.  The State Council of the People's Republic of China. *The announcement on strengthening community prevention and control of pneumonia epidemic situation of new coronavirus infection* [in Chinese] http://www.gov.cn/zhengce/2020-01/27/content_5472516.htm (2020).
15.  The State Council of the People's Republic of China. *The State Council's announcement on extending the 2020 Spring Festival Holiday* [in Chinese] http://www.gov.cn/zhengce/content/2020-01/27/content_5472352.htm (2020).
16.  The People's Government of Shanghai Municipality. *The announcement on postponing the reoperation date of companies and the reopening date of schools* [in Chinese] http://www.shanghai.gov.cn/nw2/nw2314/nw2315/nw43978/u21aw1423601.html (2020).
17.  Li, X., Zhao, X. & Sun, Y. The lockdown of Hubei Province causing different transmission dynamics of the novel coronavirus (2019-nCoV) in Wuhan and Beijing. Preprint at *medRxiv* https://doi.org/10.1101/2020.02.09.20021477 (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.  Chinazzi, M. et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020).
20.  Quilty, B. J., Clifford, S., CMMID nCoV working group 2, Flasche, S. & Eggo, R. M. Effectiveness of airport screening at detecting travellers infected with novel coronavirus (2019-nCoV). *Euro Surveill*. **25**, 2000080 (2020).
21.  Hellewell, J. et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob. Health* **8**, e488–e496 (2020).
22.  Lai, S. et al. Assessing spread risk of Wuhan novel coronavirus within and beyond China, January–April 2020: a travel network-based modelling study. Preprint at *medRxiv* https://doi.org/10.1101/2020.02.04.20020479 (2020).
23.  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. *Lancet* **395**, 689–697 (2020).
24.  National Health Commission of the People's Republic of China. *Updates on pneumonia of new coronavirus infections as of March 31, 2020* [in Chinese] http://www.nhc.gov.cn/xcs/yqtb/202004/28668f987f3a4e58b1a2a75db60d8cf2.shtml (2020).
25.  Wang, M. et al. Temperature significant change COVID-19 transmission in 429 cities. Preprint at *medRxiv* https://doi.org/10.1101/2020.02.22.20025791 (2020).
26.  Luo, W. et al. The role of absolute humidity on transmission rates of the COVID-19 outbreak. Preprint at *medRxiv* https://doi.org/10.1101/2020.02.12.20022467 (2020).
27.  Lai, S., Farnham, A., Ruktanonchai, N. W. & Tatem, A. J. Measuring mobility, disease connectivity and individual risk: a review of using mobile phone data and mHealth for travel medicine. *J. Travel Med.* **26**, taz019 (2019).

# Article

## Methods

### Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

### Model summary

An SEIR model based on travel networks was built to simulate the spread of COVID-19 between and within all prefecture-level cities in mainland China. This model has been made openly available for further use at https://github.com/wpgp/BEARmod. Population movement data across the country were used to estimate the intensity of travel restrictions and contact reductions. Data from illness onset to reporting of the first index case for each county were used to infer the changing timeliness of case identification and isolation across the course of the outbreak. The outputs of the model under NPIs were validated by using daily numbers of new cases reported across all regions in mainland China. On the basis of this modelling framework, the efficacy of applying or lifting non-pharmaceutical measures under various scenarios and timings were tested and quantified.

### Data sources

Three datasets on population movement, which were obtained from Baidu location-based services that provide over 7 billion positioning requests per day[4,28], were used in this study to measure travel restrictions and social distancing across time and space. The first is an aggregated and de-identified dataset on near-real-time daily relative outbound and inbound flow of smartphone users for each prefecture-level city in 2020 (340 cities in mainland China were included) to understand patterns of mobility during the outbreak. The daily outflow from each city since the lockdown of Wuhan and the travel restrictions that were applied on 23 January 2020 were rescaled by the mean daily flow for each city from 20 to 22 January 2020 for comparing travel reductions across cities and years (Fig. 1).

The second Baidu dataset is a historical relative movement matrix with daily total number of users at the city level from 26 December 2014 to 26 May 2015, aligning with the 2020 Chinese New Year holiday period, for which the corresponding period is 1 December 2019 to 30 April 2020. We assumed that the pattern of population movements was the same in years when there were no outbreaks and interventions. Adjusted by the level of travel reductions derived from the 2020 dataset where applicable, the second dataset was used to simulate the spread of COVID-19 and predict transmission via population movements under various scenarios, with or without intercity travel restrictions. Corresponding city-level population data in 2015 for modelling were obtained from the Chinese Bureau of Statistics[29].

The third Baidu dataset measures daily population movements at the county level (2,862 counties in China) from 26 January to 30 April 2014, as described elsewhere[30]. On the basis of the assumption that the pattern of population contact was consistent across years when there were no interventions, it was used to estimate within-city travel and contact reduction during the outbreak and interventions. First, we aggregated data from county to city level and rescaled the daily flows from 29 January 2014 by the mean of the daily flow for 26–28 January period, aligning with the date of Wuhan's lockdown and the 2020 Chinese New Year holiday period. Then, the rescaled first dataset for 2020 under interventions was compared with the 2014 dataset to derive the percentage of travel decline for each city. The percentages for cities were averaged by day to preliminarily quantify the intensity of contact reduction in China under NPIs (Supplementary Table 2), as the policies of travel restriction and social distancing measures were implemented and occurred at the same time across the country.

We also collated data of the first case reported by county across mainland China to measure the delay from illness to case report as a reference of the improved timeliness of case identification, isolation and reporting during the outbreak (Supplementary Table 1). The daily number of COVID-19 cases by date of illness onset in the city of Wuhan, Hubei province and other provinces as of 13 February 2020 were used to further validate the epidemic curves estimated in this study across time. There was an abnormal increase of cases in Wuhan and Hubei province on 1 February 2020, on the basis of the date of illness onset[2]. We interpolated the number on 1 February 2020 by using the mean of numbers of cases reported on 31 January and 2 February 2020 in the epidemic curve. The number of cases reported by city across mainland China as of 29 February 2020 was used to define the predictability of our model across space. These case data were collated from the websites of national and local health authorities, news media and publications[2,3,31] (Supplementary Information).

### Data analysis

We constructed a travel-network-based SEIR modelling framework (BEARmod) for before-and-after comparable analyses on the efficacy of NPIs. This model was extended from a typical SEIR model to specifically incorporate movement between locations that varied with each time step. In this model, each city was represented in the model as a separate subpopulation, with its own susceptible ($S$), exposed ($E$), infected ($I$) and recovered/removed ($R$) populations.

### Exposure, infection and recovery

During each time step, infected people first recovered or were removed at an average rate $r$, where $r$ was equal to the inverse of the average infectious period, and removal represents self-isolation and effective removal from the population as a potential transmitter of disease. This was incorporated as a Bernoulli trial for each infected person with a probability of recovering of $1 - \exp(-r)$. We used the median of time lags from illness onset to reported case as a proxy of the average infectious period, indicating the improving identification and isolation of cases under improved interventions (Supplementary Table 1). Then, the model converted exposed people to infectious by similarly incorporating a Bernoulli trial for each exposed individual, where the daily probability of becoming infectious $1 - \exp(-\varepsilon)$, where $\varepsilon$ was the inverse of the average time spent exposed but not infectious, on the basis of the estimated incubation period (5.2 days, 95% confidence interval (CI) 4.1–7.0)[5]. Finally, to end the exposure, infection and recovery step of the model, the number of newly exposed people was calculated for each city on the basis of the number of infectious people in the city ($I_i$) and the average number of daily contacts that lead to transmission that each infectious person has ($c$). We simulated the number of exposed individuals in a patch on a given day through a random draw from a Poisson distribution for each infectious person, in which the mean number of new infections per person was $c$, which was then multiplied by the fraction of people in the city that were susceptible. We calculated the daily contact rate $c$ using the basic reproduction rate that has been calculated in other studies ($R_0 = 2.2$ (95% CI 1.4–3.9)) divided by the average days (5.8, 95% CI 4.3–7.5) from onset to first medical visit and isolation[5], weighted by the relative level of daily contact where relevant, based on the Baidu movement data (Supplementary Table 2). Because simulation runs were not extended beyond five months, we did not include the addition of new susceptible people, or the conversion of recovered people back to susceptible.

The infection processes within each patch therefore approximate the following deterministic, continuous-time model, where $c$ and $r$ varied through time:

$$\frac{dS}{dt} = S - c\frac{SI}{N}$$

$$\frac{dE}{dt} = c\frac{SI}{N} - \varepsilon E$$

$$\frac{dI}{dt} = \varepsilon E - rI$$

$$\frac{dR}{dt} = rI$$

## Movement

After the model completed the infection-related processes, we moved infectious people between cities. To do this, we moved infected people from their current location to each possible destination (including remaining in the same place) using Bernoulli trials for each infected person, and each possible destination city. We parameterized the probability of moving from city $i$ to city $j$ ($p_{ij}$), which was equal to the proportion of smartphone users who went from city $i$ to city $j$ in the corresponding day from the Baidu dataset in 2015, accounting for the travel restrictions in 2020. This included modelling the numbers of people who stayed in the same location using $p_{ii}$, the proportion of users who did not move to a new location on that day. This allowed us to incorporate variance in the actual composition of travellers (infected versus non-infected), but because movement numbers were generated independently, it was possible for the number of infected people who stayed and the number who move in each patch to exceed or be fewer than the number of infected people in the patch. As we only wanted to incorporate variance into relative patterns of movement and not absolute numbers (particularly because the underlying values are proportions of people who moved and therefore cannot influence the total numbers of people infected), in any case in which the number of infected people who moved and the number who stayed differed from the total number of infected people in the origin patch, we rescaled values to the total number of infected people. Rescaling in this way meant the variance introduced by the Bernoulli trials could only influence relative movement patterns, and not actual numbers of infected people. Further, because we explicitly model the number of stayers in the same way as movers, rescaling should not introduce any bias in terms of the final relative movement patterns.

Through this model, stochasticity in the numbers and in the places with COVID-19 infections appears between simulation runs owing to variance in numbers of people becoming exposed, infectious and removed/recovered, as well as variance in numbers of people moving from one city to another. By modelling the COVID-19 epidemic in this way, we could simulate the incidence of COVID-19 cases, accounting for variance in recovery, infection and movement across many simulation runs (1,000). In addition, this allowed for us to account for uncertainty in contact rates after NPIs were implemented or lifted.

## Simulation runs

Using this model, we quantified how the transmission of COVID-19 varied with different intervention scenarios and timings, as well as the potential of further transmission after the lifting of travel restrictions and contact distancing measures on 17 February 2020. As the earliest date of illness onset in cases was 2 December 2020 (ref. [3]), considering the underreporting of cases and the delay from infection to onset and identification of this novel virus, we started our simulations by infecting five people in Wuhan on 1 December 2019 and propagating the epidemic through time, varying factors including the timing and types of interventions used, assumed contact and recovery rates, and movement. We initially infected five people as a minimum number of infected people that prevented stochastic extinction of the epidemic during the initial days of simulation, and found no significant difference after three months, over simulation runs that started with three, five and eight people initially infected (though with three people initially infected, 50% of runs led to zero cases over the first week of simulation). When using data from other years we fixed the simulation dates

around Chinese New Year and adjusted the start date of the epidemic accordingly.

The estimates of the model for the outbreak under NPIs as the baseline scenario were compared with reported COVID-19 cases across time and space. The sensitivity and specificity were also calculated to examine the performance of the model in predicting the occurrence of COVID-19 cases at the city level across China. The relative effects of NPIs were quantitatively assessed by comparing estimates of cases under various NPIs and timings with that of the baseline scenario. We also conducted a series of sensitivity analyses to understand the effect that changing epidemiological parameters had on the estimates and uncertainties of intervention efficacy. The software R v.3.6.1 (R Foundation for Statistical Computing) was used for data collation and analyses.

## Ethical approval

Ethical clearance for collecting and using secondary data in this study was granted by the institutional review board of the University of Southampton (no. 48002). All data were supplied and analysed in an anonymous format, without access to personal identifying information.

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

The data on the number of cases of COVID-19 reported by county, city and province across China are available from the data sources listed in the Supplementary Information, and the average days from illness onset to reporting of the first case by each county used in the modelling are detailed in Supplementary Table 1. The mobile phone datasets analysed during the current study are not publicly available as this would compromise the agreement with the data provider; however, information on the process of requesting access to the data that support the findings of this study is available from S.L., and the data on travel and contact reductions that were derived from the datasets and used in our model are detailed in Supplementary Table 2.

## Code availability

The code for the model built in this study has been made openly available for further use at https://github.com/wpgp/BEARmod.

28. Wang, X., Liu, C., Mao, W., Hu, Z. & Gu, L. Tracing the largest seasonal migration on Earth. Preprint at https://arxiv.org/abs/1411.0983 (2014).
29. National Bureau of Statistics of China. *China Statistical Yearbook 2014* http://www.stats.gov.cn/english/Statisticaldata/AnnualData/ (2020).
30. Kraemer, M. U. G. et al. Past and future spread of the arbovirus vectors *Aedes aegypti* and *Aedes albopictus*. *Nat. Microbiol.* **4**, 854–863 (2019).
31. Zhang, J. et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *Lancet Infect. Dis.* **20**, 793–802 (2020).

# Article

findings, commented on and revised drafts of the manuscript. J.R.F., O.P. and W.L. built the model, commented on and revised drafts of the manuscript. C.Z. collected data, interpreted the findings and commented on and revised drafts of the manuscript. A.J.T. interpreted the findings and revised drafts of the manuscript. A.W., M.S., X.D. and H.Y. interpreted the findings and commented on and revised drafts of the manuscript. All authors read and approved the final manuscript. The corresponding authors had full access to all of the data in the study and had final responsibility for the decision to submit for publication.

**a** Wuhan City, Hubei Province

**b** Other cities in Hubei Province

**c** Other provinces

**d** Estimated versus reported

**Extended Data Fig. 1 | Estimated and reported epidemic curves of the COVID-19 outbreak in mainland China. a**, The city of Wuhan in Hubei province. **b**, Other cities in Hubei province. **c**, Thirty other provincial regions in mainland China. The orange vertical lines indicate the date on which Wuhan's lockdown began (23 January 2020). The estimated epidemic curves of COVID-19 cases show the median (dark-blue line) and IQR (light-blue shading) of estimates (1,000 simulations), and the Pearson's correlation between the median of daily estimates and the number of daily reported cases by region as of 13 February 2020 is also shown. **d**, The Pearson's correlation between the total number of estimated cases and the total number of reported cases by province as of 29 February 2020. *P* value calculated by two-sided *t*-test.
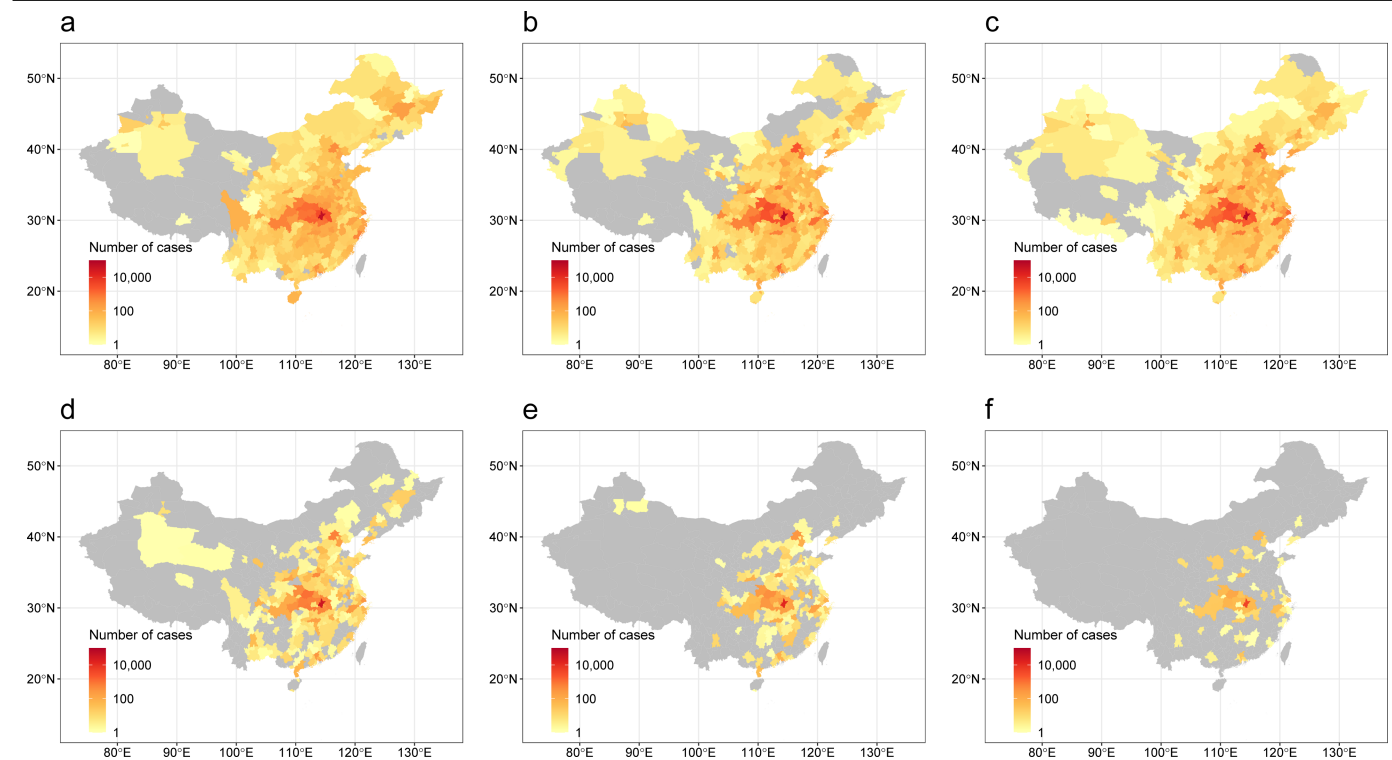
**Extended Data Fig. 2 | Areas affected by COVID-19 in mainland China under various intervention timings. a**, A total of 308 cities reported COVID-19 cases, on the basis of data obtained from national and local health authorities, as of 29 February 2020. **b**, Affected areas (298 cities) estimated by models under interventions implemented at actual timing. **c**, Estimated affected areas (326 cities) under interventions implemented at actual timing, but without intercity travel restrictions. **d**, Estimated affected areas (192 cities) under interventions implemented one week earlier than actual timing. **e**, Estimated affected areas (130 cities) under interventions implemented two weeks earlier than actual timing. **f**, Estimated affected areas (61 cities) under interventions implemented three weeks earlier than actual timing. The administrative boundary maps were obtained from the National Platform of Common Geospatial Information Services of China (www.tianditu.gov.cn).
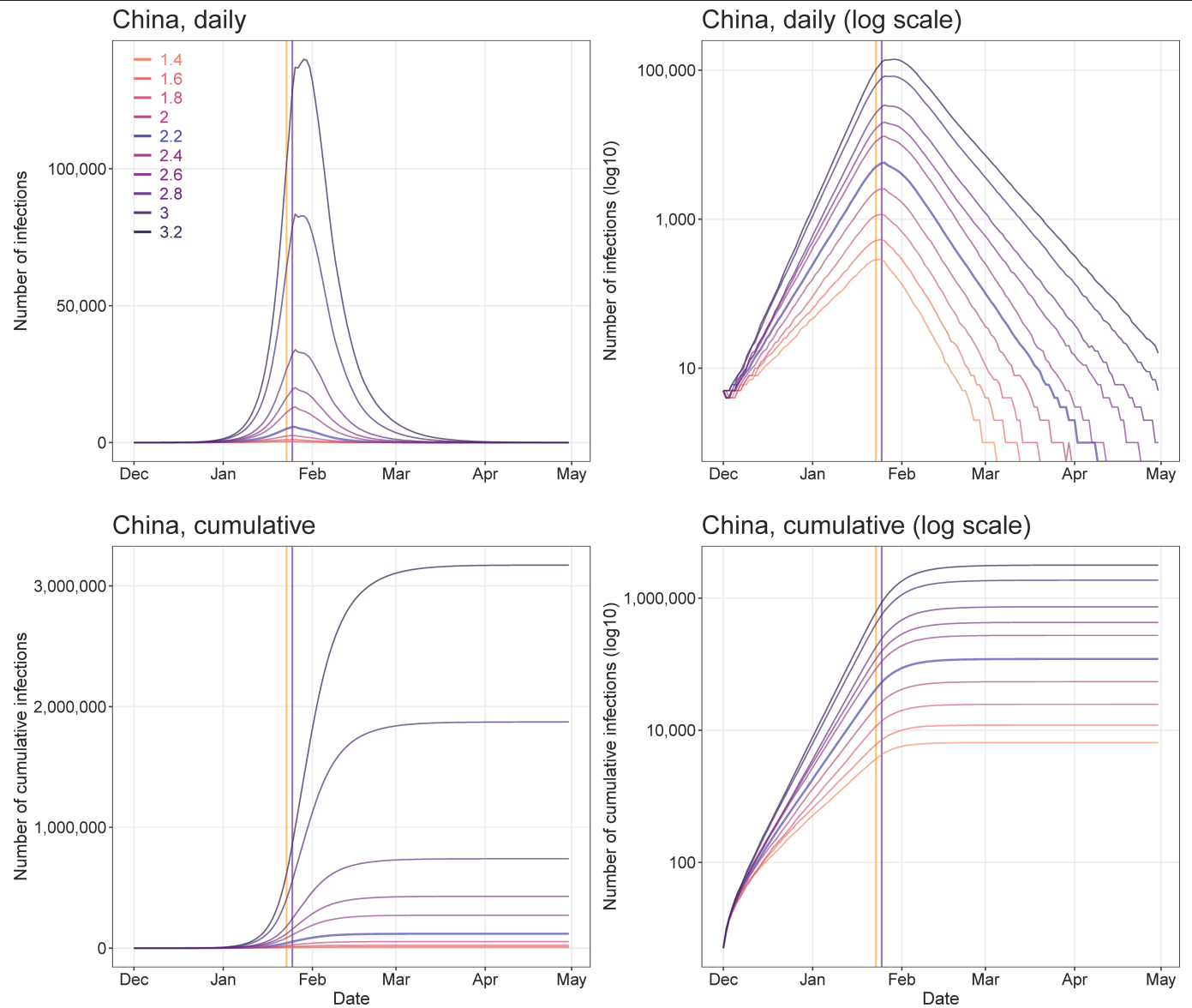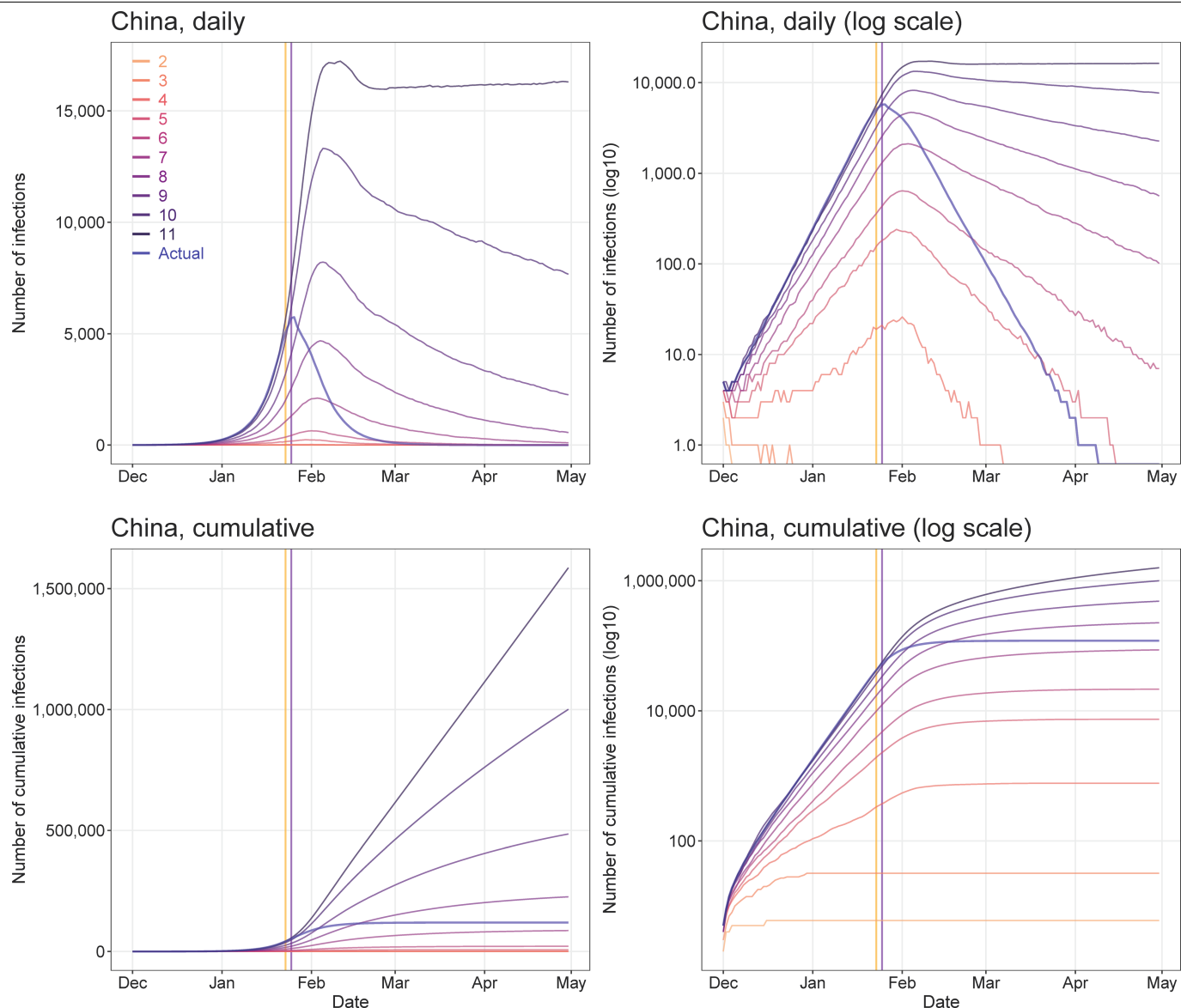
**Extended Data Fig. 3 | Sensitivity of estimates of COVID-19 epidemics for various values of $R_0$.** All other parameters, NPIs and input data were the same as the baseline model with $R_0 = 2.2$. Orange vertical lines indicate the date on which the lockdown of Wuhan began (23 January 2020); purple vertical lines indicate the date on which the Chinese New Year began (25 January 2020).
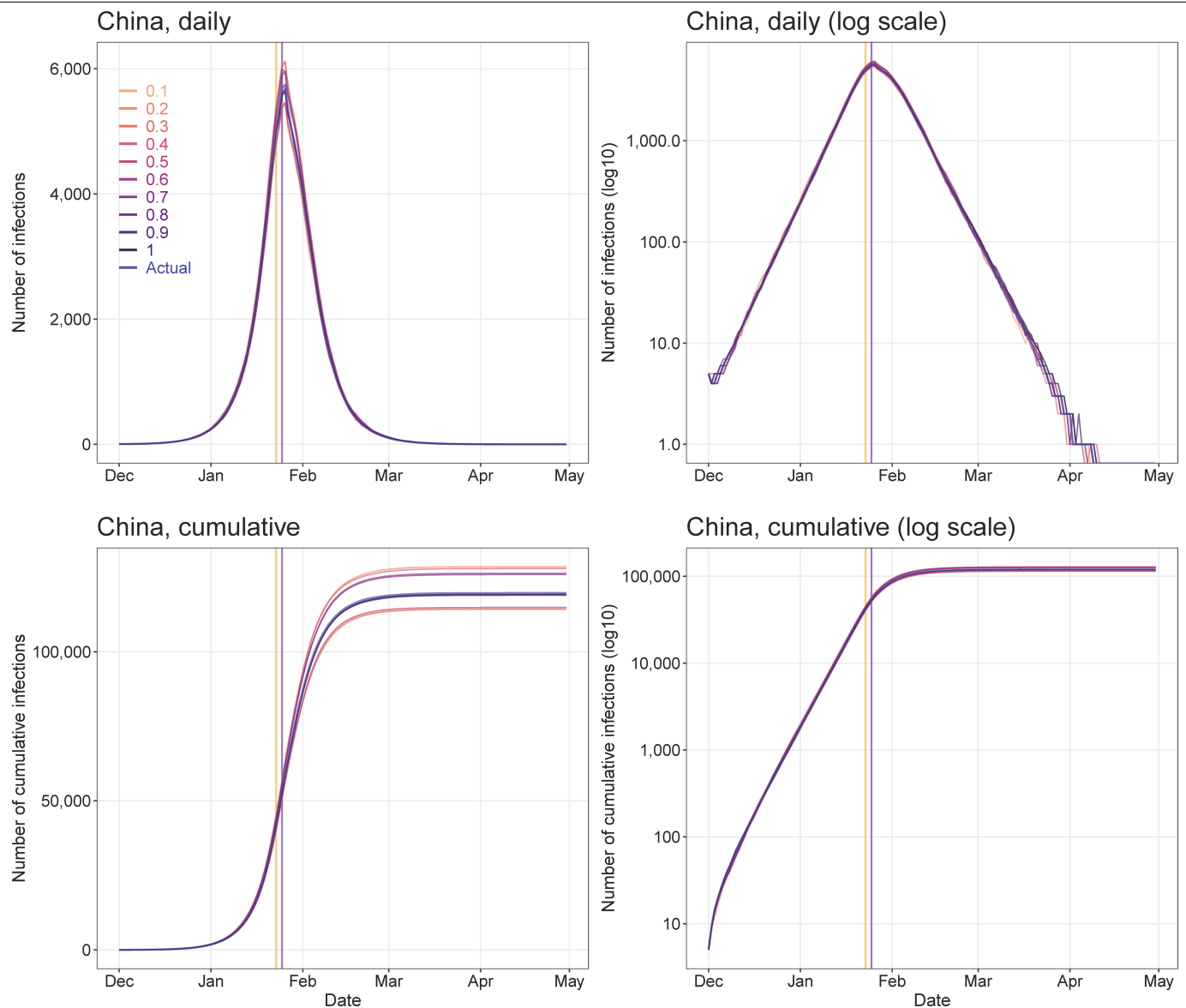
**Extended Data Fig. 4 | Sensitivity of estimates of COVID-19 epidemics for various levels of intercity travel restrictions from 23 January 2020.** All other parameters, NPIs and input data were the same as the baseline model with $R_0$ = 2.2. The actual percentages of intercity travel restrictions changed day by day across cities in China (0.1 indicates a 90% reduction from normal travel; 1 indicates no travel restrictions). Orange vertical lines indicate the date on which the lockdown of Wuhan began (23 January 2020); purple vertical lines indicate the date on which the Chinese New Year began (25 January 2020).

**Extended Data Fig. 5 | Sensitivity of estimates of COVID-19 epidemics for various numbers of days from illness onset to report or isolation.** All other parameters, NPIs and input data were the same as the baseline model with $R_0 = 2.2$. The actual delays of illness onset to report or isolation changed day by 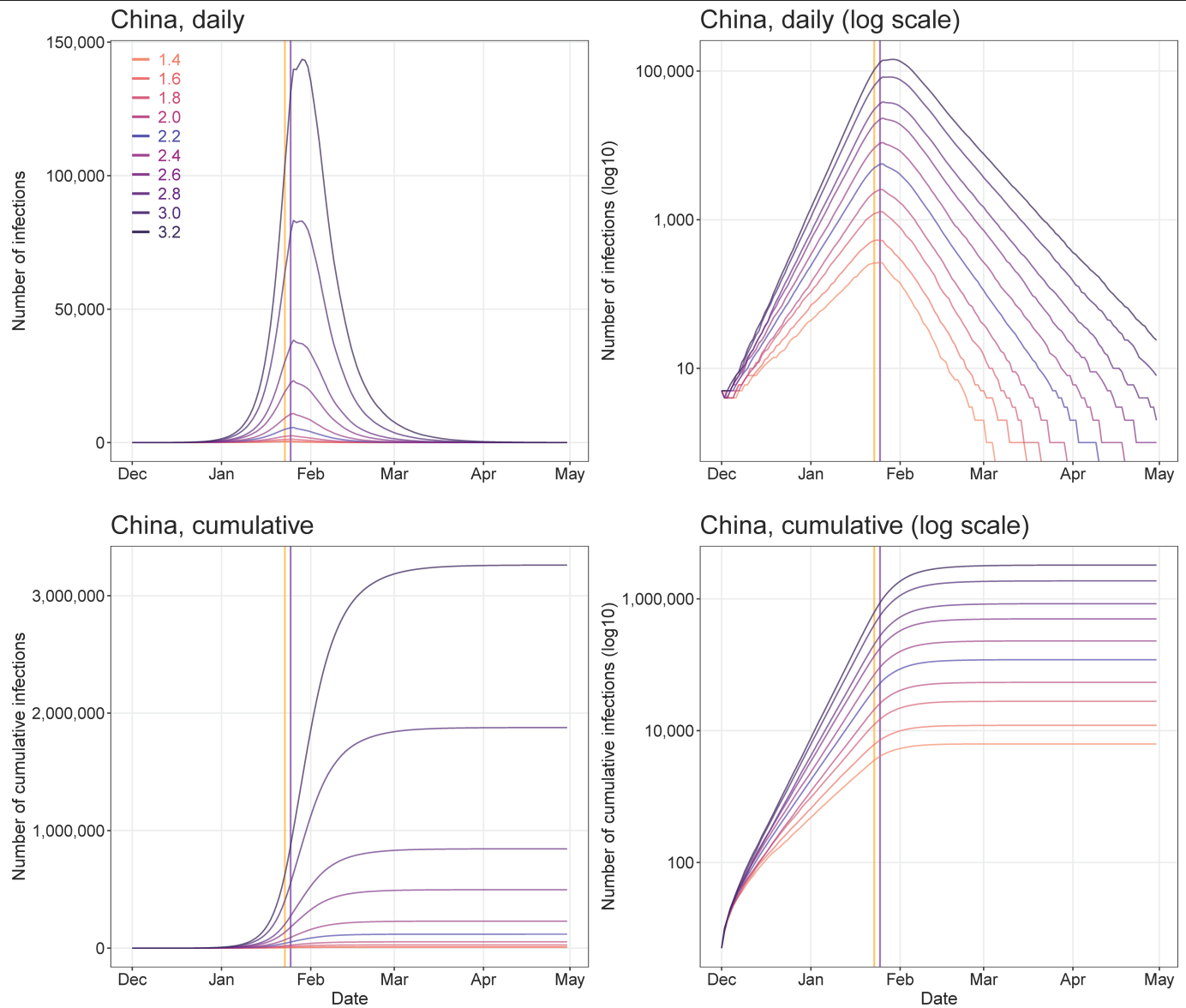day (Supplementary Table 2). Orange vertical lines indicate the date on which the lockdown of Wuhan began (23 January 2020); purple vertical lines indicate the date on which the Chinese New Year began (25 January 2020).
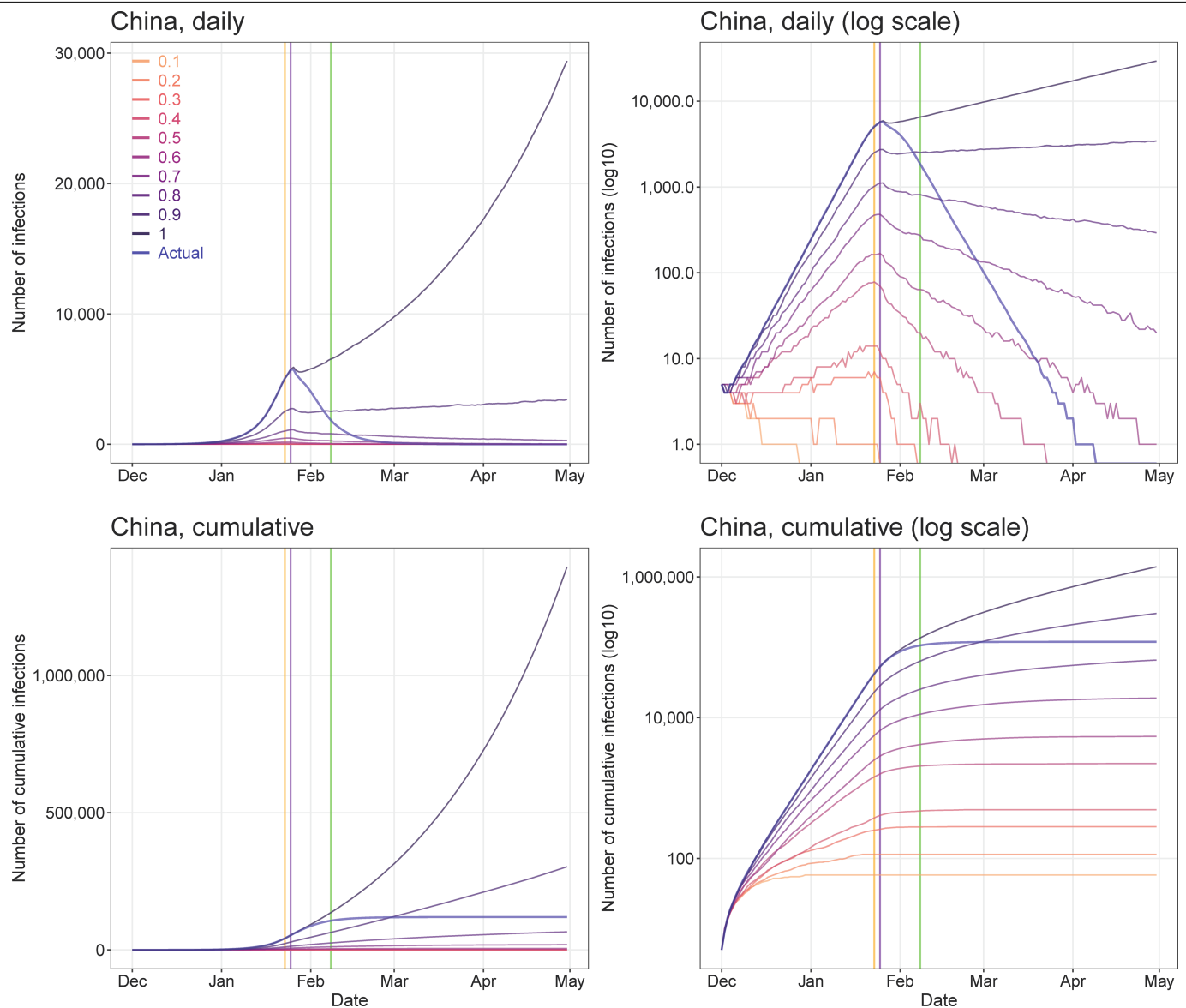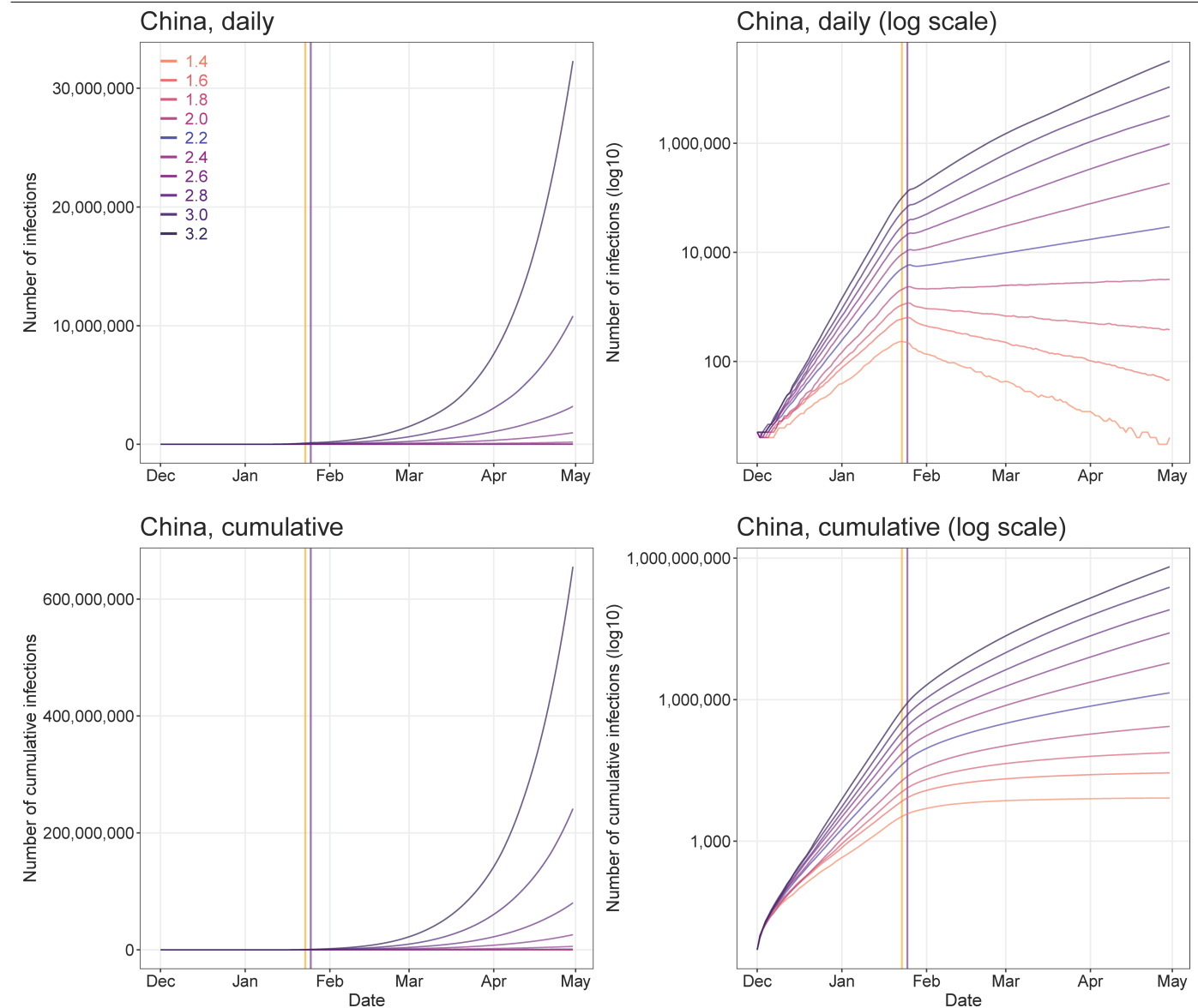
**Extended Data Fig. 6 | Sensitivity of estimates of COVID-19 epidemics for various rates of contact.** All other parameters, NPIs and input data were the same as the baseline model with $R_0 = 2.2$. The actual percentage of population contact (0.1 indicates 10% of usual contact, 1 means no contact restrictions) changed day by day across the country (Supplementary Table 1). Orange vertical lines indicate the date on which the lockdown of Wuhan began (23 January 2020); purple vertical lines indicate the date on which the Chinese New Year began (25 January 2020).

**Extended Data Fig. 7 | Sensitivity of estimates of COVID-19 epidemics for various values of $R_0$ and without intercity travel restrictions.** All other parameters, NPIs and input data were the same as the baseline model with $R_0 = 2.2$. Orange vertical lines indicate the date on which the lockdown of Wuhan began (23 January 2020); purple vertical lines indicate the date on which the Chinese New Year began (25 January 2020).

**Extended Data Fig. 8 | Sensitivity of estimates of COVID-19 epidemics for various values of $R_0$ and without contact restrictions within cities.** All other parameters, NPIs and input data were the same as the baseline model with $R_0 = 2.2$. Orange vertical lines indicate the date on which the lockdown of Wuhan began (23 January 2020); purple vertical lines indicate the date on which the Chinese New Year began (25 January 2020).
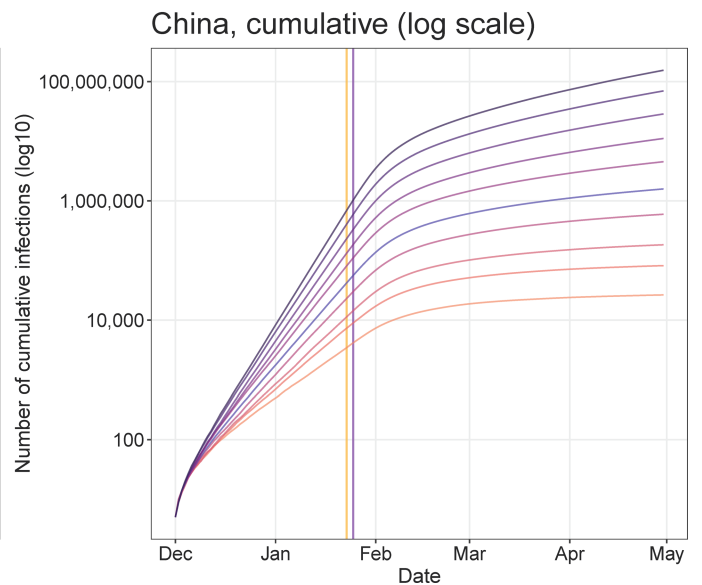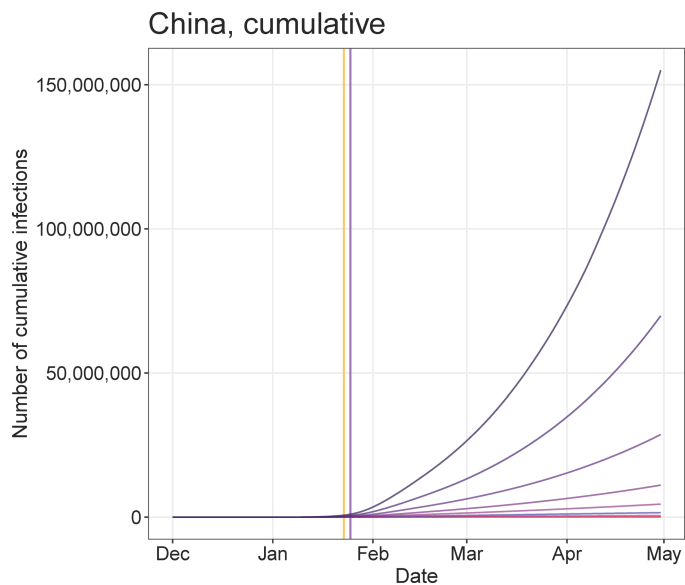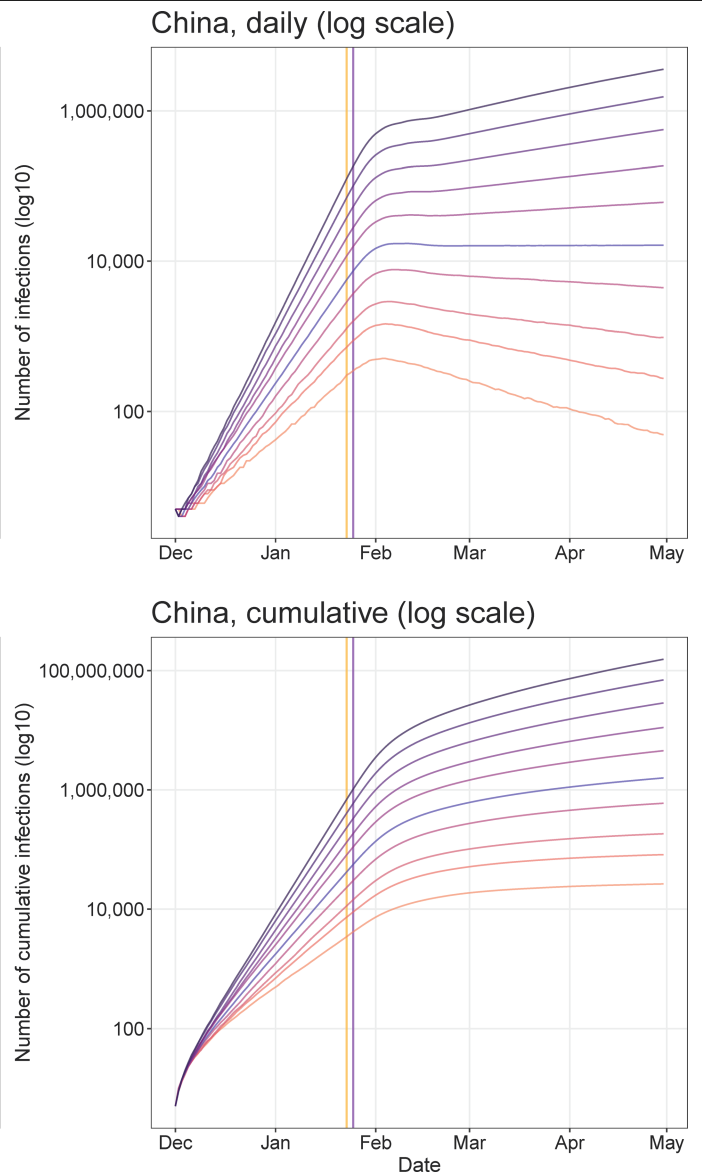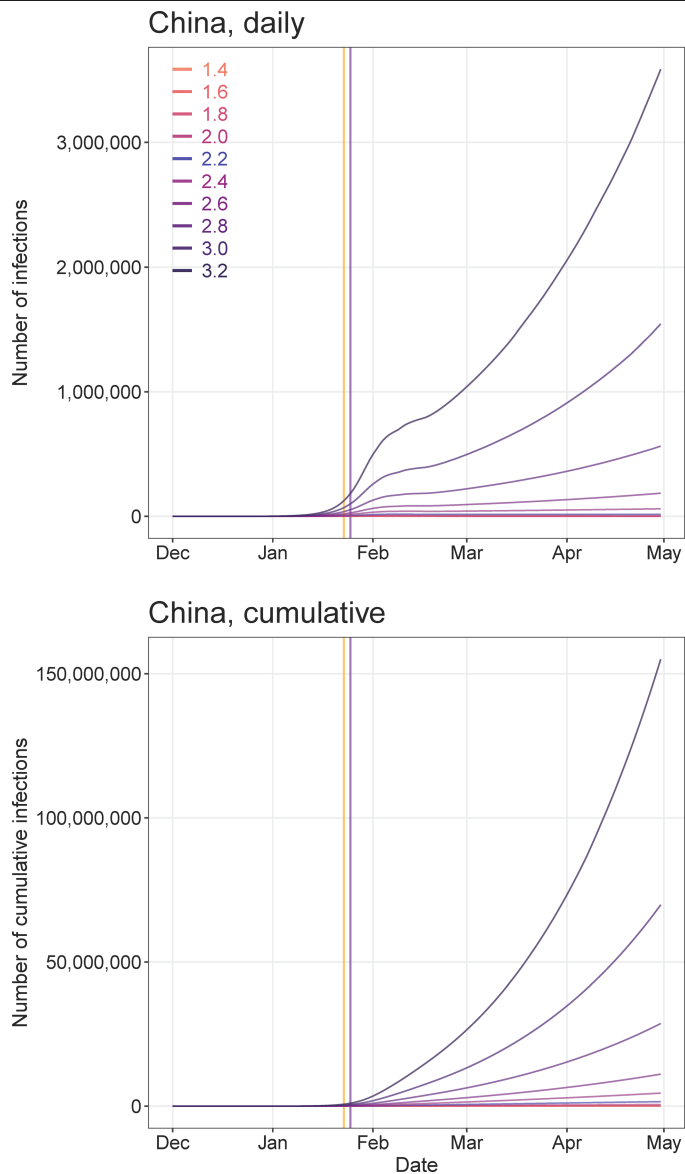
**China, daily**

1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0
3.2

**China, daily (log scale)**

**China, cumulative**

**China, cumulative (log scale)**

**Extended Data Fig. 9 | Sensitivity of estimates of COVID-19 epidemics for various values of $R_0$ and without improved timeliness of case detection and isolation.** The delay from illness onset to detection and isolation was set as a constant of 11 days, which was the average delay from 16–18 January 2020. All other parameters, NPIs and input data were the same as the baseline model with $R_0 = 2.2$. Orange vertical lines indicate the date on which the lockdown of Wuhan began (23 January 2020); purple vertical lines indicate the date on which the Chinese New Year began (25 January 2020).

**Extended Data Table 1 | Reports and estimates of COVID-19 cases in mainland China as of 29 February 2020**

| Interventions and timing | Wuhan City, Hubei Province | Other cities in Hubei Province | Other provinces | Mainland China |
|---|---|---|---|---|
| **Under current non-pharmaceutical interventions (NPIs)** | | | | |
| No. of cases reported (%)[a] | 49,122 (62) | 17,785 (22) | 12,917 (16) | 79,824 (100) |
| Estimates of cases (%) | 78,910 (69) | 18,503 (16) | 16,912 (15) | 114,325 (100) |
| Interquartile range | 51,952-111,280 | 11,029-28,685 | 9,499-27,033 | 76,776-164,576 |
| Dates of estimated peak | Jan 25-27 | Jan 24-26 | Jan 24-26 | Jan 25-27 |
| Interval between NPIs and epidemic peak[b] | 7 days | 6 days | 6 days | 7 days |
| **Percentage (%) of cases that could have been prevented with earlier interventions** | | | | |
| One week ahead | 61 (45-79) | 71 (55-86) | 78 (62-90) | 66 (50-82) |
| Two weeks ahead | 84 (78-89) | 90 (82-94) | 91 (84-95) | 86 (81-90) |
| Three weeks ahead | 94 (92-96) | 97 (95-99) | 98 (97-99) | 95 (93-97) |
| **Estimated relative no. of cases with later interventions [c]** | | | | |
| One week delay | 2.4 (1.6-3.5) | 3.1 (1.8-4.6) | 3.3 (2-5.4) | 2.6 (1.8-3.8) |
| Two weeks delay | 5.8 (4.0-8.6) | 8.6 (5.3-12.8) | 9.4 (6.1-14.6) | 6.7 (4.6-10.0) |
| Three weeks delay | 15.1 (9-21.1) | 22.6 (13.5-33.9) | 27.9 (17.5-42.8) | 17.6 (11.2-25.5) |
| **Estimated relative no. of cases under various NPIs [c]** | | | | |
| Without inter-city travel restriction | 1.0 (0.6-1.3) | 1.1 (0.7-1.7) | 1.1 (0.7-1.7) | 1.0 (0.6-1.4) |
| Without inner-city contact reduction | 2.5 (1.7-3.7) | 2.6 (1.5-4.2) | 2.4 (1.2-4.0) | 2.6 (1.7-3.7) |
| Without case early detection and isolation | 5.0 (3.3-6.9) | 5.6 (3.2-8.4) | 5.1 (2.5-8.4) | 5.0 (3.3-7.1) |
| Without all interventions above | 51.4 (33.2-71.2) | 91.6 (57.6-132.5) | 124.7 (77.4-180) | 67.3 (43.7-93.7) |

[a]The reported data on COVID-19 cases were obtained from the Chinese National Health Commission as of 29 February 2020.
[b]The timeliness of case identification and reporting improved from 19 January 2020 and the travel restrictions and social distancing were implemented from 23 January 2020. We compared the peak dates by region with 19 January 2020 to define the interval from NPIs to epidemic peak.
[c]Referring to the median of estimates under actual interventions and timing.
The median and IQR of estimates are shown.

# nature research

| | |
|---|---|
| Corresponding author(s): | Shengjie Lai |
| Last updated by author(s): | Apr 17, 2020 |

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size ($n$) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. $F$, $t$, $r$) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ | Estimates of effect sizes (e.g. Cohen's $d$, Pearson's $r$), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | R version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria) was used to perform data collation and analyses. |
| Data analysis | R version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria) was used to perform data collation and analyses. The model built by this study has been made openly available for further use at https://github.com/wpgp/BEARmod. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data of COVID-19 cases reported by county, city, and province across China are available from data sources detailed in the Supplementary, and the average days from illness onset to report of the first case by each county used in the modelling are detailed in Supplementary Table 2. The mobile phone datasets analysed during the current study are not publicly available since this would compromise the agreement with the data provider, but the information on the process of requesting access to the data that support the findings of this study are available from Dr Shengjie Lai (Shengjie.Lai@soton.ac.uk), and the data of travel and contact reductions derived from the datasets and used in our model are detailed in Supplementary Table 1.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Study description | Quantitative observational and modelling study |
| Research sample | COVID-19 cases reported across mainland China as of February 29, 2020 were included in this study. As public awareness and enhanced case searching remained high throughout the study period, a high proportion of cases with symptoms was likely to have been detected, with nearly all reported cases eventually subjected to laboratory testing. However, the reported data of COVID-19 cases might not include asymptomatic and mild infections, and our model may have underestimated the total number of infections. The data on COVID-19 cases reported by county, city, and province across China are available from the data sources listed in the Supplementary Information File 3. This study also used population movement data across the country in 2020 and previous years, obtained from Baidu Location-based service. However, coverage biases of smartphone and Baidu users in population likely exist. Though a high percentage of the population owns smartphones in China, the mobile user group still does not cover specific subgroups of the population, particularly children. Therefore, our population movement data may provide an incomplete picture, and differences between the characteristics of smartphone owners and non-owners may also bias estimates in this study. Additionally, the magnitude and patterns of movements could change year by year. |
| Sampling strategy | This study included the numbers of all COVID-19 cases reported across mainland China as of February 29, 2020. Population movement data on human mobility of all Baidu users across the country were obtained from Baidu location-based service in 2014-2015 and 2020. |
| Data collection | We collated data of the first case reported by county across mainland China to measure the delay from illness to case report as a reference of the improved timeliness of case identification, isolation and reporting during the outbreak (Supplementary Information File 1). The daily number of COVID-19 cases by date of illness onset in Wuhan City, Hubei Province and other provinces as of February 13, 2020 were used to further validate the epicurves estimated in this study across time. The number of cases reported by city across mainland China as of February 29 were used to define the predictability of our model across space. These case data were collated from the websites of national and local health authorities, news media, and publications (Supplementary Information File 3). The epidemiological parameters estimated for the early stage of the outbreak in Wuhan from previous study (reference #5) were initially collected used to parameterise the epidemic before widely implementing interventions. Three population movement datasets, obtained from Baidu location-based services, were used in this study: 1) daily relative outbound and inbound flow of mobile phone users for each prefecture-level city (340 cities in mainland China) in 2020; 2) historical relative movement matrix with daily total number of users at city level from December 26, 2014 to May 26, 2015, aligning with the 2020 Chinese new year holiday period; 3) daily population movements at county level (2862 counties in China) from January 26 through April 30, 2014, aligning with the 2015 and 2020 Chinese new year holiday period. |
| Timing | COVID-19 cases: December 2, 2019 - February 29, 2020. Three Baidu population movement datesets: 1) January 26, 2014- April 30, 2014; 2) December 26, 2014 - May 26, 2015; 3) January 1, 2020 - April 13, 2020. |
| Data exclusions | Before conducting this study, we already noticed that there was an abnormal increase of cases in Wuhan City and Hubei Province on February 1, 2020, based on the date of illness onset. The case definition has been adjusted and a large number of clinically diagnosed cases before laboratory confirmation have been retrospectively reported into the information system since 12 February. However, the spike on February 1 might not represent the actual infection patterns. We have discussed this issue and underlying causes (e.g. changes of definitions, reporting delay, system error, incorrect reporting of the onset date) with epidemiologists in China, but exact reasons remain unclear. Therefore, before comparing reported data with estimates in our study, we interpolated the number on February 1 by using the mean of the numbers of cases reported on January 31 and February 2 in the epicurves of Wuhan and Hubei Province. |
| Non-participation | As this study collected and used secondary data from disease surveillance and Baidu location-based service, we did not access to the raw data and we don't know how many participants dropped out/declined participation. However, the number of COVID-19 cases might be unreported as asymptomatic and mild infections exist. The mobile user group does not cover specific subgroups of the population, particularly children, and not all mobile owners use the Baidu location-based service. Therefore, our population movement data may provide an incomplete picture of movement of all population in China, and the spatiotemporal and demographic variations in the behaviour of phone users could have biased population distribution and travel estimates. |
| Randomization | We did not randomly sample COVID-19 cases and movement data from population. In our SEIR modelling framework, we conducted 1000 simulations to account for the uncertainty of estimates. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|-----|----------------------|
| ☒ | Antibodies |
| ☒ | Eukaryotic cell lines |
| ☒ | Palaeontology |
| ☒ | Animals and other organisms |
| ☒ | Human research participants |
| ☒ | Clinical data |

## Methods

| n/a | Involved in the study |
|-----|----------------------|
| ☒ | ChIP-seq |
| ☒ | Flow cytometry |
| ☒ | MRI-based neuroimaging |