Data Science and Management 5 (2022) 212–218

Contents lists available at ScienceDirect

Data Science and Management

journal homepage: www.keaipublishing.com/en/journals/data-science-and-management

Assessing spread risk of COVID-19 within and beyond China in early 2020

Shengjie Lai [a,b,*], Isaac I. Bogoch [c], Nick W. Ruktanonchai [a,d], Alexander Watts [e,f], Xin Lu [g,h], Weizhong Yang [i], Hongjie Yu [b,j], Kamran Khan [c,e,f], Andrew J. Tatem [a,**]

[a] WorldPop, School of Geography and Environmental Science, University of Southampton, Southampton, SO17 1BJ, UK
[b] Shanghai Institute of Infectious Disease and Biosecurity, Fudan University, Shanghai, 200433, China
[c] Department of Medicine, University of Toronto, Toronto, M5S2E8, Canada
[d] Population Health Sciences, Virginia Tech, Blacksburg, VA, 24061, USA
[e] Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, M5B1W8, Canada
[f] Bluedot, Toronto, M5J1A7, Canada
[g] College of Systems Engineering, National University of Defense Technology, Changsha, 410073, China
[h] Department of Public Health Sciences, Karolinska Institutet, Stockholm, 17177, Sweden
[i] School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, 100730, China
[j] School of Public Health, Fudan University, Key Laboratory of Public Health Safety, Ministry of Education, Shanghai, 200433, China

A R T I C L E I N F O

Keywords:
COVID-19
Human mobility
Pandemic
Mobile phone
Air travel

A B S T R A C T

A novel coronavirus emerged in Wuhan in late 2019 and has caused the coronavirus disease 2019 (COVID-19) pandemic announced by the World Health Organization (WHO) on March 12, 2020. This study was originally conducted in January 2020 to estimate the potential risk and geographic range of COVID-19 spread within and beyond China at the early stage of the pandemic. A series of connectivity and risk analyses based on domestic and international travel networks were conducted using historical aggregated mobile phone data and air passenger itinerary data. We found that the cordon sanitaire of Wuhan was likely to have occurred during the latter stages of peak population numbers leaving the city, with travellers departing into neighbouring cities and other megacities in China. We estimated that 59,912 air passengers, of which 834 (95% uncertainty interval: 478–1,349) had COVID-19 infection, travelled from Wuhan to 382 cities outside of Chinese maniland during the two weeks prior to the city's lockdown. Most of these destinations were located in Asia, but major hubs in Europe, the U.S. and Australia were also prominent, with a strong correlation seen between the predicted risks of importation and the number of imported cases found. Given the limited understanding of emerging infectious diseases in the very early stages of outbreaks, our approaches and findings in assessing travel patterns and risk of transmission can help guide public health preparedness and intervention design for new COVID-19 waves caused by variants of concern and future pandemics to effectively limit transmission beyond its initial extent.

1. Introduction

In December 2019, a cluster of patients with pneumonia of unknown cause was reported in the city of Wuhan, Hubei Province, China, epidemiologically associated with a seafood wholesale market (Huang et al., 2020; Tan et al., 2020). It has been determined that the pathogen causing the viral pneumonia among affected individuals was a new coronavirus (SARS-CoV-2) (Tan et al., 2020). The pathogen exhibited high human-to-human transmissibility and has spread rapidly within and beyond Wuhan city in early January of 2020 (Li et al., 2020). World Health Organization (WHO) declared the coronavirus disease 2019 (COVID-19) outbreak a Public Health Emergency of International

* Corresponding author. WorldPop, School of Geography and Environmental Science, University of Southampton, UK.
** Corresponding author.
E-mail addresses: Shengjie.Lai@soton.ac.uk (S. Lai), A.J.Tatem@soton.ac.uk (A.J. Tatem).

ELSEVIER Publishing services by Elsevier on behalf of KeAi

https://doi.org/10.1016/j.dsm.2022.08.004
Received 16 March 2022; Received in revised form 23 May 2022; Accepted 19 August 2022
Available online 26 August 2022

Concern on January 30th, 2020.

Wuhan is central China's transportation hub with a population of 11 million residents and a large number of higher-education students (~1.3 million in 89 universities and colleges), a particularly mobile population (Wuhan data). Beyond these factors, viral spread was likely exacerbated further by the surge in domestic and international travel during the 40-day Lunar New Year (LNY) celebrations (from January 10th, 2020 to February 18th, 2020), the largest annual human migration in the world, comprised of hundreds of millions of people travelling across China. As of February 4th, 2020, China has reported 20,530 confirmed cases and 23, 314 suspected cases with COVID-19 infections (National Health Commission of China, 2020). Of the confirmed cases, 2,788 are severe and 426 people have died. Most cases were reported from Wuhan and other cities in Hubei Province, and all provinces have confirmed cases imported from Wuhan and secondary transmission has been reported in some provinces. Additionally, there were 153 cases reported in 23 countries outside of China, with most having a travel history involving Wuhan (World Health Organization, 2020).

The potential pathway from this local outbreak in Wuhan to a pandemic might involve four steps: (i) local transmission in Wuhan (primary city of epidemic); (ii) spread from Wuhan to other cities within and beyond China via infected travellers, causing sporadic secondary transmission in these cities (secondary cities of epidemic), (iii) further spreading from secondary cities with local transmission to other tertiary cities in China and international cities via returning travellers after the LNY holiday; (iv) onward transmission across multiple countries, leading to a pandemic. To interrupt the spread, a cordon sanitaire of Wuhan and surrounding cities in Hubei Province has been in place since January 23rd, 2020, just two days before Chinese Lunar New Year's Day. However, significant numbers of people had likely travelled back to their hometowns for the holiday by this time. According to Wuhan authorities, it was likely that more than five million residents had already left the city before the lockdown, but where they went and how high the risk of spreading the virus remains an open question in early 2020 (CGTN, 2020).

In late January 2020, we conducted a travel network-based analysis to explore patterns of domestic and international population movements from high-risk cities in China, to provide preliminary estimates of the potential risk of COVID-19 spreading across and beyond the country. Given the limited understanding of the epidemiology of COVID-19 at the very early stage of the pandemic, our approaches of analysis and findings on travel patterns from historical mobility data can help contribute to identifying the spread risk and tailoring public health interventions for future COVID-19 waves or new pandemics caused by emerging human-to-human transmitted diseases.

2. Material and methods

To identify the areas that were most vulnerable to virus importation, we performed and integrated a series of analyses, by using de-identified and aggregated mobile phone-based population movement data, air passenger itinerary data, and case reports. We defined the potential risk and geographic range of COVID-19 virus spread across three scenarios: (1) from the primary city (Wuhan) into other cities in Chinese maniland (31 provincial regions), (2) from high-risk secondary cities into other cities across China, and (3) from high-risk cities in Chinese maniland into cities in other countries or regions during the LNY holiday in January 2020 and the following three months. We also estimated the number of airline travellers likely needing to be quarantined or screened to capture travellers potentially exposed to COVID-19 in high-risk cities of Chinese maniland.

2.1. Spread risk and destinations of COVID-19 from Wuhan

To define daily patterns and the connectivity of population movements at county and prefecture (city) level across Chinese maniland from January to April, we used the aggregated and de-identified daily flow of the users of Baidu, a Chinese search engine (https://www.baidu.com/). Baidu offers location-based service (LBS), based on the global positioning system, Internet Protocol address, location of signalling towers and Wi-Fi, for online searching, mapping, and a large variety of apps and softwares for mobile devices. These data have been used to visualize population migration around Chinese Lunar New Year.

Two Baidu datasets were used in this study. The first one covered daily movement data at county level from December 1st, 2013 to April 30th, 2014, as described elsewhere (Lai et al., 2020b). We calculated relative netflow following the equation below to extract daily patterns during LNY holiday at county level, with the population in 2014 obtained from National Bureau of Statistics of China (2020).

$$\text{Relative netflow} = (\text{inflow} - \text{outflow}) / \text{population of each county}$$

The second dataset was a daily movement matrix at the city level based on data from Baidu's search app from January 1st, 2015 to April 30th, 2015. The last recorded locations of a user (or device) for each day were compared, and if the location changed, we counted the user (or device) as someone who had moved from one city to another city. As the cordon sanitaire of Wuhan took place on January 23rd, 2020, just 2 days before LNY's day, and given that LNY's day in 2015 was February 19th, we took February 17th, 2015, as a reference of the lockdown day in our 2015 dataset. To understand the spread risk of COVID-19 from Wuhan into other cities via domestic population movement for the two weeks (the quarantine period of the virus) before the travel ban, we aggregated the historical daily outflows of people from Wuhan to other cities across Chinese maniland for the two weeks prior to February 19th, 2015. Based on the second dataset, assuming that the usual seasonal patterns of domestic and international population movements without travel restrictions were consistent during the LNY holiday across years in China, the risk of importation for each destination city or province was preliminarily defined as the percentage of travellers received by each city or province out of the total volume of travellers leaving Wuhan during the two weeks before the city's lockdown.

2.2. Spread risk from high-risk secondary cities in Chinese maniland

Without stringent non-phamaceutical interventions like Wuhan's, these secondary cities might have a high risk of community-level transmission through introducing infected travellers from Wuhan, and then spread the virus to other tertiary cities by returning population movements after the LNY holiday, causing an even wider spread of the virus. As most of the cities in Hubei province have implemented travel restrictions as Wuhan before LNY, we defined the high-risk secondary cities outside of Hubei province as the cities within top 30 ranked cities (Table S1 of Supplementary data) with the highest risk of importation from Wuhan defined above. Based on the 2015 Baidu dataset on population movement, the risk of spreading the virus from high-risk secondary cities to tertiary cities was preliminarily calculated as the averaged percentage of travellers received by each tertiary city out of the total volume of travellers leaving each high-risk secondary city during the four weeks following LNY's Day. We chose a period of four weeks because the returning flow of LNY's population movement, Chunyun, generally lasts four weeks.

2.3. Destinations of the virus spreading beyond Chinese maniland

To define the connectivity and risk of COVID-19 spreading from Wuhan and high-risk secondary cities defined above, into the cities beyond Chinese maniland, we obtained aggregated itinerary data from the International Air Travel Association (IATA). IATA data account for approximately 90% of passenger travel itineraries on commercial flights, and these data represent direct origin (Wuhan) to destination trips, and indirect trips that originated in Wuhan, but have connecting flights to a final destination. We quantified monthly volumes of airline travellers departing Wuhan and high-risk secondary cities from February 1st, 2018, through April 30th, 2018. With the assumption that the population movements around the LNY holiday in 2020 was consistent with the pattern in 2018, all final destinations were ranked by volumes of airline travellers, and the relative risk of importation was defined as the percentage of airline travellers received by each destination city out of the total volume of travellers leaving high-risk cities in China.

We also estimated the number of airline travellers that might have needed to be kept in quarantine from Wuhan during the two weeks prior to the city's travel ban. The LNY's day in 2018 was on February 16th and the lockdown of Wuhan happened two days before LNY's day, corresponding to the date of February 14th, 2018. We therefore defined the number of travellers needing to be quarantined as half of the volume of airline travellers from Wuhan in February 2018, representing the 2-week total number of travellers for the first half of February. We then estimated the number of infections and its 95% uncertainty interval (UI) in these airline travellers from Wuhan, based on a binomial distribution and the proportion of COVID-19 infections in the citizens evacuated from Wuhan reported by Singapore, South Korea, Japan, and Germany, as of January 31st, 2020 (World Health Organization, 2020; Ministry of Health of Singapore, 2019; Japanese Ministry of Health, 2020).

Additionally, to capture travellers potentially exposed to virus, we also estimated the volume of airline travellers that would be required to be screened at origin high-risk cities in China and destinations across the globe for the following three months of February to April, in the absence of a full international travel bans. Considering air traffic flows from China have changed due to the airline flight cancellations, travel restrictions imposed by countries or regions, or changes in travel behaviours, we calculated the volume of travellers using different scenarios of reductions (50%, 75%, and 90%) in total passenger volumes.

2.4. Validation

We compared the spatial patterns of the risks of Chinese cities importing the virus from Wuhan, estimated by the population movement data in 2014 and 2015, and more recent data on the top 50 ranked origin and destination cities in January 2020, available from the Baidu Migration site (https://qianxi.baidu.com/). To futher validate our results, we also compared the importation risk estimated in this study with the number of cases imported from Wuhan to other provinces reported in Chinese maniland, as of January 25th, 2020, and the number of imported cases reported by other countries or regions by February 3rd, 2020 (World Health Organization, 2020). The distribution between days of travelling from Wuhan, illness onset, first medical visit, and hospitalization of imported cases, as of January 25th, 2020, were also analysed. These case data were collated from the websites of WHO, national and local health authorities or new agencies within and beyond China (Supplementary data). R version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria) was used to perform data analyses.

3. Results

3.1. Risk and destinations of virus spread within Chinese maniland

Significant migratory flows occurred in opposite directions before and after LNY's Day. In Wuhan City, mass movements of people began about three weeks prior to LNY, with the first peak of population leaving the city before the start of the winter holiday for universities, especially in the three counties that contain many universities and students (Fig. 1). Although a cordon sanitaire of Wuhan and some cities in Hubei Province has been in place since January 23rd, 2020, the timing of this might have occurred during the latter stages of peak population numbers leaving Wuhan, as another peak of movements out of the city was seen 2 days before LNY's Day.

We found that a large number of travellers were likely departing

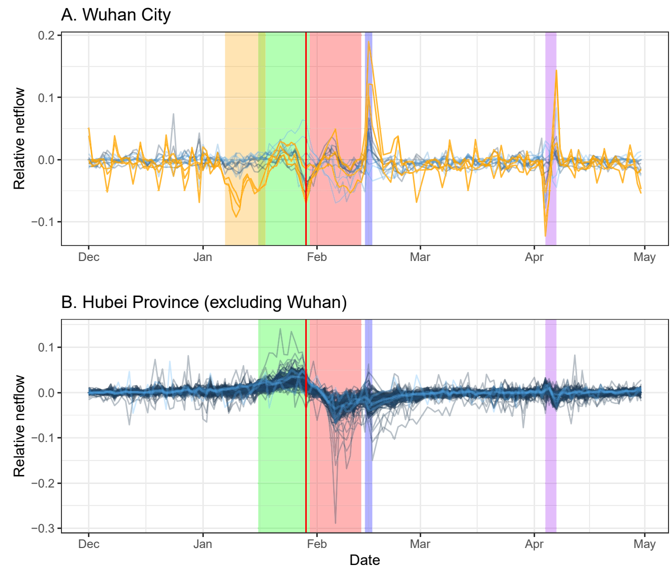

Fig. 1. Patterns of daily human movement across county in Wuhan City and Hubei Province across five months. (a) Wuhan City. (b) Hubei province (excluding Wuhan). Each blue line represents the netflow of population movement in each county. Yellow lines represent the netflow of population movement of three counties (Wuchang, Hongshan, and Jiangxia) with more universities or collegues in Wuhan. Vertical red line shows the day of cordon sanitaire in place in cities of Hubei. Shadow colours: yellow represents 2 weeks before the start of winter break of universities in Wuhan; green represents 2 weeks before LNY's Day; red represents 2 weeks since LNY's Day; blue represents Lantern Festival and weekend; purple represents Tomb Sweeping holiday and weekend. Relative netflow = (Inflow – Outflow)/population, based on the population movement data from 2013 to 2014 obtained from Baidu, Inc.

Wuhan into neighbouring cities and other megacities in China before Wuhan's lockdown (Table S1 of Supplementary data). Other cities in Hubei Province received huge amounts of people during the two weeks before LNY and showed a decreasing population since LNY, following a peak of outflow at the end of the LNY holiday (Fig. 1b and Table S2 of Supplementary data). If Wuhan's lockdown had not have been undertaken, our analyses suggest the main destination cities of population outflow since LNY would have been similar to the situation two weeks prior to LNY (Table S3 of Supplementary data).

In terms of the initial importation risk of virus for each city during the two weeks before Wuhan's lockdown, nearly all other cities in Hubei Province were estimated to be high-risk areas. Other places with high risks were Beijing, Shanghai, Guangzhou and other large cities. In terms of the provincial level, the risks were high in Guangdong and Hunan, followed by Henan and Zhejiang. There was a significant correlation (R-squared = 0.59, $p < 0.001$) between the number of imported cases and the risk of importation estimated from traveller numbers from Wuhan within the two weeks before LNY's Day (Fig. 2a). Further, a high proportion of cases travelled with symptoms at the early stage of the outbreak, and the lag from illness onset to hospitalization decreased from a median of 6 days (interquartile range: 4–7 days) in the first half of January 2020 to 3 days (1–5 days) in the second half (Fig. 3).

According to our definition outlined in the methods, the 17 high-risk secondary cities outside of Hubei Province were identified as: Beijing, Shanghai, Guangzhou, Zhengzhou, Tianjin, Hangzhou, Jiaxing, Changsha, Xi'an, Nanjing, Shenzhen, Chongqing, Nanchang, Chengdu, Hefei, Fuzhou, and Dongguan. Should community-level outbreaks occur in these cities, they could contribute to further spread of infection to other highly connected cities within China via movement after the holiday (Fig. S1 of Supplementary data).

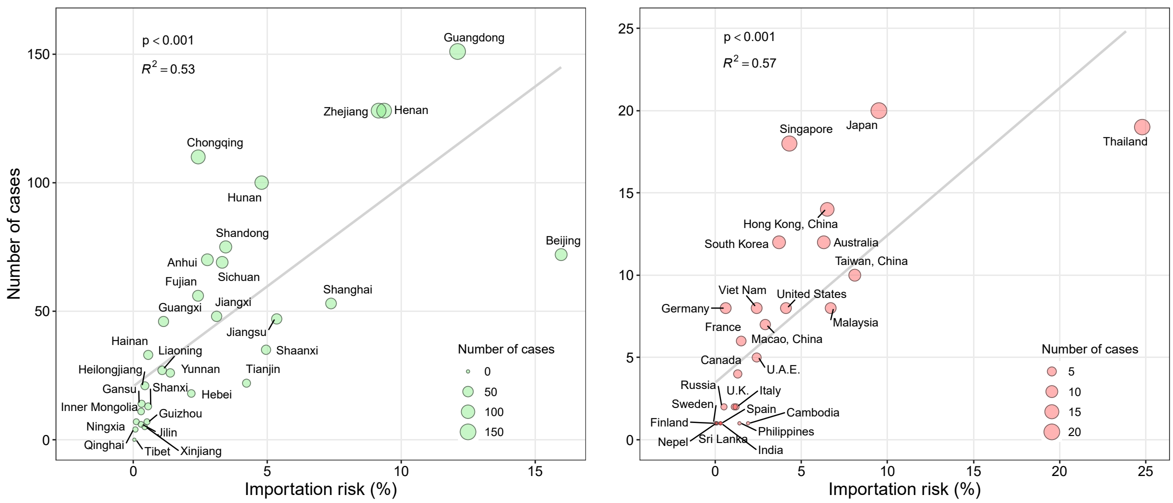

Fig. 2. Correlation between the number of cases reported and the risk of importation via travel. (a) Number of imported cases reported by each province (excluding Hubei), as of January 25th, 2020, versus the risk of importation from Wuhan. The risk of importation for each province was defined as the percentage of travellers received by each province out of the total volume of travellers leaving Wuhan during the two weeks before the city's lockdown. (b) Number of imported cases reported by each country or region, as of February 3rd, 2020, versus the risk of importation from Wuhan. The risk of importation for each country or region was defined as the percentage of travellers received by each destination out of the total volume of airline travellers leaving Wuhan from February to April 2018. Grey lines represent linear regression of importation risk against the number of cases reported, with *R*-squared and *p*-values are indicated on the graphs.

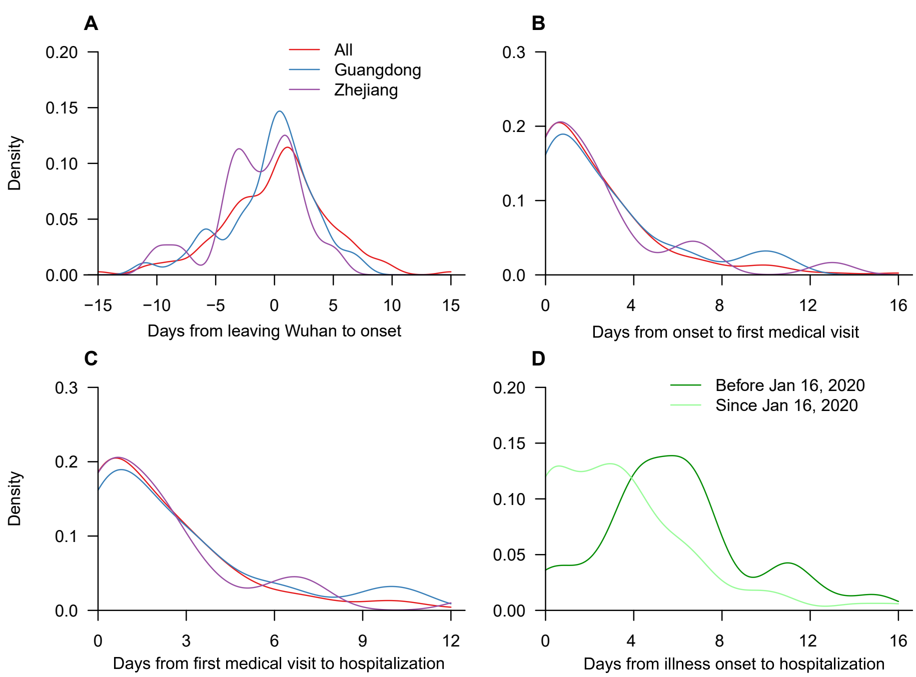

Fig. 3. Comparisons of time-delay distributions for cases imported from Wuhan before the city's lockdown on January 23rd, 2020. (a) Days from leaving Wuhan to illness onset ($N = 145$). The negative day means the illness onset before travel. (b) Days from illness onset to the first medical visit ($N = 164$). (c) Days from the first medical visit to hospitalization ($N = 164$). (d) Days from illness onset to hospitalization during the first half ($N = 67$) and second half (N = 80) of January 2020, respectively. A total of 164 cases with available data as of January 25th, 2020, were included. The Weibull distribution was used to estimate time-delay distributions.

3.2. *International spread risk and destinations*

Based on historical air travel data, the connectivity and spread risk between high-risk cities in Chinese maniland and cities in other countries or regions was defined for the three months around the LNY holiday (Appendix A Fig. S3 and Tables S4 and S5). Bangkok, Hong Kong, China and Taipei, China ranked in the top three, followed by Seoul, Tokyo and Singapore. The main destinations were presented by region in the supplemental materials (Tables S4–S7 of Supplementary data). During the two weeks before Wuhan's lockdown, there were an estimated total 59,912 airline travellers from Wuhan that might have needed to be kept in quarantine at the 382 destinations outside of Chinese maniland (Table S8 of Supplementary data). Thailand, Japan, and Taiwan, China ranked in the top three, followed by Malaysia, Hong Kong, China and Singapore. Based on an overall infection rate of 1.39% (16/1149; 95% UI: 0.80%–2.25%) in citizens evacuated from Wuhan before February 1st, 2020, reported by Singapore (1.08%, 1/92), South Korea (1.36%, 5/368), Japan (1.42%, 8/565), and Germany (1.61%, 2/124), we made preliminary estimates of a total of 834 (95% UI: 478–1349) airline travellers that might have been infected with COVID-19 from Wuhan two weeks prior to the city's lockdown. If adjusted by the estimated doubling time of the virus transmission from the lockdown to the evaualtion, 297 (170, 480) airline travellers could have been infected (Table S8 of Supplementary data). As of February 3rd, 2020, a significant correlation (*R*-squared = 0.58, $p < 0.001$) was seen between the number of imported cases reported in those countries or regions and the risk of importation via travellers defined in our study (Fig. 2b).

Because significant intranational and international spread from Wuhan has already occurred, a very large number of airline travellers (6.5 million under the scenario of 50% travel reduction as usual, 3.3 million under 75% reduction, and 1.3 million under 90% reduction, respectively) would be required to be screened at origin high-risk cities in China and destinations across the globe for the following three months of February to April 2020 (Table 1 and Tables S9 and S10 of Supplementary data), to ensure that all travellers from high risk cities were covered.

As seasonal patterns of domestic and international population movements across years were highly consistent with the timing and duration of public and school holidays in different countries (Lai et al., 2022), we also found similar spatial patterns in the risks of Chinese cities importing the virus from Wuhan via population movements estimated by both the Baidu data in 2014 and 2015, and the more recent data covering the top 50 ranked destinations in 2020, (Fig. 1a and Appendix A Fig. S10), highlighting the value of using historical data to rapidly assess present day risks.

4. Discussion

Mobile phone-based population movement data and air passenger itinerary data have been widely used to quantify the connectivity and transmission risk of pathogens via domestic and international human travel (Bogoch et al., 2016; Lai et al., 2018, 2019; Tatem, 2014; Tatem et al., 2012). Given the rapidly growing number of confirmed COVID-19 infections, increasing evidence of human-to-human transmission within and beyond China (Chan et al., 2020; Phan et al., 2020), and our limited understanding of this novel virus in the beginning of the pandemic (Callaway and Cyranoski, 2020; Munster et al., 2020), the approaches used in this study and our findings of travel patterns in historical data and spread risk estimation could help guide public health preparedness and intervention design for future COVID-19 waves caused by new SARS-CoV-2 variants of concern with high transmissibility or immunity escape (Lai et al., 2021a, 2021b, 2021c).

In terms of domestic connectivity and risk, the high population outflows from three counties in Wuhan with many colleges and universities in the first two weeks of January in 2020 were likely college students leaving the city to avoid the peak traffic just one week before the New Year (Tan et al., 2021). Because of the early timing of this movement, many students might have avoided the period when the virus spread rapidly in Wuhan, and their risk of spreading the virus may be low as well. However, our results suggest that during the two weeks prior to Wuhan's travel ban on January 23rd, 2020, a large number of travellers still departed Wuhan into neighbouring cities and other megacities in China and might have spread the virus to new areas, as the timing of the lockdown occurred during the latter stages of peak population numbers leaving Wuhan. Further exacerbating this risk, we found that during the outbreak's initial stages, a particularly high proportion of cases travelled with illness caused by the virus, together with the transmissibility of COVID-19 through asymptomatic contacts, potentially causing additional transmission during travel (Rothe et al., 2020).

Moreover, several destination cities outside of Hubei Province that received high volume of travellers from Wuhan prior to LNY's day could serve as significant secondary cities in the outbreak. Most of these cities have large populations and international airports, highly connected with other regions within and beyond China. The initial imported seed cases likely caused the local community transmission, and further spread the virus into wider geographical ranges following the population flows occurring due to the LNY holiday. Therefore, substantial public health interventions have been immediately applied across the country during the first wave of the pandemic (Lai et al., 2020b, 2021d), including the cordon sanitaire in several of the most affected cities, cancellation of mass gatherings, reduction of travel and contact rate, as well as the extension of the LNY and school winter holiday, which has mitigated subsequent local establishment of COVID-19 introduced by travellers.

Table 1
Top 30 ranked cities across the globe receiving airline travellers from 18 high-risk cities in Chinese maniland from February to April, representing three-month air traffic after LNY's holiday with travel banned from Wuhan and 75% reduction of travel from other cities.

Rank	Countries/ regions (top 30)	Volume (%)[a]	City (top 30)	Countries/ regions (top 30)	Volume (%)[a]
1	Thailand	485.6 (14.5)	Bangkok	Thailand	254.2 (7.7)
2	Japan	382.8 (11.5)	Hong Kong	Hong Kong, China	244.4 (7.4)
3	Hong Kong, China	244.4 (7.4)	Taipei, China	Taiwan, China	209.3 (6.4)
4	Taiwan, China	237.7 (7.3)	Seoul	South Korea	185.8 (5.7)
5	South Korea	230.2 (7.1)	Tokyo	Japan	173.4 (5.3)
6	U.S.	189.1 (5.1)	Singapore	Singapore	138.6 (4.2)
7	Malaysia	150.6 (4.4)	Phuket	Thailand	118.8 (3.6)
8	Singapore	138.6 (4.2)	Osaka	Japan	107.0 (3.3)
9	Vietnam	115.6 (3.5)	Kuala Lumpur	Malaysia	92.2 (2.8)
10	Australia	109.8 (3.2)	Macau, China	Macau, China	62.3 (1.9)
11	Indonesia	99.6 (2.8)	Denpasar Bali	Indonesia	53.1 (1.6)
12	Cambodia	64.3 (1.9)	Sydney	Australia	49.9 (1.5)
13	Macau, China	62.3 (1.9)	Chiang Mai	Thailand	37.8 (1.2)
14	Germany	58.0 (1.8)	Los Angeles	U.S.	38.0 (1.2)
15	Philippines	61.6 (1.8)	Melbourne	Australia	37.3 (1.1)
16	U.K.	46.7 (1.4)	Nagoya	Japan	34.6 (1.1)
17	Canada	50.7 (1.4)	London	U.K.	34.9 (1.1)
18	Italy	36.9 (1.1)	New York	U,S.	36.0 (1.1)
19	U.A.E	38.7 (1.1)	Ho Chi Minh City	Vietnam	34.6 (1.1)
20	Russia	36.9 (1.0)	Nha Trang	Vietnam	35.7 (1.1)
21	France	32.9 (0.9)	Dubai	U.A.E	34.4 (1.0)
22	India	26.0 (0.8)	Phnom Penh	Cambodia	31.2 (0.9)
23	New Zealand	29.5 (0.8)	Siem Reap	Cambodia	30.2 (0.9)
24	Spain	26.1 (0.7)	Paris	France	28.5 (0.9)
25	Egypt	14.3 (0.4)	Kota Kinabalu	Malaysia	29.6 (0.9)
26	Maldives	12.2 (0.4)	Manila	Philippines	29.5 (0.9)
27	Sri Lanka	13.8 (0.4)	Krabi	Thailand	29.8 (0.9)
28	Turkey	16.5 (0.4)	Frankfurt	Germany	25.5 (0.8)
29	Laos	8.6 (0.3)	Jakarta	Indonesia	27.8 (0.8)
30	Myanmar	10.4 (0.3)	Kaohsiung	Taiwan, China	24.8 (0.8)
	Other	254.6 (10.2)	Other		1015.8 (30.8)
	Total	3,285 (100)	Total		3,285 (100)

[a] In thousand. Based on air travel data from February to April 2018, obtained from the International Air Travel Association (IATA).

However, beyond the cases that have occurred in China, air passengers have spread COVID-19 across countries and continents within a short time period (Bogoch et al., 2020; Rothe et al., 2020). In particular, a high volume of international airline travellers left Wuhan for hundreds of destination cities across the world during the two weeks prior to the travel restriction implemented in the city. For example, we estimated that a huge volume of airline travellers would be required to be screened in February to April, even under the scenario of significant reduction (90%) in air passengers compared with the same period of previous years. Subsequently, a pandemic become inevitable as it was challenging to have substantial and timely responses in many destinations to prevent and contain further international seeding and local transmission (Yang et al., 2020). Of additional concern was that the anticipated destinations of hundreds of thousands of travellers were low-income or lower-middle income countries, where inadequately resourced medical and public health systems might be unable to detect and adequately manage the rapid spread of COVID-19. Therefore, ensuring that surveillance and health systems around the world are ready and sufficiently strong to detect and deal with emerging infections at the early stage is a priority for future pandemic response.

5. Limitations and conclusion

It is important to note that our study has several major limitations. Firstly, while we did present simple scenarios of reduced air travel volumes, our primary analyses assumed "business as usual" travel based on previous non-outbreak years, when significant spatio-temporal changes to human travel behaviours across and beyond China have occurred. Second, the mobile phone data used might provide an incomplete and biased picture of travellers, as the data only cover the population owning a smart phone and using Baidu app. Third, the case data used in this study likely varied in quality and completeness due to the timeliness of reporting, varying laboratory diagnosis capacities, and differences in details announced on health authority websites. Fourth, compared with airline travellers leaving Wuhan prior to January 23rd evacuees from Wuhan from the January 29th to 31st period might have a higher risk of infection due to their longer stay in Wuhan during the potential continued spread of the virus since January 23rd. This may result in overestimates of the number of infections in airline travellers from Wuhan prior to the city's lockdown.

Due to limited knowledge of the epidemiology of the virus (e.g., the proportion and infectiousness of asymptomatic or subclinical infections) at the time of writing in January of 2020 (Lai et al., 2020a), as well as the rapidly changing situation of a pandemic and interventions (Ge et al., 2022; Li et al., 2021), the simplicity of our approach to define importation risk can help to quickly update risk assessments, prioritise surveillance, target limited resources and understand the potential of pathogene introduction at specific destinations. Compared with other earlier studies (Pullano et al., 2020; Wu et al., 2020), we explored the various scenarios of travel restrictions and used a more comprehensive and spatio-temporally detailed population movement matrix, together with details on the actual final destination cities of air passengers based on the global itinerary dataset. These novel datasets provide new insights on the impacts of internal and international connectivity on the transmission and interventions of this emerging pathogen (Ruktanonchai et al., 2020), and our methods for assessing spread risk via travellers can help guide public health preparedness and intervention design for future pandemics.

Ethical statement

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.

Authors' contributions

SL designed the study, collected data, finalised the analysis, wrote the manuscript, and interpreted the findings. IIB, NWR, AW, and XL did data collection and analysis. AJT wrote the manuscript, interpreted the findings and revised drafts of the manuscript. IIB, NWR, WY, HY, and KK 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 the data in the study and had final responsibility for the decision to submit for publication. The corresponding authors attest that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Data sharing

The datasets on monthly air passenger data in February–April 2018 used in this study are available from Dr. Kamran Khan (Kamran@bluedot.global). The case data are available from Dr. Shengjie Lai (Shengjie.Lai@soton.ac.uk). The mobile phone datasets analysed during the current study are not publicly available, but 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).

Declaration of competing interest

The authors declare that there are no conflicts of interest.

Acknowledgments

We thank Baidu Inc. and IATA prodiving the data. We also thank staff members at disease control institutions, hospitals, and health administrations across China where outbreaks occurred for field investigation, administration, and data collection. This study was supported by the grants from the Bill & Melinda Gates Foundation (Grant Nos.: INV-024911 and OPP1134076); the European Union Horizon 2020 (Grant No.: MOOD 874850); the National Natural Science Fund of China (Grant Nos.: 81773498, 71771213 and 91846301); National Science and Technology Major Project of China (Grant No.: 2016ZX10004222-009); Program of Shanghai Academic/Technology Research Leader (Grant No.: 18XD1400300); Hunan Science and Technology Plan Project (Grant Nos.: 2017RS3040 and 2018JJ1034). AJT is supported by funding from the Bill & Melinda Gates Foundation (Grant Nos.: OPP1106427, OPP1032350, OPP1134076, and OPP1094793), the Clinton Health Access Initiative, the UK Department for International Development (DFID) and the Wellcome Trust (Grant Nos.: 106866/Z/15/Z and 204613/Z/16/Z). HY is supported by funding from the National Natural Science Fund for Distinguished Young Scholars of China (Grant No.: 81525023); Program of Shanghai Academic/Technology Research Leader (Grant No.: 18XD1400300); and the United States National Institutes of Health (Comprehensive International Program for Research on AIDS grant U19 AI51915). The research team members were independent from the funding agencies. The funders had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; and the preparation, review, or approval of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dsm.2022.08.004.

References

Bogoch, Ii, Brady, O.J., Kraemer, M.U.G., et al., 2016. Potential for Zika virus introduction and transmission in resource-limited countries in Africa and the Asia-Pacific region: a modelling study. Lancet Infect. Dis. 16 (11), 1237–1245.

Bogoch, Ii, Watts, A., Thomas-Bachli, A., et al., 2020. Potential for global spread of a novel coronavirus from China. J. Trav. Med. 27 (2), taaa011.

Callaway, E., Cyranoski, D., 2020. China coronavirus: six questions scientists are asking. Nature 577 (7792), 605–607.
CGTN, 2020. Five million people left Wuhan before the lockdown, where did they go? Available at: https://news.cgtn.com/news/2020-01-27/5-million-people-left-Wuhan-before-the-lockdown-where-did-they-go–NACCu9wItW/index.html.
Chan, J.F., Yuan, S., Kok, K.H., et al., 2020. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 395 (10223), 514–523.
Ge, Y., Zhang, W.B., Liu, H., et al., 2022. Impacts of worldwide individual non-pharmaceutical interventions on COVID-19 transmission across waves and space. Int. J. Appl. Earth Obs. Geoinf. 106 (1), 106102649.
Huang, C., Wang, Y., Li, X., et al., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395 (10223), 497–506.
Japanese Ministry of Health, 2020. Present situation of new type coronavirus infectious disease. Available at: https://www.mhlw.go.jp/stf/newpage_09290.html.
Lai, S., Bogoch, I.I., Watts, A., et al., 2020a. Preliminary risk analysis of 2019 novel coronavirus spread within and beyond China. Available at: https://www.worldpop.org/events/china.
Lai, S., Farnham, A., Ruktanonchai, N.W., et al., 2019. Measuring mobility, disease connectivity and individual risk: a review of using mobile phone data and mHealth for travel medicine. J. Trav. Med. 26 (3), taz019.
Lai, S., Floyd, J., Tatem, A., 2021a. Preliminary risk analysis of the international spread of new COVID-19 variants. Available at: https://www.worldpop.org/events/covid_variants.
Lai, S., Li, Z., Cleary, E., et al., 2021b. Exploring international travel patterns and connected communities for understanding the spreading risk of VOC Omicron. Available at: https://www.worldpop.org/events/covid_omicron. (Accessed 16 March 2022).
Lai, S., Loveridge, A., Tatem, A., 2021c. Indian international travel patterns for assessing the spreading risk of new COVID-19 variant. Available at: https://www.worldpop.org/events/covid_india. (Accessed 10 March 2022).
Lai, S., Ruktanonchai, N.W., Carioli, A., et al., 2021d. Assessing the effect of global travel and contact restrictions on mitigating the COVID-19 pandemic. Engineering 7 (7), 914–923.
Lai, S., Ruktanonchai, N.W., Zhou, L., et al., 2020b. Effect of non-pharmaceutical interventions to contain COVID-19 in China. Nature 585 (7825), 410–413.
Lai, S., Sorichetta, A., Steele, J., et al., 2022. Global holiday datasets for understanding seasonal human mobility and population dynamics. Sci. Data 9 (1), 17.
Lai, S.J., Johansson, M.A., Yin, W.W., et al., 2018. Seasonal and interannual risks of dengue introduction from South-East Asia into China, 2005-2015. PLoS Neglected Trop. Dis. 12 (11), e0006743.
Li, J., Lai, S., Gao, G.F., et al., 2021. The emergence, genomic diversity and global spread of SARS-CoV-2. Nature 600 (7889), 408–418.
Li, Q., Guan, X., Wu, P., et al., 2020. Early transmission dynamics in wuhan, China, of novel coronavirus-infected pneumonia. N. Engl. J. Med. 382 (13), 1199–1207.
Ministry of Health of Singapore, 2019. Updates on 2019 novel coronavirus. Available at:. https://www.moh.gov.sg/2019-ncov-wuhan.
Munster, V.J., Koopmans, M., Van Doremalen, N., et al., 2020. A novel coronavirus emerging in China - key questions for impact assessment. N. Engl. J. Med. 382 (8), 692–694.
National Bureau of Statistics of China, 2014. China Statistical Yearbook 2020. Available at: http://www.stats.gov.cn/english/Statisticaldata/AnnualData/.
National Health Commission of China, 2020. Updates on pneumonia of new Coronavirus infections as of March 31, 2020. Available at: http://www.nhc.gov.cn/xcs/yqtb/202004/28668f987f3a4e58b1a2a75db60d8cf2.shtml.
Office of Wuhan Statistics, 2020. Wuhan data. Available at: http://tjj.wuhan.gov.cn/.
Phan, L.T., Nguyen, T.V., Luong, Q.C., et al., 2020. Importation and human-to-human transmission of a novel coronavirus in vietnam. N. Engl. J. Med. 382 (9), 872–874.
Pullano, G., Pinotti, F., Valdano, E., et al., 2020. Novel coronavirus (2019-nCoV) early-stage importation risk to Europe, January 2020. Euro Surveill. 25 (4), pii=2000057.
Rothe, C., Schunk, M., Sothmann, P., et al., 2020. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. N. Engl. J. Med. 382 (10), 970–971.
Ruktanonchai, N.W., Floyd, J.R., Lai, S., et al., 2020. Assessing the impact of coordinated COVID-19 exit strategies across Europe. Science 369 (6510), 1465–1470.
Tan, S.-Y., Lai, S., Fang, F., et al., 2021. Mobility in China, 2020: a tale of four phases. Natl. Sci. Rev. 8 (11), nwab148.
Tan, W., Zhao, X., Ma, X., et al., 2020. A novel coronavirus genome identified in a cluster of pneumonia cases—wuhan, China 2019−2020. China CDC Weekly 2 (4), 61–62.
Tatem, A.J., 2014. Mapping population and pathogen movements. Int Health 6 (1), 5–11.
Tatem, A.J., Huang, Z., Das, A., et al., 2012. Air travel and vector-borne disease movement. Parasitology 139 (14), 1816–1830.
World Health Organization, 2020. Coronavirus disease (COVID-19) pandemic. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019.
Wu, J.T., Leung, K., Leung, G.M., 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 395 (10225), 689–697.
Yang, J., Li, J., Lai, S., et al., 2020. Uncovering two phases of early intercontinental COVID-19 transmission dynamics. J. Trav. Med. 27 (8), taaa200.