The results in this note were based on information as of 8:30 am on 21 January 2020.

There are 3 cases confirmed outside mainland China on 21 January 2020. As of 12:00 pm on 22 January 2020, there are 7 cases confirmed outside mainland China, including 2 in Thailand, 1 in Japan, 1 in South Korea, 1 in Taiwan, 1 in United States and 1 in Macau. We are updating the results by incorporating the new information.

Nowcasting and forecasting the Wuhan 2019-nCoV outbreak

Kathy Leung[1], Joseph T. Wu[1], Gabriel M Leung[1]

[1]WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China

In this note, the meaning of "cases" differs by context as follows:

Case definition by Chinese Center for Disease Control and Prevention (China CDC)

Definition of **suspected** cases*:

- Fever ≥ 38°C
- Radiographic findings of pneumonia or acute respiratory distress syndrome
- Normal or reduced white blood cell counts or reduced lymphocyte subset counts
- No improvement or deteriorate after 3-5 days of antibiotics treatment

Definition of **probable** cases:

- Suspected case
- Epidemiologic link or history

Definition of **confirmed** cases:

1st case in the province:

- Probable case
- Detection of virus nucleic acid at the City, Provincial and National CDC

2nd case or after in the province:

- Probable case
- Detection of virus nucleic acid at the City and Provincial CDC

Case definition used in the modelling study

Symptomatic cases who could be detected by temperature screening at international borders and/or with disease severity of a level requiring hospitalization plus travel history to Wuhan.

*"Definition of suspected cases of unexplained pneumonia" were from National Health Commission of the People's Republic of China on http://www.nhc.gov.cn/.

Summary

Background

Since December 31, 2019, the Chinese city of Wuhan has reported an outbreak of atypical pneumonia likely caused by the 2019 novel coronavirus (2019-nCoV). Cases have been exported to other Chinese cities as well as internationally.

Methods

Using the number of cases exported from Wuhan to other countries (see case definitions on the title page), we first inferred the number of cases that had occurred in Wuhan during 1-17 January 2020. We then estimated the number of cases that had been exported to other Chinese cities during the same period. Next, we forecasted the national and international spread of 2019-nCoV by air, train and road traffic during one week immediately before and after the first day of Spring Festival or Lunar New Year (January 25), accounting for the effect of wet market interventions and temperature screening of air and train passengers that had been recently implemented in Wuhan (exit port) and elsewhere (entry ports). Finally, we presented a statistical tool for predicting in real time the final outbreak cluster size in the event of a super-spreading event (SSE) and demonstrated its performance in the 2015 nosocomial outbreaks of MERS in Seoul and the 2003 Amoy Gardens community outbreak of SARS in Hong Kong.

Findings

We estimated that 1,343 (547-3,446) cases had had onset of symptoms in Wuhan during 1-17 January 2020. We estimated that Beijing, Shanghai, Chongqing, Guangzhou and Shenzhen had imported 17 (6-46), 15 (5-41), 15 (5-40), 14 (4-38) and 10 (3-29) cases from Wuhan during this period, respectively. If wet market interventions had reduced the force of infection by 75% and temperature screening could detect 50% of symptomatic passengers, the expected total number of international case exportation during the two weeks straddling the first day of Spring Festival would be 0.81 (0-3). However, cases would likely continue to be exported from Wuhan to Beijing, Shanghai, Chongqing, Guangzhou and Shenzhen during this period. If SSEs were observed in the nosocomial or community settings (as seen in previous outbreaks of SARS and MERS), our statistical tool can accurately and precisely predict the final outbreak size within 2-5 days after three or more cases of the cluster has been observed.

Conclusions

Situational awareness by nowcasting with real time updates is critical to public health control of 2019-nCoV, as for other newly emerging pathogens. Our forecasts, based on actual travel patterns by different

modes of transport, are dependent on some key assumptions including the absence of superspreading events, containment of a single epicentre at Wuhan and robust surveillance and satisfactory adherence to public health control measures everywhere. Further delineation of actual epidemiological parameters for 2019-nCoV would improve model performance.

Introduction

Wuhan, provincial capital of Hubei in China, is investigating an outbreak of atypical pneumonia caused by the zoonotic 2019 novel coronavirus (2019-nCoV). As of 21 January 2020 (8:30 am), there have been 219 cases of 2019-nCoV infections confirmed in mainland China, including 14 cases in Guangdong, 5 cases in Beijing and 2 in Shanghai. Four exported cases to other countries have been confirmed, including 2 in Thailand, 1 in Japan and 1 in South Korea. While details remain scanty in the public domain, Figures 1A summarises a chronology of events and 1B provides a rough epidemic curve of cumulative incidence by date of announcement. The title page of this note gives the case definition provided by National Health Commission of the People's Republic of China.

Two other novel coronaviruses (CoV) have emerged as major global epidemics since 2002, namely SARS-CoV (2002) and MERS-CoV (2012), respectively spreading to 37 and 27 countries, causing 8,000+ and 2,494 cases and almost 800 and 858 deaths worldwide to date. Both are zoonoses and epidemiologically similar, except that SARS-CoV has virtually no subclinical cases whereas MERS-CoV behaves more like the other four commonly circulating human CoVs with a substantial proportion of asymptomatic infections (Table 1). Both usually present with moderate to severe respiratory symptoms which often progress to severe pneumonia.

A notable common characteristic of both SARS-CoV and MERS-CoV is that they have limited potential for sustained community transmission (i.e. low basic reproductive number) [1-3] . However, the most worrisome aspect is the potential for large case clusters apparently seeded by a single index which can exceed 100 in size, caused by "superspreading" [2-6]. Such superspreading events (SSEs) had mostly involved nosocomial clusters in healthcare facilities as observed in Hong Kong, Taipei, Singapore and Toronto for SARS and in Jeddah and Seoul for MERS [7, 8]. Nevertheless, the Amoy Gardens example during Hong Kong's 2003 SARS outbreak recorded 321 cases in a single SSE community cluster.

Using the number of cases exported from Wuhan to other countries, we first inferred the number of cases that had occurred in Wuhan during 1-17 January 2020. We then estimated the number of cases that had been exported to other Chinese cities during the same period. We then forecasted the domestic and international spread of 2019-nCoV by air, train and road traffic during one week before and one week after the first day of Spring Festival, accounting for the effect of wet market interventions and temperature screening that had been recently implemented in Wuhan. Finally, we presented a statistical tool for predicting in real time the final outbreak cluster size in the event of a super-spreading event (SSE) and demonstrated its performance in the 2015 nosocomial outbreaks of MERS in Seoul and the 2003 Amoy Gardens community outbreak of SARS in Hong Kong.

Methods

Data

We obtained the population size of Wuhan (including migrant workers) from the National Bureau of Statistics. To date, while it is generally acknowledged that there is human-to-human transmission, its extent has yet to be quantified. We assumed the incubation period of this novel coronavirus to be similar to that of SARS and MERS (mean = 6 days; coefficient of variation = 0.67). We estimated the daily number of outbound travellers from Wuhan by air, train and road with data from three sources (See Supplementary Information for details):

1) The monthly number of domestic and international flight bookings from Wuhan in January to February 2019 obtained from Official Airline Group (OAG).
2) The daily number of domestic passengers by means of transportation (i.e. air, train and road) recorded by Tencent's LBS (location-based services) database from Wuhan to 334 prefecture-level cities in mainland China from 6 January to 7 March 2019 (https://heat.qq.com/).
3) The domestic passenger volumes from and to Wuhan during *chungyun* 2020 estimated by Wuhan Municipal Transportation Management Bureau in the press releases in December 2019 [9].

Estimating the outbreak size in Wuhan and the number of cases that had been exported to other cities in China

We used the number of international case exportations reported up to January 20, 2020 (2 to Thailand, 1 to Japan and 1 to South Korea) to infer the total number of cases in Wuhan up to January 17, 2020 (the onset date of the most recent international case exportation). Our methodology was essentially the same as that reported by Imai et al (https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/news--wuhan-coronavirus/). In this calculation, we assumed that the catchment population size of the Tianhe International Airport at Wuhan was 11.08 million and cases could travel during their incubation period and within 4 days after symptoms onset. Cases in this calculation corresponded to 2019-nCoV infections in Wuhan that were clinically comparable to the cases that had been exported internationally, i.e. infections with very mild symptoms would not be included. Using the resulting estimated outbreak size in Wuhan, we then estimated the number of cases that had been exported domestically to other Chinese cities by air, train and road travel.

Forecasting international and domestic case exports

We assumed that human-to-human transmission would not be self-sustaining (i.e. $0 \leq R_0 < 1$) and hence the outbreak size would scale linearly with the number of zoonotic infections. Wuhan health officials have closed and disinfected local wet markets several times, as well as stepped up public hygiene campaigns and instituted enhanced surveillance and infection control procedures at health care facilities since 1 January 2020 (http://wjw.wuhan.gov.cn/front/web/showDetail/2020010509020); and implemented temperature screening at the international airport and train stations since 15 January. Domestic and international destination ports have variously adopted similar entry temperature screening measures. We projected the impact of these interventions on the number of international and domestic case exportation during one week before and one week after the first day of Spring Festival under various assumptions regarding their effectiveness. In the base case, we assumed that the wet market interventions and public health control measures had reduced the force of infection in Wuhan by 75% and temperature screening at exit +/- entry could detect 50% of symptomatic travellers (i.e. screening sensitivity = 50%). To account for the uncertainty regarding intervention effectiveness, we also considered 60% and 90% reduction in the force of infection in Wuhan and temperature screening sensitivity of 25% and 75%. Note that temperature screening would not be able to detect infected individuals before they developed symptoms, and hence its effectiveness might be limited given that the mean incubation period for typical CoV was 6 days.

SSE cluster size estimation

SSEs had been repeatedly observed in SARS and MERS outbreaks, but little is known regarding their cause and probability of occurrence [10]. Although the probability of SSEs with more than 20 secondary cases per index case was estimated to be on the order of 0.001-0.01 for MERS [11], the 2015 MERS outbreak in Seoul comprised three SSE clusters of size 27, 29 and 125, leading to widespread public anxiety in Seoul. Given that the aetiological pathogen for the Wuhan outbreak was likely a coronavirus and potential public misinterpretation of the explosive epidemic growth associated with SSEs as intrinsically high transmissibility, it would be prudent not to underestimate the threat of SSE when there was large uncertainty in the transmissibility and superspreading potential of this novel coronavirus. To help mitigate the risk of SSEs, we proposed a Bayesian statistical framework for estimating SSE cluster size in real-time and retrospectively demonstrated its performance in the nosocomial outbreaks of MERS in Seoul in 2015 and the Amoy Gardens outbreak of SARS cases in Hong Kong in 2003.

Say that a cluster alert would be issued when three or more cases were confirmed to be epidemiologically linked to the same index case. Cluster size estimation would begin after a cluster alert has been issued. We assumed that a cluster size estimate was robust if (i) the width of the 95% credible interval (CrI) was smaller than 10 or the posterior mean (i.e. sufficiently precise); and (ii) its 95% CrI covered the true value (i.e. sufficiently accurate).

We formulated a Bayesian framework for estimating the size of an SSE cluster based on the back-calculation techniques used for estimating the size of non-transmissible acute infectious disease outbreaks, bioterrorist anthrax attacks and infectious diseases with long incubation period (e.g. HIV, BSE, vCJD) [12]. In brief, we provide only a simplified formulation here (see Supplementary Information for the full details).

Suppose all n cases of an SSE cluster were infected by the index case at time 0 and k cases had been observed by time t. Let X and Y denote the incubation period and delay from symptoms onset to case observation for a given case. The likelihood function was

$$L(n,\theta) = \binom{n}{k}\left(1 - Prob(X + Y > t|\theta)\right)^{n-k} \prod_{i=1}^{k} Prob(X = x_i, Y = y_i|\theta)$$

where x_i and y_i were the observed values of X and Y for the ith case, and θ comprised the parameters for the joint probability distribution of X and Y. We estimated n and θ using Markov Chain Monte Carlo (MCMC) methods with Gibbs sampling and non-informative flat prior for n. For the MERS case study, we assumed that independent data from MERS cases in the Middle East were available for informing the prior distribution of θ. For the Amoy Garden SARS case study, we adopted non-informative flat priors for θ but augmented the likelihood with data from the non-Amoy-Gardens cases in Hong Kong that had been observed up to time t.

Findings

Estimating the outbreak size in Wuhan and the number of cases that had been exported to other domestic cities

Using the posterior mode (the value with the highest probability in the posterior distribution) as our point estimate, we estimated that 1,343 (95% credible interval: 547-3,446) cases had had onset of symptoms in Wuhan by 17 January 2020; the point estimate would be higher at 1,678 if the posterior mean was used instead because the posterior distribution was right-skewed. Figure 3 shows that the number of cases that had been exported to cities to which Wuhan had the highest outbound travel volumes during 1-17 January 2020. We estimated that Beijing, Shanghai, Chongqing, Guangzhou and Shenzhen had imported 17 (6-46), 15 (5-41), 15 (5-40), 14 (4-38) and 10 (3-29) cases from Wuhan, respectively (the corresponding posterior means were 22.0, 19.5, 18.6, 17.8 and 13.5).

Forecasting international and domestic case exports

Figure 4 shows the probability distribution of the number of international case exportation that would be expected to occur during one week before and one week after the first day of Spring Festival. See Table 2 for the cities which are at the highest risk of international case exportation from Wuhan. In the base case, the expected total number of international case exportation during this period would be 0.81 (0-3). If the force of infection was reduced by 60% and 90% instead of our base case value of 75%, then the expected number of international case exportation would be 1.29 (0-4) and 0.32 (0-2), respectively. If the sensitivity of temperature screening was 25% and 75% instead of our base value of 50%, then the expected number of international case exportation would be 0.91 (0-4) and 0.71 (0-3), respectively.

Figure 5 summarizes the number of domestic case exportation that would be expected to occur during the same period. Although the number of case exportation would be substantially reduced in the base case, case exportation would likely continue to occur in Beijing, Shanghai, Chongqing, Guangzhou and Shenzhen. Sporadic case exportation from Wuhan remained possible even when the reduction in the force of infection was increased to 90% or the sensitivity of temperature screening increased to 75%.

SSE cluster size estimation.

The 2015 MERS outbreak in Seoul included three nosocomial clusters which comprised 29, 125 and 27 cases, respectively [7]. Using the incubation period and onset-to-confirmation delay estimates from 39 MERS cases in the Middle East [3, 13-15] to inform the priors for the MERS outbreak in Seoul, the model generated robust cluster size estimates for the first, second and third cluster 5, 4 and 2 days after the respective cluster alerts had been issued and 7, 17 and 12 days before the last case of the respective cluster was observed (Figure 6). At those points in time, the observed cases in each cluster corresponded to around 41% (12/29), 45% (57/125) and 26% (7/27) of the true cluster size. The cluster size estimates became increasingly accurate and precise thereafter. The model generated robust size estimates for clusters 2 and 3 sooner because the data from cluster 1 lessened the impact of the differences in incubation period and onset-to-confirmation delay between the MERS cases in the Middle East and Seoul. This highlights the importance of priors on incubation period and onset-to-confirmation delay in cluster size estimation.

The Amoy Garden cluster comprised 321 cases and was the largest cluster in the 2003 SARS outbreak in Hong Kong [1]. The model generated robust cluster size estimates 3 days after the cluster alert had been issued and 19 days before the last case of the cluster was observed. At that point in time, only 29% of all cases (89/310) in the cluster had been observed. The cluster size estimate became increasingly accurate and precise thereafter (Figure 7).

Discussion

During the present period shortly after the first outbreak alert having been issued on December 31, 2019 and the identification of the likely causative agent announced on January 9, 2020 but with few clinical or epidemiologic details otherwise, public health authorities would find fulfilling the critical tasks of nowcasting and forecasting difficult [16, 17]. Here we have estimated the outbreak size thus far in Wuhan and the likely extent of disease spread to other cities domestically. However, even with exported infectors to destination locations, the likelihood of them successfully seeding a major outbreak, especially in places with a robust surveillance and public health control infrastructure that have already been placed on high alert, is likely modest subject to the uncertain extent of human-to-human transmission and the absence of SSEs.

The examples of SARS and MERS have highlighted the critical threat of SSEs in both nosocomial and community settings even when $R_0 < 1$ (Table 1). We therefore developed predictive analytics that could provide real time estimates of the eventual cluster size of SSEs for service planning and cross-validation by public health authorities (available for online access at https://www3.influenza.hk/wuhan/). This statistical tool only requires daily incidence data as input and can be easily deployed in the field setting. The general public often interpret SSEs as imminent widespread community transmission although sustained occurrence of such events is rare. As such, the capacity for accurately nowcasting and forecasting SSE cluster size could be useful for not only enhancing situational awareness by health officials but also alleviating public unease with regard to the uncertainty and significance of such events [18].

Reassuringly for the Wuhan outbreak so far, local, provincial and national authorities appear to have understood the epidemiologic importance of closing down the index wet market (and others as well as their supply chains), isolating patients in designated hospitals with proper infection control processes in place, keeping close contacts under medical surveillance for rapid quarantine and isolation potentially, requiring real time reporting of all suspicious clusters, and screening for symptoms among outbound travellers [19].

Our models rest on a series of assumptions, namely: 1) case definitions as per the title page which do not necessarily tally with the official working definitions or the ongoing refinement of such given changing circumstances and emerging evidence; 2) arbitrary reductions in the force of infection which is dependent on the effectiveness of control measures so far implemented; 3) arbitrary sensitivity of exit and entry screening in detecting symptomatic cases with fever; 4) Wuhan would remain the only epicentre; 5) no significant SSEs in Wuhan or elsewhere with confirmed cases; and 6) robust surveillance and high degree of alertness to detect potential cases everywhere.

The overriding epidemiologic priority currently to inform public health control would be to compile and release a line list of suspected/possible/probable/confirmed cases and close contacts that is updated daily and linked to clinical outcomes and laboratory test results. A robust line list is essential for the generation of accurate and precise epidemiologic parameters as inputs into transmission model for informing situational awareness and optimizing epidemic response [20].

Author contributions

JTW, GML and KL designed the experiments. KL collected data. JTW and KL analysed data. KL, JTW and GML interpreted the results and wrote the manuscript.

Declaration of interests

We declare no competing interests.

Acknowledgement

We thank Chi-Kin Lam and Miky Wong from School of Public Health, The University of Hong Kong for technical support. This research was supported by a commissioned grant from the Health and Medical Research Fund from the Government of the Hong Kong Special Administrative Region. The funding body had no role in study design, data collection and analysis, preparation of the manuscript, or the decision to publish.

Table 1. Epidemiologic characteristics of human coronaviruses (HCoVs)

Parameter	SARS-CoV	MERS-CoV	Commonly circulating HCoVs – 229E, NL63, OC43 and HKU1
Basic reproductive number R_0, mean (95% CI) for SARS-CoV and MERS-CoV or transmissibility in the case of HCoVs	**Average** Beijing [10]: 1.88 Hong Kong [21]*: 1.70 (0.44-2.29) Singapore [10]: 1.63 Singapore [21]*: 1.83 (0.47-2.47) Toronto [21]*: 0.86 (0.24-1.18) Worldwide [6]: 0.95 (0.67-1.23) **Early phase (including SSE):** Singapore [22]: 2.2-3.6 **Early phase (excluding SSE):** Hong Kong [23]: 2.7 (2.2-3.7) **Later phase (excluding SSE):** Hong Kong [23]: 0.14-1 **After generation 1 (excluding SSE)** Beijing [10]: 0.94	**Average** Middle East [11]: 0.47 (0.29-0.80) Saudi Arabia [24]: 0.45 (0.33-0.58) Middle East and South Korea [6]: 0.91 (0.36-1.44) **Early phase (including SSE):** South Korea [25]: 2.0-8.1	229E and OC43 and United States [26]: Annual infection rates ranged from 2.8% to 26% in prospective cohorts. Guangzhou in China [27]: Detected in 2.25% adults and children with fever and upper respiratory infection symptoms, among which 60% were OC43, 17% were 229E, 15% were NL63 and 7.8% were HKU1. United Kingdom [28]: Coronaviruses were detected in all age groups, most frequently (4.86%) in the 7- to 12-month age category.
Incubation period (days), mean (SD) or mean (95% CI)	Hong Kong [1]: 4.6 (3.8-5.8) Hong Kong [29]: 4.4 (4.6) Beijing [29]: 5.7 (9.7) Taiwan [29]: 6.9 (6.1)	Saudi Arabia [15]: 5.0 (4.0-6.6) South Korea [15]: 6.9 (6.3-7.5)	OC43 and other common HCoVs [30]: 2-4 Common HCoVs [31]: 2-5 Common HCoVs [32]: 3-4
Serial interval (days), mean (SD)	Singapore [22]: 8.4 (3.8)	Saudi Arabia [24]: 6.8 (4.1) South Korea [7]: 12.4 (2.8)	--
Seroprevalence amongst non-cases	Hong Kong, amongst close contacts [33]: ~ 0%	Qatar [34]: 0.21% (10/4719) among healthy blood donors; 0.74% (1/135) among close contacts of cases but not sick Arabian Peninsula [35]: 0.15% (15/10365) among general population 6.2% (68/1090) among individuals exposed to camels	OC43 and 229E [36]: 86-100% HKU1, S-protein based ELISA [37]: 0% in <10 years old to a plateau of 21.6% in 31-40 years old
Case-hospitalization probability, mean (95% CI)	~ 100%	South Korea [8]: ~100%	OC43 in Canada [38]****: 12.6% 229E and OC43 and United States [26]: Prevalence ranged from 3.3% to 11.1% in the hospitalized cohort.

			Brazil [39]: 11% among children younger than three years attending the pediatric emergency room with ALRI and hospitalized
Case-fatality proportion	Worldwide WHO**: 9.6% Mainland China [40]**: 6.4% Hong Kong [1]***: 17%	Worldwide WHO***: 34.5% South Korea [8]***: 20.4%	--

* Mean (IQR)
** Among probable cases
*** Among lab-confirmed cases
**** Among elderly and disabled adults in a long-term care facility

City	Monthly no. of air passengers
Bangkok	16202
Hong Kong*	7531
Seoul	5982
Singapore	5661
Tokyo	5269
Taipei	5261
Kota Kinabalu	4531
Phuket	4411
Macau	3731
Ho Chi Minh City	3256
Kaohsiung	2718
Osaka	2636
Sydney	2504
Denpasar-Bali	2432
Phnom Penh	2000
London	1924
Kuala Lumpur	1902
Melbourne	1898
Chiang Mai	1816
Dubai	1799

*Due to the ongoing social unrest since June 2019, we used actual flight volume based on local estimates in the models

Table 2. Cities to which Wuhan has the highest international outbound air travel volume in 2019 based on the data from OAG.

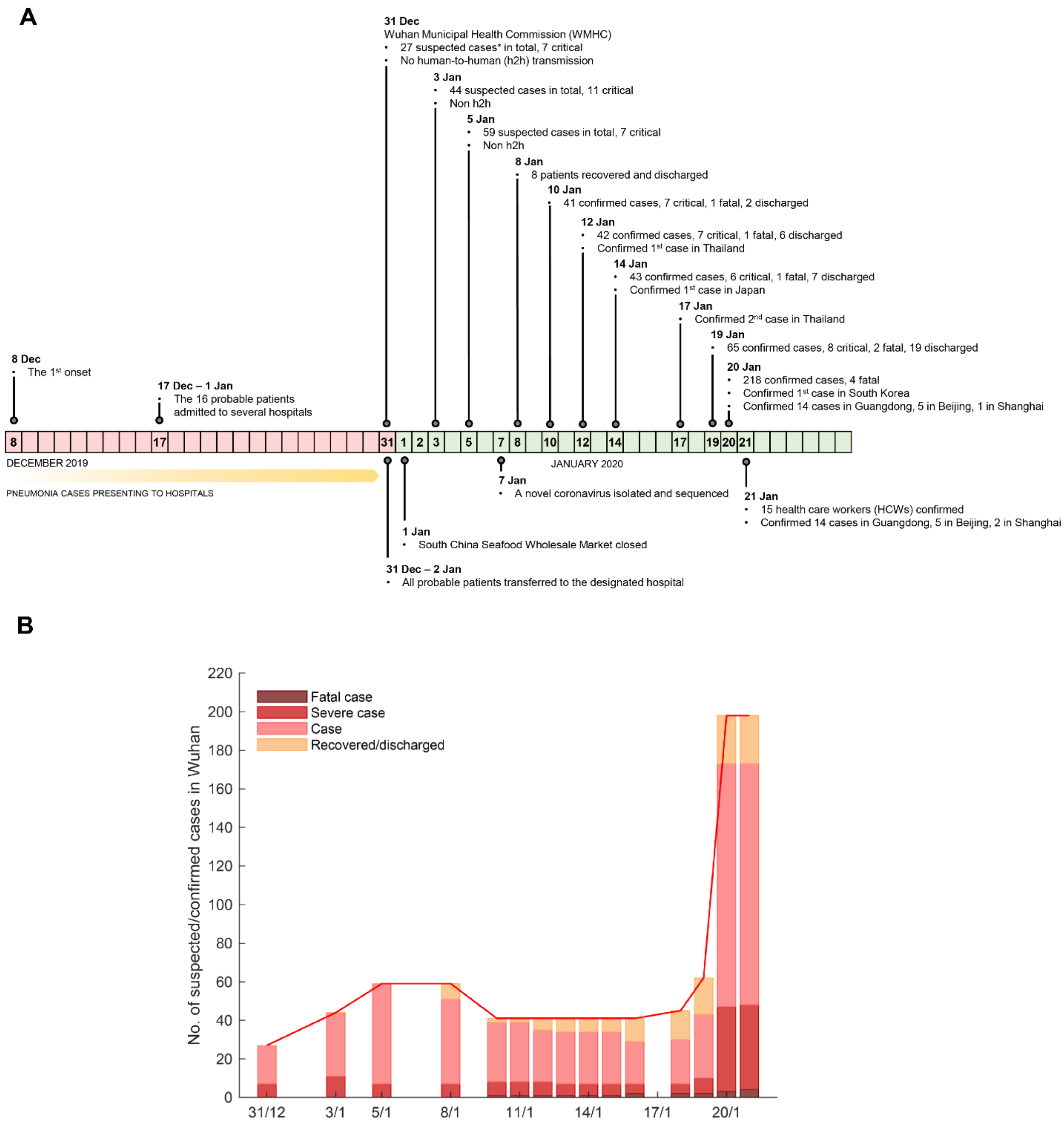

Figure 1. Timeline of Wuhan pneumonia outbreak.

Source of information:

1. “Wuhan Municipal Health Commission's briefing on the current pneumonia epidemic in the city” by Wuhan Municipal Health Commission on http://wjw.wuhan.gov.cn/
2. Information of the 16 “probable” cases were from “Wuhan "viral pneumonia" patients mostly work in the South China Seafood Market” on Wenweipo (http://news.wenweipo.com/2020/01/05/IN2001050012.htm) and “Unexplained pneumonia event in the Wuhan South China Seafood Market” on People.cn (http://yuqing.people.com.cn/n1/2020/0102/c429781-31531561.html).
3. “Definition of suspected cases of unexplained pneumonia” were from National Health Commission of the People’s Republic of China on http://www.nhc.gov.cn/

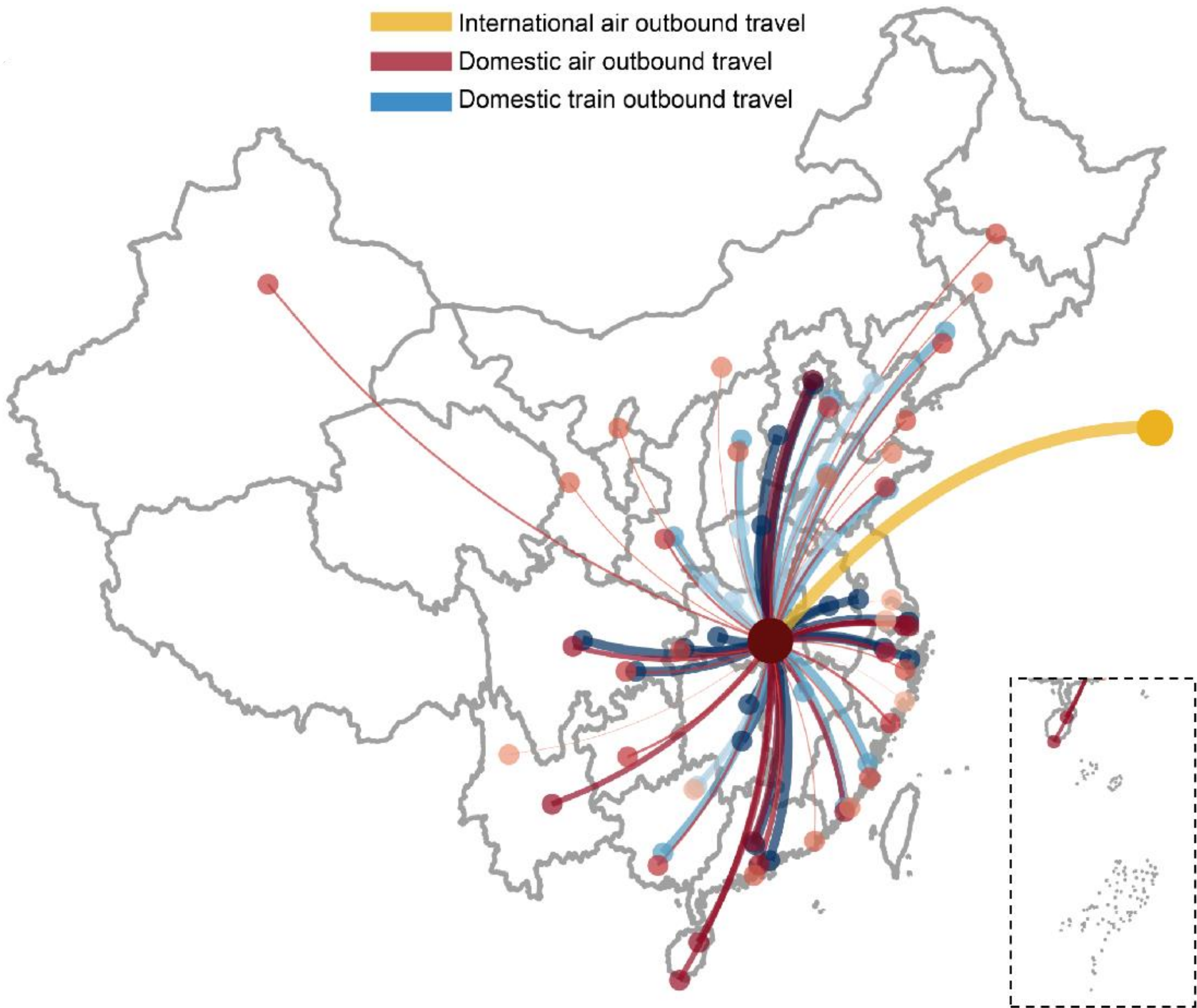

Figure 2. Risk of spread outside Wuhan. Major routes of outbound air and train travel originating from Wuhan during *chunyun* 2019. Wuhan is a hub for both air and train transportation in central China. Darker and thicker edges represent higher volumes. International outbound air travel (yellow) and the top 40 domestic (red) outbound air routes constituted 13.5% and 81.3% of all outbound air volume.

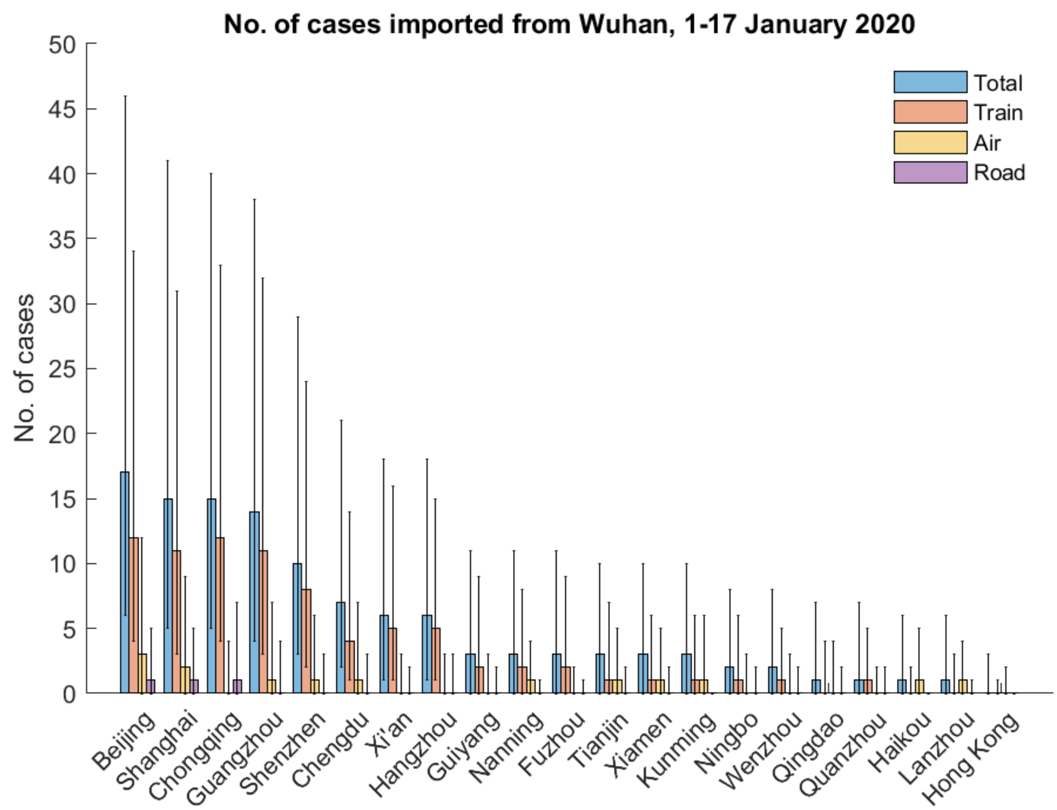

Figure 3. Estimated number of cases that had been exported to cities to which Wuhan had the highest outbound travel volumes during 1-17 January 2020. Bars correspond to posterior modes and vertical lines indicate the 95% credible intervals.

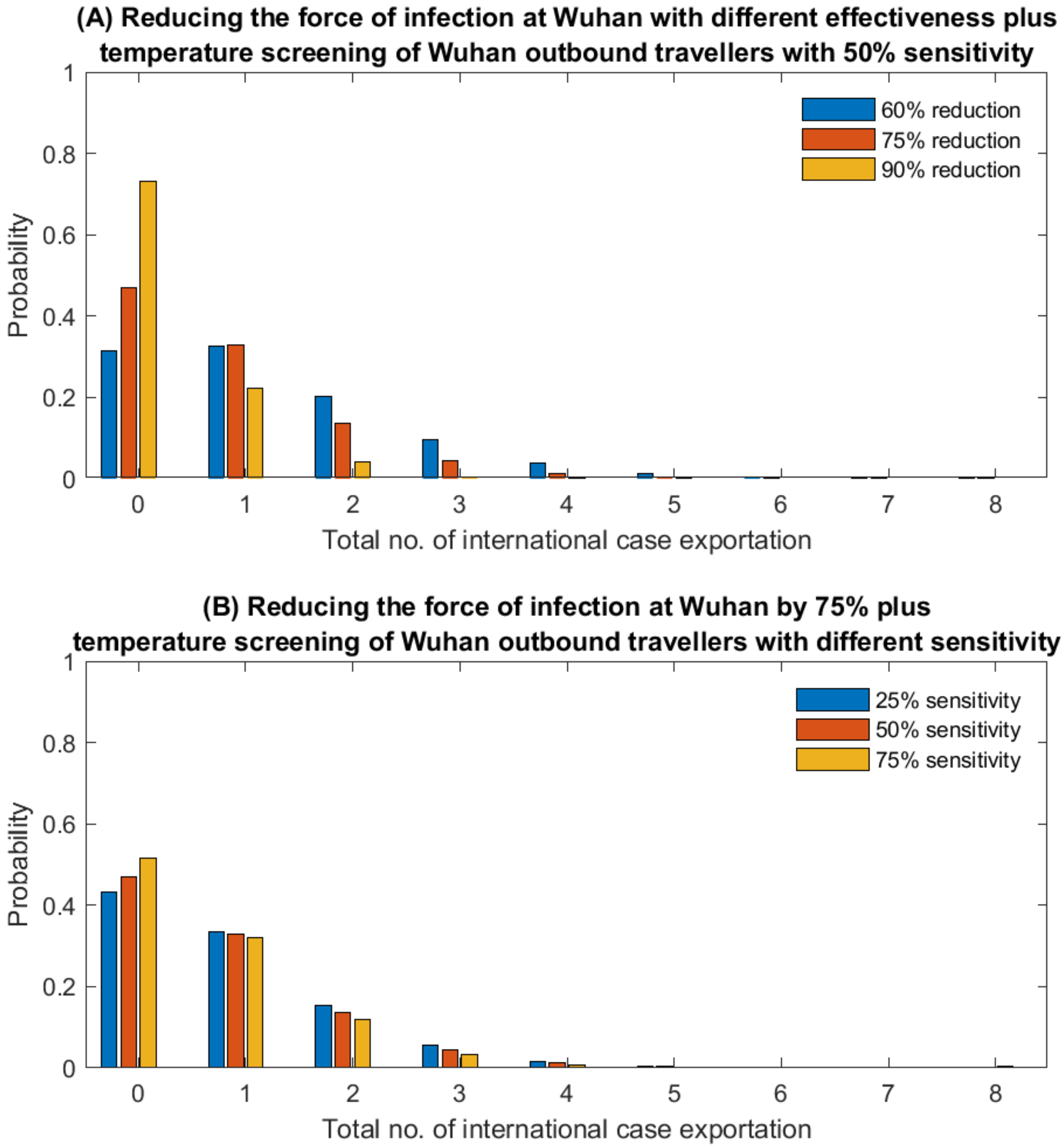

Figure 4. Probability distribution of the total number of international case exportation during one week immediately before and one week after the first day of Spring Festival. See Table 2 for the cities which are at the highest risk of international case exportation from Wuhan.

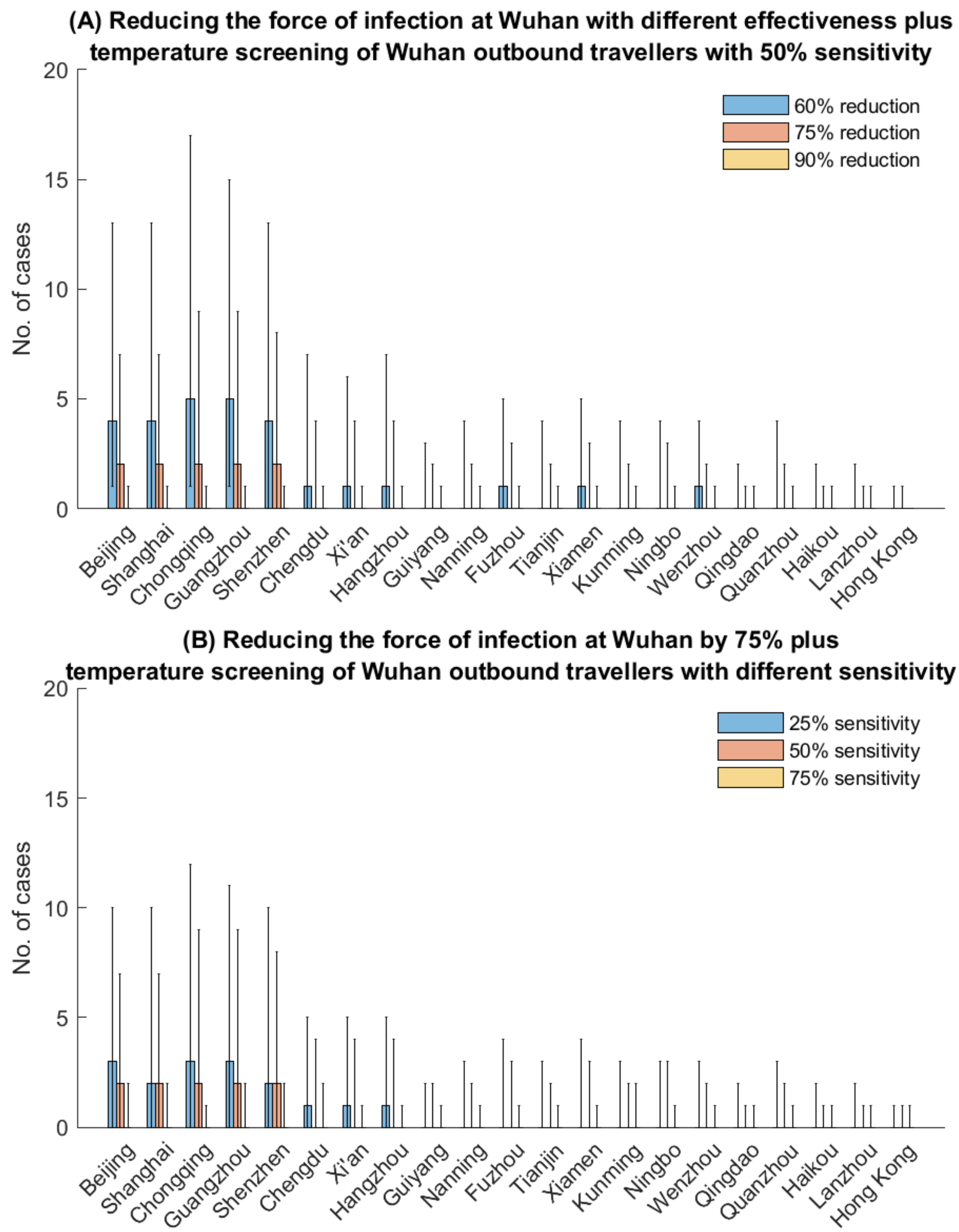

Figure 5. The total number of case exportation to Chinese cities to which Wuhan had the highest outbound travel volumes during one week immediately before and one week after the first day of Spring Festival. Before the first day of Spring Festival, passengers usually leave from developed coastal cities and railway interchange cities such as Beijing, Shanghai, Shenzhen and Guangzhou to less developed regions. The passenger flow direction is reversed after the Spring Festival.

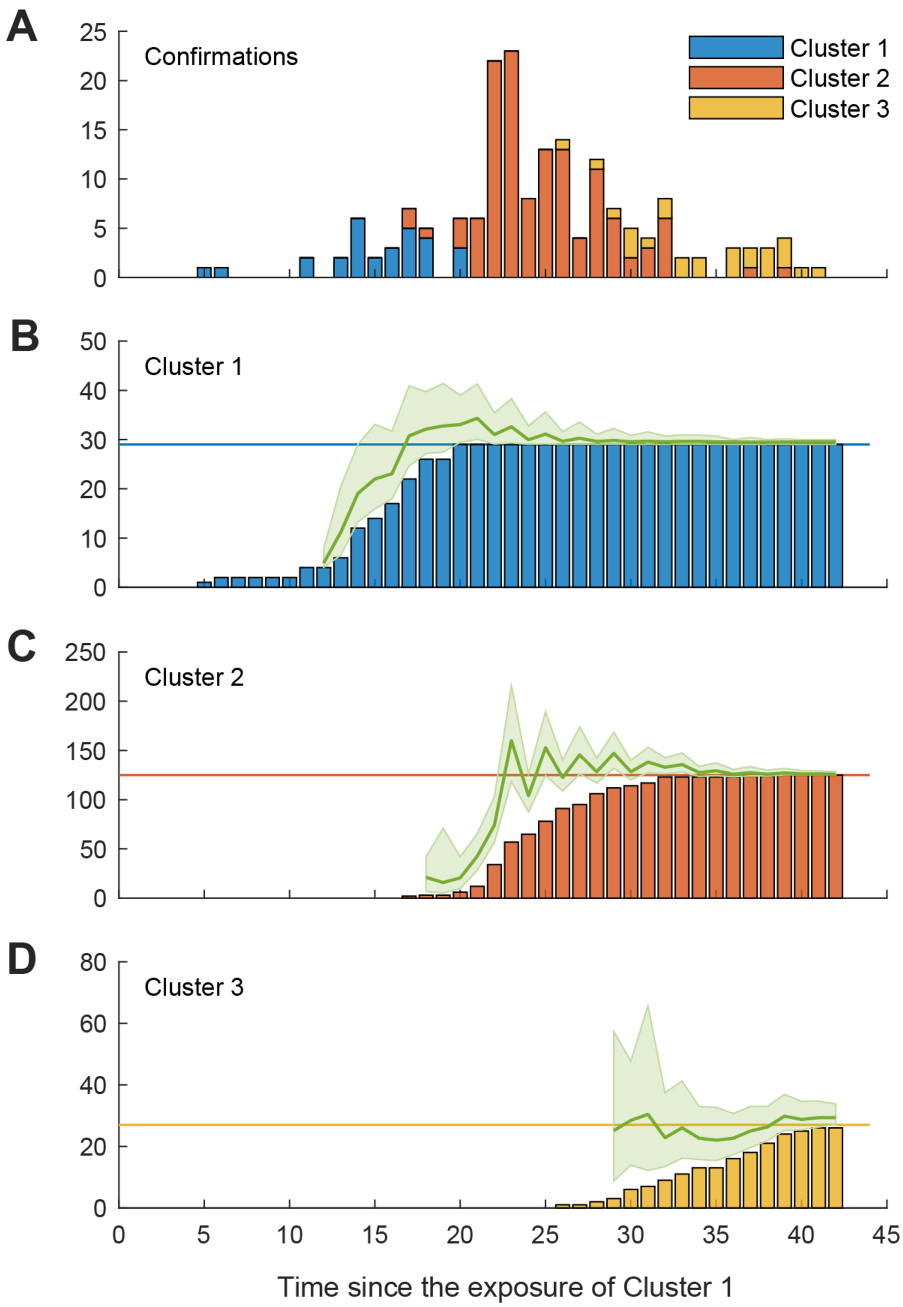

Figure 6. Retrospective real-time estimation of the final size of the three nosocomial MERS SSE clusters in Seoul in 2015. (A) Time of laboratory confirmation of MERS infection of the 181 cases. (B-D) Real-time estimation of the final size of Clusters 1-3. Cluster size estimation was performed at the end of each day. Green lines and shades indicate the posterior medians and the 95% credible intervals of the cluster size estimate, respectively. Bars indicate the cumulative number of lab-confirmed MERS cases. The horizontal line indicates the final size of the SSE cluster.

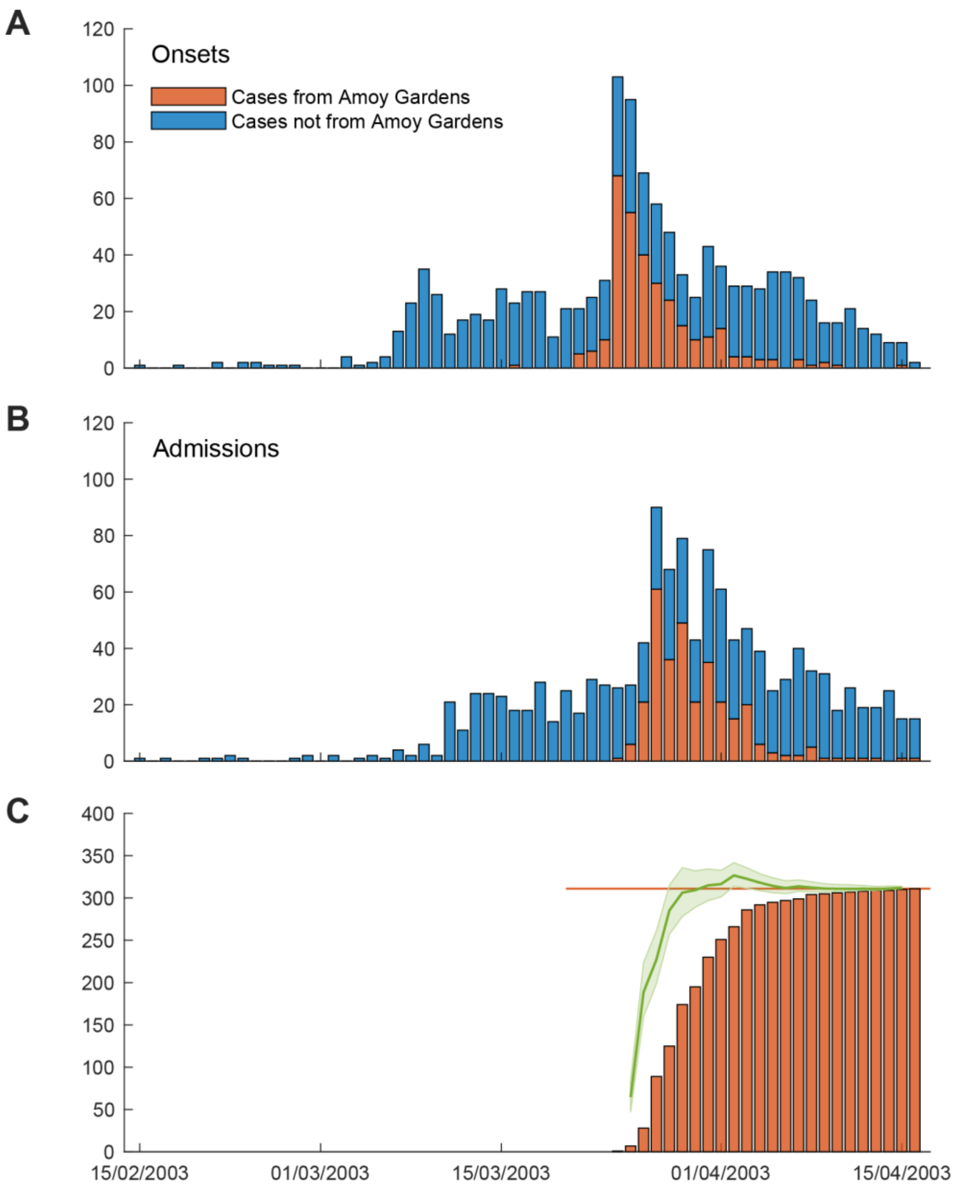

Figure 7. Retrospective real-time estimation of the final size of the SARS SSE cluster in Amoy Gardens in Hong Kong in 2003. The index patient developed symptoms on 14-15 Mar and visited Amoy Gardens on 14 and 19 Mar after his stay in Prince Wales Hospital on 15-18 Mar. We assumed that all cases in the Amoy cluster were infected on 19 Mar. (A) Time of symptoms onset of the 311 lab-confirmed cases from Amoy Gardens and 933 non-Amoy cases by 16 Apr with explicit onset time in our line list [1]. (B) Time of hospital admission of the Amoy cases and non-Amoy cases. (C) Real-time estimation of the final size. Cluster size estimation was performed at the end of each day. Green lines and shades indicate the posterior medians and the 95% credible intervals of the cluster size estimate, respectively. Bars indicate the cumulative number of lab-confirmed SARS cases. The horizontal line indicates the final size of the SSE cluster.

References

1. Leung, G.M., et al., *The epidemiology of severe acute respiratory syndrome in the 2003 Hong Kong epidemic: an analysis of all 1755 patients.* Annals of internal medicine, 2004. **141**(9): p. 662-673.
2. Breban, R., J. Riou, and A. Fontanet, *Interhuman transmissibility of Middle East respiratory syndrome coronavirus: estimation of pandemic risk.* The Lancet, 2013. **382**(9893): p. 694-699.
3. Cauchemez, S., et al., *Middle East respiratory syndrome coronavirus: quantification of the extent of the epidemic, surveillance biases, and transmissibility.* The Lancet Infectious Diseases, 2014. **14**(1): p. 50-56.
4. Hung, L.S., *The SARS epidemic in Hong Kong: what lessons have we learned?* Journal of the Royal Society of Medicine, 2003. **96**(8): p. 374-378.
5. McDonald, L.C., et al., *SARS in healthcare facilities, Toronto and Taiwan.* Emerging infectious diseases, 2004. **10**(5): p. 777.
6. Chowell, G., et al., *Transmission characteristics of MERS and SARS in the healthcare setting: a comparative study.* BMC Medicine, 2015. **13**(1).
7. Cowling, B.J., et al., *Preliminary epidemiological assessment of MERS-CoV outbreak in South Korea, May to June 2015.* Eurosurveillance, 2015. **20**(25).
8. Oh, M.-d., et al., *Middle East respiratory syndrome: what we learned from the 2015 outbreak in the Republic of Korea.* 2018. **33**(2): p. 233.
9. Jingchu.com. *Wuhan is expected to send 15 million passengers during Spring Festival in 2020.* 2019; Available from: http://hb.people.com.cn/n2/2019/1228/c194063-33671662.html.
10. Lloyd-Smith, J.O., et al., *Superspreading and the effect of individual variation on disease emergence.* Nature, 2005. **438**: p. 355.
11. Kucharski, A. and C. Althaus, *The role of superspreading in Middle East respiratory syndrome coronavirus (MERS-CoV) transmission.* Euro surveillance, 2015. **20**(25): p. pii= 21167.
12. Egan, J.R. and I.M. Hall, *A review of back-calculation techniques and their potential to inform mitigation strategies with application to non-transmissible acute infectious diseases.* Journal of The Royal Society Interface, 2015. **12**(106): p. 20150096.
13. Assiri, A., et al., *Hospital Outbreak of Middle East Respiratory Syndrome Coronavirus.* New England Journal of Medicine, 2013. **369**(5): p. 407-416.
14. World Health Organization. *Middle East respiratory syndrome coronavirus (MERS-CoV).* 2020 [cited 2020; Available from: http://www.who.int/emergencies/mers-cov/en/.
15. Virlogeux, V., et al., *Comparison of incubation period distribution of human infections with MERS-CoV in South Korea and Saudi Arabia.* Scientific Reports, 2016. **6**(1).

16. Desai, A.N., et al., *Real-time Epidemic Forecasting: Challenges and Opportunities.* Health security, 2019. **17**(4): p. 268-275.

17. Polonsky, J.A., et al., *Outbreak analytics: a developing data science for informing the response to emerging pathogens.* Philosophical Transactions of the Royal Society B, 2019. **374**(1776): p. 20180276.

18. Leung, G.M., et al., *Longitudinal assessment of community psychobehavioral responses during and after the 2003 outbreak of severe acute respiratory syndrome in Hong Kong.* Clinical Infectious Diseases, 2005. **40**(12): p. 1713-1720.

19. Wuhan Municipal Health Commission. *Wuhan Municipal Health Commission's briefing on the current pneumonia epidemic in the city*. Available from: http://wjw.wuhan.gov.cn/.

20. Lipsitch, M., et al., *Improving the evidence base for decision making during a pandemic: the example of 2009 influenza A/H1N1.* Biosecurity and bioterrorism: biodefense strategy, practice, and science, 2011. **9**(2): p. 89-115.

21. Chowell, G., et al., *Model parameters and outbreak control for SARS.* Emerging Infectious Diseases, 2004. **10**(7): p. 1258.

22. Lipsitch, M., et al., *Transmission dynamics and control of severe acute respiratory syndrome.* Science, 2003. **300**(5627): p. 1966-1970.

23. Riley, S., et al., *Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions.* Science, 2003. **300**(5627): p. 1961-1966.

24. Cauchemez, S., et al., *Unraveling the drivers of MERS-CoV transmission.* Proceedings of the National Academy of Sciences, 2016. **113**(32): p. 9081-9086.

25. Park, J.-E., S. Jung, and A. Kim, *MERS transmission and risk factors: a systematic review.* BMC public health, 2018. **18**(1): p. 574.

26. Walsh, E.E., J.H. Shin, and A.R. Falsey, *Clinical impact of human coronaviruses 229E and OC43 infection in diverse adult populations.* The Journal of infectious diseases, 2013. **208**(10): p. 1634-1642.

27. Zhang, S.-f., et al., *Epidemiology characteristics of human coronaviruses in patients with respiratory infection symptoms and phylogenetic analysis of HCoV-OC43 during 2010-2015 in Guangzhou.* PloS one, 2018. **13**(1): p. e0191789.

28. Gaunt, E.R., et al., *Epidemiology and clinical presentations of the four human coronaviruses 229E, HKU1, NL63, and OC43 detected over 3 years using a novel multiplex real-time PCR method.* Journal of clinical microbiology, 2010. **48**(8): p. 2940-2947.

29. Lau, E.H., et al., *A comparative epidemiologic analysis of SARS in Hong Kong, Beijing and Taiwan.* BMC infectious diseases, 2010. **10**(1): p. 50.

30. Vijgen, L., et al., *Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event.* Journal of virology, 2005. **79**(3): p. 1595-1604.
31. Lessler, J., et al., *Incubation periods of acute respiratory viral infections: a systematic review.* The Lancet infectious diseases, 2009. **9**(5): p. 291-300.
32. Mason, R.J., et al., *Murray and Nadel's textbook of respiratory medicine*. 2010: Elsevier Saunders.
33. Leung, G.M., et al., *Seroprevalence of IgG antibodies to SARS-coronavirus in asymptomatic or subclinical population groups.* Epidemiology & Infection, 2006. **134**(2): p. 211-221.
34. Al Kahlout, R.A., et al., *Comparative Serological Study for the Prevalence of Anti-MERS Coronavirus Antibodies in High-and Low-Risk Groups in Qatar.* Journal of immunology research, 2019. **2019**.
35. Abbad, A., et al., *Middle East respiratory syndrome coronavirus (MERS-CoV) neutralising antibodies in a high-risk human population, Morocco, November 2017 to January 2018.* Eurosurveillance, 2019. **24**(48).
36. Gill, E.P., et al., *Development and application of an enzyme immunoassay for coronavirus OC43 antibody in acute respiratory illness.* Journal of clinical microbiology, 1994. **32**(10): p. 2372-2376.
37. Chan, C., et al., *Examination of seroprevalence of coronavirus HKU1 infection with S protein-based ELISA and neutralization assay against viral spike pseudotyped virus.* Journal of Clinical Virology, 2009. **45**(1): p. 54-60.
38. Patrick, D.M., et al., *An outbreak of human coronavirus OC43 infection and serological cross-reactivity with SARS coronavirus.* Canadian Journal of Infectious Diseases and Medical Microbiology, 2006. **17**(6): p. 330-336.
39. Matsuno, A.K., et al., *Human coronavirus alone or in co-infection with rhinovirus C is a risk factor for severe respiratory disease and admission to the pediatric intensive care unit: A one-year study in Southeast Brazil.* PloS one, 2019. **14**(6): p. e0217744.
40. Jia, N., et al., *Case fatality of SARS in mainland China and associated risk factors.* Tropical Medicine & International Health, 2009. **14**: p. 21-27.

Supplementary Information

Definition of chunyun

Chunyun, also known as the Spring Festival travel season, is a period in China with extremely high traffic load around the Lunar New Year. It begins 15 days before the Lunar New Year's Day and lasts for 40 days. *Chunyun* 2019 was from 21 January 2019 to 1 March 2019 while *chungyun* 2020 is from 10 January 2020 to 18 February 2020.

Mobility data.

The mobility data were from three sources:

1) The monthly number of domestic and international flight bookings from Wuhan in January to February 2019 obtained from Official Airline Group (OAG)
2) The daily number of domestic passengers by means of transportation (i.e. air, train and road) recorded by Tencent's LBS (location-based services) database from Wuhan to 334 prefecture-level cities in mainland China from 6 January to 7 March 2019 (https://heat.qq.com/)
3) The domestic passenger volumes from and to Wuhan during *chungyun* 2020 estimated by Wuhan Municipal Transportation Management Bureau in the press releases in December 2019 [9]

Since the passengers recorded by the LBS database were only those using Tencent's services (e.g. WeChat, QQ, PUBG and Tencent Maps etc.), we calibrated the passenger volumes recorded in 2) by comparing the number of passengers by air in 2) with the actual number of domestic flight bookings in 1). We found that the number of passengers in 1) and 2) were highly correlated, with the Pearson's correlation coefficient of 0.944 (among the top 40 domestic destinations).

Domestic outbound passengers by air consists 86.5% of all outbound air volume. We first selected the top 40 destination cities which accounted for 94% of all outbound domestic flight volume of Wuhan and calculated the ratio of daily number of domestic passengers by air between OAG and Tencent. Then the ratios of the top 40 cities were weighted by the number of bookings of OAG and averaged during the period from 6 January to 7 March 2019 to obtain a scaling ratio γ to upscale the number of domestic passengers by air, train and road recorded in Tencent's LBS database. The domestic passenger volumes from all 334 prefecture-level cities recorded in 2) were scaled up with γ as the estimates of actual passenger volumes by air, train and road. The estimated passenger volumes during *chunyun* 2019 were used to estimate passenger volumes in *chungyun* 2020 from Wuhan by adjusting the difference in the overall volumes between 2019 and 2020.

Likelihood formulation.

Let X and Y be the incubation period and the delay from symptoms onset to case observation. Let $f(x,y|\theta)$ be the joint probability density function (pdf) of X and Y with parameters θ. Let $Z = X + Y$ (i.e. the time from infection to case observation) with cumulative density function (cdf) $Q(z|\theta) = \int_0^z \int_0^{z-y} f(x,y|\theta)dxdy$ and pdf $q(z|\theta) = Q'(z|\theta)$.

Let n_j be the size of cluster j, and $x_{i,j}$ and $y_{i,j}$ be the values of X and Y for the i^{th} case in cluster j. At time t, let $K_{j,1}$ be the set of observed cases in cluster j for whom the dates of onset and observation are both known and $K_{j,2}$ be the set of observed cases for whom only the date of observation is known, i.e. the total number of observed cases for cluster j is $|K_{j,1}| + |K_{j,2}|$. The likelihood function for cluster j is then

$$L_j(n_j,\theta) = \binom{n_j}{|K_{j,1}| + |K_{j,2}|}(1 - Q(t|\theta))^{n_j - |K_{j,1}| - |K_{j,2}|} \prod_{i \in K_{j,1}} f(x_{i,j}, y_{i,j}|\theta) \prod_{i \in K_{j,2}} q(z_{i,j}|\theta)$$

Similarly, let U_1 be the set of observed cases who do not belong to any cluster and for whom the dates of onset and observation are both known; and U_2 be the set of observed cases who do not belong to any cluster and for whom only the date of observation is known.

The overall likelihood function with W clusters is

$$\prod_{j=1}^{W} L_j(n_j,\theta) \prod_{i \in U_1} f(x_i, y_i|\theta) \prod_{i \in U_2} q(z_i|\theta)$$

We made the following specific assumptions in the MERS and SARS case studies:

1. X and Y are independent with pdf $g(x)$ and $h(y)$, respectively, such that $f(x,y) = g(x)h(y)$.
2. The pdf $g(x)$ and $h(y)$ are lognormal distributions.