Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Assessing regional risk of COVID-19 infection from Wuhan via high-speed rail

Tao Li\textsuperscript{a,}\textsuperscript{*}, Lili Rong\textsuperscript{a}, Anming Zhang\textsuperscript{b}

\textsuperscript{a} Institute of Systems Engineering, Dalian University of Technology, PR China
\textsuperscript{b} Sauder School of Business, University of British Columbia, Canada

\textbf{ARTICLE INFO}

\textbf{Keywords:}
High-speed rail
Train and station
COVID-19
Regional infection risk
Probabilistic risk model

\textbf{ABSTRACT}

This paper demonstrates that transportation networks may be used to assess and predict the regional risk of COVID-19 infection from the outbreak. We use China’s high-speed rail (HSR) network at the scale of prefecture level to assess, based on a probabilistic risk model, the risk of COVID-19 infection from Wuhan to the country’s 31 province-level regions at the early stage of domestic spread. We find that the high-risk regions are mainly distributed along the southern half of Beijing-Hong Kong HSR line, where a large number of infection cases have been confirmed at the early stage. Furthermore, the two components of the infection risk, namely, the probability (proxied by the region’s correlation with Wuhan through HSR) and the impact (proxied by the region’s population with mobility), can play different roles in the risk ranking for different regions. For public health administrators, these findings may be used for better decision making, including the preparation of emergency plans and supplies, and the allocation of limited resources, before the extensive spread of the epidemic. Moreover, the administrators should adopt different intervention measures for different regions, so as to better mitigate the epidemic spread according to their own risk scenarios with respect to the probability of occurring and, once occurred, the impact.

1. Introduction

In December 2019, a cluster of atypical pneumonia cases was detected and reported in Wuhan, China, which was caused by an unknown pathogen (Zhang et al., 2020a; Oum and Wang, 2020; Sun et al., 2021a). The disease soon spread throughout China before Chinese Lunar New Year (January 25, 2020). In order to reduce the spread and mitigate the epidemic, the travel ban was implemented in Wuhan on January 23, 2020, and shortly after was extended to Hubei province where Wuhan is its capital city, and other regions. Outdoor movement restrictions were also adopted in a number of cities. The novel coronavirus (unknown pathogen) was named COVID-19 by the World Health Organization (WHO) on February 11, 2020. Initially, the COVID-19 epidemic was declared a public health emergency of international concern by WHO on January 30, 2020. On March 11, 2020, COVID-19 was declared a pandemic by WHO (WHO, 2020a).\textsuperscript{1} As of March 29, 2021, more than 126.89 million people have been confirmed to be infected (across over 223 countries, areas or territories), of which more than 2.78 million people died (WHO, 2020b).

Transportation network has been regarded as an important channel of spreading an epidemic, as human mobility has become increasingly dependent on it (Colizza et al., 2006; Jones et al., 2009; Bogoch et al., 2015; Zhao et al., 2020; Nakamura et al., 2020; Gaskin et al., 2021; Dong et al., 2021; Cui et al., 2021). For example, relying on global transportation network, SARS spread to more than 25 countries in 2003, MERS-CoV to 27 countries in 2012, and Ebola to more than eight countries in 2014. To prevent the spread of COVID-19 pandemic, therefore, numerous transportation facilities and services around the world have been shut down; relatedly, temperature screening measures have been implemented at airports and train/coach stations (Wu et al., 2020b; Bian et al., 2021; Mitrega and Choi, 2021; Kallbekken and Sælen, 2021). For example, the airport and stations in Wuhan have been temporarily closed (until April 8, 2020, 76 days after Wuhan lockdown); and airlines in countries such as the United States, the Philippines, Uzbekistan, and Singapore have progressively suspended flights to and from China.

\textsuperscript{*} Corresponding author. Linggong Rd., Dalian, 116024, PR China.
\textsuperscript{E-mail addresses:} taoli1987@outlook.com (T. Li), llrong@dlut.edu.cn (L. Rong), anming.zhang@sauder.ubc.ca (A. Zhang).

\footnotesize{\textsuperscript{1} In between the two declarations, WHO assessed the risk of spread of COVID-19 and the risk of its impact as very high at the global level on 28 February 2020 (WHO, 2020a).}

\url{https://doi.org/10.1016/j.tranpol.2021.04.009}
Received 30 March 2021; Accepted 10 April 2021
Available online 14 April 2021
0967-070X/© 2021 Elsevier Ltd. All rights reserved.
More specifically, transportation networks may be used to predict and mitigate an epidemic, such as by forecasting its potential dissemination (Bogoch et al., 2015; Wu et al., 2020b; Jia et al., 2020), assessing the risk of infection from outbreak (Chen et al., 2011; Christidis and Christodoulou, 2020; Gilbert et al., 2020), and investigating the effect of intervention measures on mitigation of the epidemic spread (Kraemer et al., 2020; Chinazzi et al., 2020; Oum and Wang, 2020). Linking transportation networks with epidemic control has become a research focus that has received more and more attention from scholars and practitioners in recent years. For the COVID-19 pandemic most studies, so far, have investigated the risk of COVID-19 infection from Wuhan to other domestic regions (Zhao et al., 2020; Zhang et al., 2020b; Wei et al., 2021) or from one country to other countries via civil aviation (Christidis and Christodoulou, 2020). Some researchers have found that there is a strong correlation between the cumulative number of confirmed cases of COVID-19 and travel by train in China, while travel by car and flight is not significantly correlated with the cumulative number (Zhao et al., 2020).

In 2019, the total number of passengers transported by railway reached 3.66 billion, of which high-speed rail (HSR) accounts for 64.4%. In comparison, the total number of passengers transported by civil aviation only reached 660 million, about a quarter of the HSR traffic (NBSC, 2020). At the end of 2019, the high-speed rail network (HSRN) length of China was 35,000 km, accounting for more than 60% of the world’s total (NRAC, 2020), and there were more than 900 stations and more than 4700 high-speed trains operating in the network each day. HSR has become one of the main choices for interregional travel in China, especially for short- and medium-distance travel (Wang et al., 2017; Zhu et al., 2018, 2019; Liu et al., 2019, 2020; Li et al., 2020a). As a core HSR hub, Wuhan plays a major role in China high-speed rail network (CHSRN), bridging both the North-South parts, and the East-West parts, of China (Jiao et al., 2017, 2020; Li et al., 2019; Li and Rong, 2020; Zhang et al., 2019; Wang et al., 2020; Yang et al., 2020). As a result, the outbreak of COVID-19 in Wuhan can rapidly and easily spread to other regions when effective control measures cannot be taken in time, owing to the lack of awareness of COVID-19. In other words, HSRN can be a key channel of exporting cases from Wuhan to other domestic regions.

In this study, we assess the regional risk of COVID-19 infection from Wuhan during the initial period of pandemic spread in mainland China, based on the operation data of HSR. We develop a probabilistic risk model in which the regional risk of COVID-19 infection from Wuhan is taken as a function of both the probability that a region is infected with COVID-19 from Wuhan, and the impact of the infection. More specifically, we measure the probability by the correlation of each region with Wuhan in terms of connectivity in CHSRN, and the impact by the region’s population with mobility. We then rank the risk of COVID-19 infection from Wuhan for the 31 province-level regions in mainland China, and compare this ranking with the actual, cumulative number of confirmed cases of COVID-19 for five different periods during the early transmission.

We find that regions with a high risk of infection mainly distribute along the southern half of Beijing-Hong Kong HSR line, and these regions have a large number of cumulative cases for all the five periods. Furthermore, the two components of the infection risk, namely, the probability (proxied by the region’s correlation with Wuhan through HSR) and the impact (proxied by the region’s population with mobility), can play different roles in the risk ranking for different regions. For public health administrators and policy makers, these findings regarding early-stage regional risks can be used for better decision making, including on the preparation of emergency plans and emergency supplies, and the allocation of limited resources (e.g., personal protective equipment such as N95 respirators, gowns, masks, gloves, and face shields) ahead of the ongoing outbreaks to avoid resource shortage and inefficient, unbalanced distribution. Moreover, for different regions, the administrators should adopt different intervention measures so as to better mitigate and control the epidemic spread according to their own risk scenarios with respect to the probability of occurring and, once occurred, the impact.

The rest of this paper is organized as follows: Section 2 reviews the studies that are relevant to ours. Section 3 presents our research methodology including the description of HSR network and our assessment model of infection risk. The study area and data sources are described in Section 4, followed by Section 5 that covers the results and discussion. Finally, Section 6 concludes the study and discusses avenues for further research.

2. Literature review

In recent decades, researchers from different fields have conducted numerous studies on infectious diseases (Wu et al., 2020a), such as investigating their whole spectrum and pathophysiology via retrospective studies on infected cases, including clinical spectrum, reproduction interval, incubation period, transmission modes (Garbino et al., 2006; Yin and Wunderink, 2018; Zhu et al., 2020; Huang et al., 2020; Guan et al., 2020; Li et al., 2020a; Weisberg et al., 2021; Cortinovis et al., 2021), exploring the effect of control measures on reducing their spread (Khachatrian et al., 2015; Funk et al., 2017; Tian et al., 2020; Oraby et al., 2021; Anderson et al., 2021), and examining their socio-economic impact (Hai et al., 2004; Keogh-Brown and Smith, 2008; Dénès and Guelml, 2019; Beck et al., 2020; Mueller et al., 2021).

However, many of these studies in transportation have focused on the potential dissemination and the infection risk of infectious diseases (Grais et al., 2003; Jones et al., 2009; Chen et al., 2011; Bogoch et al., 2015; Nakamura et al., 2020; Wu et al., 2020b; Zhang et al., 2020a; Sun et al., 2020, 2021a). However, the risk assessment can be taken as the indicator of the dissemination, while the pre-assumption of the indicator needs more evidence.

Based on these studies, the dissemination risk of COVID-19 pandemic has attracted increasing attention (Weible et al., 2020). Although there are many studies on the control measures of mitigating the pandemic and its impact as it spreading around the world (Kraemer et al., 2020; Chinazzi et al., 2020; ADB, 2020; Cui et al., 2021; Sun et al., 2021b), there are relatively few studies that focus on its dissemination risk in the early stage of the pandemic spread (Wu et al., 2020b; Zhang et al., 2020a; Sun et al., 2020, 2021a). However, the risk assessment can be used to better allocate limited resources ahead of the ongoing outbreaks, and better prepare the mitigation measures for policy makers. These studies focusing on the risk assessment can be summarized into two categories according to their research methods.

The first category is based on a disease transmission model, with an objective to assess the risk of COVID-19 transmission, such as SIR epidemic model, and SEIR epidemic model. Wu et al. (2020b) describe the risk of COVID-19 infection from Wuhan as the relative number of confirmed cases exported from Wuhan and extend the SEIR epidemic model into a SEIR-meta-population model to simulate the transmission dynamics of COVID-19 epidemic, and then obtain the infection risk of more than 300 prefecture-level cities by using flight bookings data and human mobility data. It is noted that the above method usually requires a lot of assumptions and parameters in advance, such as pre-assessment.
or pre-assumption of the basic reproductive number of COVID-19, and the assumptions of similar transmissibility across all areas.

The second category is based on probability models, in which the infection risk is always defined as a function of both the probability for an infection to occur and its impacts. Although Jia et al. (2020) also regard the risk of COVID-19 infection from Wuhan as infection cases, they propose a risk source model to calculate the risk by using population flow data, in which the greater the population inflow to a region from Wuhan, the greater the probability of infection and the greater the cumulative number of infection cases in the region. A limitation of this method is that it requires detailed data as support and relies on a sophisticated approach towards analyzing data, such as machine learning. In the research of Pullano et al. (2020), the risk of COVID-19 infection from China to Europe is estimated as the probability of imported infection cases via air travel data. They do not provide the risk for an individual country however, because of the difficulty in assessing travel-related imported cases. By applying a similar method, Gilbert et al. (2020) assess the importation risk from China to each country in Africa.

From the prior related works, it is easy to see that most of them are limited to use the epidemic models coupled with air travel data or use the risk assessment model with much assumptions and parameters. Based on this observation, this study firstly proposes a method for assessing the infection risk from the outbreak at the early stage of a pandemic spread, which does not depend on too many assumptions and parameters and uses the operation data of HSR. The aim of this work is to assess the risk of COVID-19 infection from Wuhan to the 31 province-level regions in China via its HSR network. It is noted that in China, prevention and control of COVID-19 epidemic are usually the responsibility of province-level governments (under the general leadership of the central government), so we focus on province-level regions rather than prefecture-level cities. Different from the previous literature, we will, as discussed more fully in Section 3, focus on the relative infection risk between regions rather than the cumulative number of confirmed cases in regions. Further, the definition of the risk will be separated from the number of infected cases, due to the difficulty in assessing the number. We also note that most of the data we will use is publicly (and easily) available, in contrast to the human mobility data and population flow data used in the previous studies. Besides, at the regular prevention and control stage of the post COVID-19 world, our method can be extended to easily combine with the main transportation network of a region and quickly forecast the regional risk under the sporadic outbreaks of the pandemic.

3. Methodology

For ease of exposition, the description of HSR network (HSRN) and the relevant notions are outlined, which can then be used for developing the assessment model of regional infection risk.

3.1. High-speed rail network

In this research, HSRN is described as $G = (S, L, T, Q)$ at the scale of prefecture-level cities. In the network, $S = \{s_i | i \in N\}$ denotes the set of HSR stations locating in the cities at prefecture-level, where $N = \{1, 2, \ldots, n\}$ denotes the set of the stations number and $n$ is the sum of the stations. And $L = \{l_{ij} | i,j \in N, i \neq j\}$ denotes the set of service links between these stations, $T = \{t_{ij} | i,j \in N, i \neq j\}$ denotes the set of weights of service links, which is indicated by the travel time spent on service links. Clearly, a link in $L$ is determined by high-speed trains and its weight is determined by the running time of these trains, which means if a high-speed train stops at two stations, there is a link connecting the stations and the weight of the link is the in-vehicle time of the train between the two stations, as shown in Fig. 1. Moreover, $Q = \{q_{ij} | i,j \in N, i \neq j\}$ denotes the set of high-speed train frequency on service links.

3.2. Assessment model of regional infection risk

The regional infection risk is evaluated on the basis of a conventionally probabilistic risk model in ISO (2009), Liao et al. (2012), Shankar et al. (2018), Koks et al. (2019) and Kang et al. (2021). In this study, the regional infection risk is defined as a function of both the probability that a region is infected with COVID-19 from Wuhan at the early stage, and the impact of the infection. In other words, the regional infection risk can be expressed as:

$$R_A = P_A \times I_A$$

where $R_A$ denotes the infection risk of region $A$ importing the virus from Wuhan, $P_A$ denotes the corresponding probability, and $I_A$ denotes the impact of the infection on region $A$. As shown below, this regional infection risk is a relative risk rather than an absolute risk (such as the risk measured by the cumulative number of confirmed cases), indicating the magnitude of the risk as compared to that of other regions.

3.2.1. Probability of the infection

In Eq. (1), probability $P_A$ is related to the correlation between region $A$ and Wuhan. The stronger the correlation between them, the higher the probability of COVID-19 infection from Wuhan to region $A$. In this study, we define $P_A$ as a function of both the scope and intensity of the correlation between region $A$ and Wuhan from the perspective of HSRN. In other words, if a region is extensively and intensively connected to Wuhan on HSRN, then the region is more probable to import COVID-19 from Wuhan and hence is to be infected. Furthermore, the scope of the correlation is measured by the number of HSR stations located in region $A$ and connected to Wuhan’s station, whereas the intensity of the correlation by the spatial distance and passenger flow from Wuhan to region $A$ on HSRN. Based on this discussion, probability $P_A$ is given by:

$$P_A = C_A^2 \times C_A^3$$

Fig. 1. A simple description example of HSR network. The left part of the figure represents the bipartite graph consisting of HSR stations and high-speed trains, which can be used to establish the service links between these stations located in different cities of different regions in the right part of the figure. In the right part, the colored lines denote the service links, whereas the bold numbers next to them represent travel time and the numbers in brackets denote the frequency of high-speed train on service links.
where \( C^S_A \) and \( C^I_A \) indicate the scope of the correlation and the intensity of the correlation between region \( A \) and Wuhan, respectively. Further, \( C^A_S \) can be calculated by the following equations:

\[
C^A_S = \begin{cases} 
\hat{k}_u, & \text{if region } A \text{ is directly connected to Wuhan} \\
\hat{k}_w, & \text{if region } A \text{ is indirectly connected to Wuhan} \\
0, & \text{if region } A \text{ is not connected to Wuhan}
\end{cases}
\]

Further, \( \hat{k}_w = \frac{k^U_w}{\max_{X \subseteq M} k^X_w} \)

with

\[
\hat{k}_w = \frac{k^U_w}{\max_{X \subseteq M} k^X_w}
\]

Fig. 2. Population and distribution of HSR stations in each province-level region at the end of 2019. The degree of each HSR station is calculated by the number of service links directly connecting it on HSRN. The background color of each region indicates its population, whereas the size and color of each point indicate the degree of each station. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Fig. 3. Daily high-speed train frequency at each province-level region, and the daily high-speed train frequency and number of HSR stations connected to Wuhan. The background color of each region indicates the daily frequency of high-speed trains stopping in this region. The size of each red point indicates the number of stations located in a region and connected to Wuhan. Besides, the width and color of each line indicate the daily frequency of high-speed trains stopping in both Wuhan and each region. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
In the above equations, $k_A^w$, $k_U^w$, and $k_X^w$ denote the number of HSR stations connected to Wuhan in regions $A$, $U$, and $X$, respectively, and region $U$ is region $A$'s nearest neighbor on the shortest path to Wuhan on HSRN. $M$ is the set of all regions in China, and $\hat{k}_A^w$ and $\hat{k}_U^w$ denote the normalization of $k_A^w$ and $k_U^w$, respectively. $a_{Xi}$ denotes whether HSR station $i$ is located in region $X$, and $b_{iw}$ denotes whether there is a service link between station $i$ and Wuhan's station. Furthermore, $C_A^I$ can be calculated as:

$$C_A^I = \begin{cases} \frac{\hat{k}_A^w}{\hat{d}_A^w}, & \text{if region } A \text{ is directly connected to Wuhan} \\ \frac{\hat{k}_U^w}{(\hat{d}_A^w \times \tau)^+}, & \text{if region } A \text{ is indirectly connected to Wuhan} \\ 0, & \text{if region } A \text{ is not connected to Wuhan} \end{cases}$$

(9)

with

$$\hat{k}_A^w = \frac{k_A^w}{\max_{X \in M} f_X^w}$$

(10)

$$\hat{k}_U^w = \frac{k_U^w}{\max_{X \in M} f_X^w}$$

(11)

$$f_X^w = \sum_{i=1}^N a_{Xi} \times b_{iw}, \text{ applying to } f_A^w \text{ and } f_U^w$$

(12)

Fig. 4. Average travel time from Wuhan to each province-level region on HSRN. The travel time is measured in minute, and the transfer time at each transfer station is set to 30 min. For Hainan province, Ningxia Hui autonomous region and Tibet autonomous region, Wuhan is unreachable.

Fig. 5. Probability of COVID-19 infection from Wuhan to each province-level region. The map on the bottom left represents the scope of the correlation between province-level regions and Wuhan on HSRN, and the map on the bottom right represents the intensity of that correlation on CHSRN. Different colors indicate different values of the probability, scope and intensity. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
In the above equations, $f_A^w$, $f_U^w$, and $f_X^w$ denote HSR passenger flows from Wuhan to regions $A$, $U$, and $X$, respectively, and $\hat{f}_A^w$ and $\hat{f}_U^w$ denote the normalization of $f_A^w$ and $f_U^w$, respectively. $f_w^u$ denotes the passenger flow from Wuhan’s station to station $i$ on HSRN, $d_{w}^i$ denotes the spatial distance from Wuhan’s station to station $i$, and is measured by $t_{w}$ which represents the shortest travel time between them. Further, $t_{w}^i$ and $t_{w}^t$ denote, respectively, the in-vehicle time measured by the running time of high-speed trains, and the transfer time at transfer station from Wuhan’s station to station $i$. Finally, $c_{A}^i$ denotes the shortest path length from Wuhan’s station to station $i$ located in region $A$ on HSRN, and $\hat{c}_A^i$ denotes the average of $c_{A}^i$.

### 3.2.2. Impact of the infection

The impact of virus infection is related to the region’s population and its mobility (Bogoch et al., 2015; Riad et al., 2019; Kraemer et al., 2020; Jia et al., 2020) in that the more population and more mobility of the population in a region, the greater the impact of virus dispersal in the region (indicating the greater number of cities infected with virus from the outbreak core region and the greater number of imported virus cases). From the perspective of HSRN, we define impact $I_A$ as the product of the population and the development state of HSRN in region $A$, with the development state indicating the population mobility. Thus,

$$I_A = H_A \times M_A$$

where $H_A$ and $M_A$ denote the population and the development state of HSRN in region $A$, respectively. They are given by:

$$\begin{align*}
H_A &= \frac{\text{pop}_A}{\max_{X \in M} \text{pop}_X} \\
M_A &= \frac{\hat{k}_A \times \hat{f}_A}{\max_{X \in M} k^X} \\
\hat{k}_A &= \frac{\sum_{i=1}^{N} a_{A}^i}{\max_{X \in M} k^X} \\
\hat{f}_A &= \frac{f^A}{\max_{X \in M} f^X}
\end{align*}$$
\[
f^A = \sum_{i=1}^{N} \sum_{j \neq i} a^i_j \times f_{ij}, \quad \text{applying to } f^A
\]

where \(\text{pop}_A\) and \(\text{pop}_X\) denote the population in regions A and X, respectively, \(k^A\) and \(k^X\) denote the number of HSR stations located in regions A and X, respectively. Furthermore, \(f^A\) and \(f^X\) denote the HSR passenger flows in regions A and X respectively, \(f_{ij}\) denotes the passenger flow from stations \(i\) to \(j\), and \(\hat{k}^A\) and \(\hat{f}^A\) denote the normalization of \(k^A\) and \(f^A\) respectively.

4. Study area and data

This study focuses on mainland China which is composed of 31 province-level regions. In addition to four municipalities, each province-level region consists of more than multiple prefecture-level cities. There are 338 prefecture-level cities and cites above covered by the study area. Due to China’s special geomorphological and demographic characteristics, HSR has become the main transportation mode between these regions or cities. Among these 338 cities, there are 230 cities operating HSR in urban areas. Whereas, two of these 31 province-level regions don’t have HSR at the end of 2019, which are Ningxia Hui autonomous region and Tibet autonomous region.

As the capital of Hubei province, Wuhan is located in the central China and is the core of the Yangtze River Economic Belt. With a population more than 11 million, Wuhan is also an important transportation hub, especially in CHSRN (State Council of the People’s Republic of China, 2017), and has a high human mobility. At the end of 2019, Wuhan had 200 civil aviation routes including 63 international routes and its international trains of Wuhan reached 11 countries. Its 2019 passenger traffic volume of Wuhan reached 253.7 million (Wuhan Bureau of Statistics, 2020). In China, Wuhan is the original place of the epidemic outbreak and is hit by the disease the hardest. The epidemiological investigations indicate that most of the confirmed cases of COVID-19 have a travel history to Wuhan or contacts with cases from Wuhan during the initial stage of COVID-19 transmission (Li et al., 2020c).

The data used in this paper includes the confirmed case data of COVID-19, the demographic data of 31 province-level regions, and the operation data of HSR. The case data of COVID-19 is obtained from the official website of the Health Commission of each province-level region. The demographic data is obtained from the Statistical Bulletin of National Economic and Social Development of each province-level region (year 2019), which is published on the official website by the Bureau of Statistics of each province-level region.

The operation data consists of HSR timetable data and passenger flow data, obtained from the 12306 China Railway website, China State Railway Group Co., Ltd., National Railway Administration of the People’s Republic of China (NRAC), Ministry of Transport of the People’s Republic of China. According to the regulation of NRAC, the HSR is designed to operate at speeds above 250 km/h, initially operated at speed above 200 km/h, and dedicated to operating passenger trains, which involves HSR lines and high-speed trains (prefixed with G, D, and C). Because several cases were reported in Wuhan at the end of 2019, Wuhan had 200 civil aviation routes including 63 international routes and its international trains of Wuhan reached 11 countries. Its 2019 passenger traffic volume of Wuhan reached 253.7 million (Wuhan Bureau of Statistics, 2020). In China, Wuhan is the original place of the epidemic outbreak and is hit by the disease the hardest. The epidemiological investigations indicate that most of the confirmed cases of COVID-19 have a travel history to Wuhan or contacts with cases from Wuhan during the initial stage of COVID-19 transmission (Li et al., 2020c).
December 2019, and the travel ban was implemented in Wuhan on January 23, 2020. Further, the purpose of this study is to assess the regional risk of COVID-19 infection from Wuhan at the early stage of domestic spread. Thus, we need to obtain the operation data of HSR before 23 January 23, 2020. Besides, the daily operation data of HSR is consistent and stable during a certain period of time. So we acquire the operation data on December 28, 2019.

In our study area, the HSRN consists of 230 nodes and 6726 service links, which is described on the basis of 4675 high-speed trains. The basic condition descriptions of each region at the end of 2019 are shown in Fig. 2, Fig. 3 and Fig. 4. From Fig. 2, it is easy to see that the East region in China has a larger population than the West. Relatedly, the number of HSR stations and the “degree” (see the definition in Fig. 2) of these stations in the East are often larger than those in the West, suggesting that HSR stations are usually located in the regions with large populations. As can be seen from Fig. 3, the daily frequency of high-speed trains stopping in the East is also greater than that in the West. Further, the closer a region is to Wuhan, the greater number of stations and daily frequency of high-speed trains it usually has connected to Wuhan. Moreover, the regions with greater frequency of high-speed trains tend to have more stations and trains connected to Wuhan. Fig. 4 shows the average travel time from Wuhan to each province-level region on HSRN. It further indicates that the spatial distance from Wuhan to frontier regions is usually far away, even unreachable.

5. Results and discussion

To obtain the results, some parameters need to be preset. Firstly, the transfer time $t_{iw}$ for each HSRN transfer is set to 30 min when station $i$ is indirectly connected to Wuhan and it takes more than one train to get to station $i$ from Wuhan. Since the passenger flow data of HSR at each station or province-level region is difficult to acquire and there is an obviously positive correlation between passenger flow and train frequency for a given period of time, we use train frequency to represent passenger flow. According to the methods proposed in Section 3, the regional risk of COVID-19 infection from Wuhan can be calculated by obtaining the probability and impact of the infection.

5.1. Probability and impact of infection

With Wuhan being its capital, Hubei province has the highest probability of COVID-19 infection from Wuhan, as shown in Fig. 5 and Fig. 6a. This is related to the fact that Wuhan has a strong correlation with the prefecture-level cities of Hubei province on the CHSRN. Although Henan, the neighbor province of Hubei, has a larger scope of correlation with Wuhan than that of Hubei, its probability of the infection is smaller than that of Hubei, due to its smaller intensity of correlation with Wuhan. From Figs. 5 and 6a, we find that the neighbor provinces of Hubei, such as Henan, Hunan, Jiangxi and Anhui, also have a high probability of infection. In addition, Jiangsu, Guangdong, Hebei,
Zhejiang, and Beijing are ranked in the top 10 in probability of infection. This is because these provinces are covered by the busier Beijing-Shanghai HSR line and Beijing-Hong Kong HSR line, which are part of the “four vertical and four horizontal” corridors and the (expanded) “eight vertical and eight horizontal” corridors of CHSRN (NDRC, 2004, 2016). Furthermore, these lines have the largest passenger flow and play key roles in the CHSRN by bridging the clusters of economy around the Jing-Jin-Ji (Beijing, Tianjin, Hebei) Economic Zone, the Yangtze River Economic Zone, and the Pearl River Economic Zone, which are the three dominant economic zones in China.

In addition, the frontier regions always have a low probability of infection. For example, as illustrated in Fig. 5, Tibet has no HSR and Xinjiang has few stations that are only indirectly connected to Wuhan, resulting in a low probability of infections. The East region’s probability of infection is commonly larger than that of the West. This is a result of more stations and high-speed trains connected to Wuhan in the East, which is not only driven by the greater demand for human mobility and economic activity, but also determined by Wuhan’s hub role in CHSRN.

We turn now to the impact of infection from Wuhan. As can be seen from Fig. 7 and Fig. 6b, there is an obvious difference among the impacts of COVID-19 infection from Wuhan on province-level regions. Guangdong has the largest population and the highest development state of HSRN of all the regions, which means that if Guangdong is infected with COVID-19 from Wuhan, it will be most seriously impacted due to the large population and high human mobility. Therefore, Guangdong has the largest infection impact of all the province-regions. Although the HSRN development state in Hubei is high, its population is relatively moderate, which results in a relatively small impact of infection on Hubei. Because the population and development state of HSRN in the frontier regions is lower than those of the other regions, the impact of infection on these regions is always smaller.

By comparing Figs. 5 and 7, we find that although the infection probability distribution is different from the impact distribution, there are still some similarities between the two. In other words, the regions with high probability of infection are mainly distributed around Wuhan and along two crucial HSR lines while the high-impacted regions are always distributed in the populous and developed areas. Nevertheless, the regions distributed along the two HSR lines not only are highly probable to be infected, but also may be seriously affected by the infection. Henan, Hunan, Jiangsu, and Guangdong are all listed in the top 5 for both probability and impact, illustrated in Fig. 6.

5.2. Regional risk of infection

As shown in Fig. 8, the regions with high risk of COVID-19 infection from Wuhan are mainly distributed along the southern half of the Beijing-Hong Kong HSR line, including Hubei, Guangdong, Henan, and Hunan. These regions not only have a strong correlation with Wuhan on CHSRN, especially through the Beijing-Hong Kong HSR line, but also have a large population and well developed HSRN. Fig. 9a demonstrates that although the impact of infection on Hubei is relatively small, its probability of infection is very high, resulting in its high risk of infection.
However, Guangdong has the opposite scenario, in which the impact of infection is larger than that of any other regions, but its probability of infection is relatively small. Some regions, such as Hainan, Tibet and Ningxia, that have no correlation with Wuhan or have no HSR (illustrated in Figs. 2 and 3), are less probable to be infected with COVID-19 from Wuhan, so their risk of infection is the smallest of all the regions. The risk of each region varies greatly, as for example the risk of Hubei is about 0.16 while that of Liaoning is approximately equal to 0 due to its low probability of infection.

Furthermore, we illustrate the cumulative number of confirmed cases in each province-level regions during the initial period of COVID-19 transmission from Wuhan, in order to compare the risk among the regions. The risk of each region varies greatly, as for example the risk of Hubei is about 0.16 while that of Liaoning is approximately equal to 0 due to its low probability of infection.

The cumulative number of confirmed cases in Hubei province excludes that in Wuhan.

Then, the distribution of cumulative number of confirmed cases over infection risk in different periods after the date of Wuhan travel ban is shown in Fig. 10. The size of each red dot indicates the cumulative number of confirmed cases in each province-level region, and the background color of each province-level region indicates its risk of infection from Wuhan. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

---

5 According to Guan et al. (2020), the incubation period is defined as the duration from the contact of the transmission source to the onset of symptoms.

6 The travel ban was implemented in Wuhan on 23 January 2020 and the purpose of this study is to assess the regional risk of COVID-19 infection from Wuhan at the early stage of the epidemic spread. Before that date, few regions reported the number of confirmed cases. Thus, January 23, 2020 is assumed to be the critical date and the initial date. We thus select 0 days, 3 days, 7 days, 14 days, and 24 days as the critical incubation periods, respectively: that is, January 23 (0 days from the date of Wuhan travel ban), January 26 (3 days), January 30 (7 days), February 6 (14 days), and February 16 (24 days), respectively.
above), while the cumulative number of the frontier regions is always the cumulative number of confirmed cases in each region at the early days of the confirmed cases of COVID-19 for five different periods during the early infection is always higher. 

cumulative confirmed cases in Southeast China, where the risk of smaller than that of other regions. There is always a larger number of cumulative confirmed cases in Hubei is always the largest of all regions (owing to its highest infection risk as discussed above), while the cumulative number of the frontier regions is always smaller than that of other regions. There is always a larger number of cumulative confirmed cases in Southeast China, where the risk of infection is always higher.

From Fig. 10 we see a significant correlation between the risk of infection and the cumulative number of confirmed cases, with the correlation coefficient being around 0.6 (all statistically significant at the 0.01 level) as shown in Table 1. Furthermore, the correlation is highest for the case of 3 days post the travel ban, and it becomes weaker and weaker as the post-ban days increase. This makes sense: as the days move further away from the Wuhan ban, factors other than the initial transportation networks can exert their influences, such as government policy in each region and the success of regional confinement efforts. For example, the top-10 regions at risk of infection tend to have a larger number of cumulative confirmed cases, illustrated in Fig. 9b. These suggest that the cumulative number of confirmed cases may be a consequence of the infection risk, and that the regional risk of infection we have assessed may serve as a good predictor for the number of cumulative confirmed cases exported from Wuhan, especially at the early days of the epidemic spread.

### Table 1

| Cumulative Confirmed Cases | January 23, 2020 | January 26, 2020 | January 30, 2020 | February 06, 2020 | February 16, 2020 |
|----------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| $R_3$                      | 0.587**         | 0.668**         | 0.622**         | 0.663**         | 0.587**         |

** Correlation is statistically significant at the 0.01 level.

(illustrated in Fig. 10a), due to lack of awareness of COVID-19 and lack of detection material. After a few days, as detection technology and material were guaranteed, more confirmed cases were detected. Obviously, the cumulative number of confirmed cases in Hubei is always the largest of all regions (owing to its highest infection risk as discussed above), while the cumulative number of the frontier regions is always smaller than that of other regions. There is always a larger number of cumulative confirmed cases in Southeast China, where the risk of infection is always higher.

From Fig. 10 we see a significant correlation between the risk of infection and the cumulative number of confirmed cases, with the correlation coefficient being around 0.6 (all statistically significant at the 0.01 level) as shown in Table 1. Furthermore, the correlation is highest for the case of 3 days post the travel ban, and it becomes weaker and weaker as the post-ban days increase. This makes sense: as the days move further away from the Wuhan ban, factors other than the initial transportation networks can exert their influences, such as government policy in each region and the success of regional confinement efforts. For example, the top-10 regions at risk of infection tend to have a larger number of cumulative confirmed cases, illustrated in Fig. 9b. These suggest that the cumulative number of confirmed cases may be a consequence of the infection risk, and that the regional risk of infection we have assessed may serve as a good predictor for the number of cumulative confirmed cases exported from Wuhan, especially at the early days of the epidemic spread.

### 6. Conclusion

Based on our probabilistic risk model, this study has proposed a method to assess regional risks of COVID-19 infection from Wuhan via HSR during the initial transmission period. By describing HSRN at the scale of a prefecture level, the development state of HSRN in each province-level region and the correlation between each region and Wuhan on HSRN were identified and further defined as the risk factors. In other words, we regarded that the probability of COVID-19 infection from Wuhan to a region is related to the correlation between the region and Wuhan on HSRN, and that the impact of the infection in a region is related to its population and human mobility measured by the development state of HSRN. With HSR operation data, demographic data, and confirmed case data of COVID-19, the risk of COVID-19 infection from Wuhan to each province-level region in mainland China was then obtained. It was further compared with the cumulative number of confirmed cases of COVID-19 for five different periods during the early stage of the epidemic spread, so as to validate the proposed method.

We note that while the probabilistic risk model developed in this paper is for assessing the regional risk of COVID-19 infection from Wuhan, the proposed method would be applicable if the outbreak were from another domestic city. In addition, we have obtained the following results and managerial and policy implications. First, the probability distribution of infection is different from the impact distribution of infection. The high-probability regions are mainly distributed around Wuhan and along the Beijing-Shanghai HSR line and the Beijing-Hong Kong HSR line, while the high-impacted regions are mainly distributed in the populous and developed areas. Therefore, the administrators may take different response measures to prepare for the pandemic according to the specific conditions of each region. Specifically, the high-probability and high-impacted regions, as such the area along the two HSR lines, require more attention to control measures. These regions also should be heavily monitored at the long regular prevention and control stage of the post COVID-19.

Second, the critical factors for infection risk are different for different regions. For example, for Hubei, the probability of infection is the critical risk factor, leading to high infection risk, while for Guangdong, the impact of infection is the critical risk factor. The Hubei government implemented a travel ban in Wuhan so as to cut off the correlation of the other areas of the province with Wuhan, thereby reducing the probability of infection. Thus, the administrator of each region should focus on its own critical risk factors when mitigating the risk of infection, which is very significant during the regular control stage of the post COVID-19. Otherwise, it will lead to excessive consumption of resources, resulting in hugely socioeconomic costs.

Third, the high-risk regions are mainly distributed along the southern half of the Beijing-Hong Kong HSR line. These regions not only have a strong correlation with Wuhan, but also have a large population and well developed HSRN. Thus, the central government should pay more attention to these regions when allocating emergency resources and preparing emergency response plans. In fact, the Beijing-Hong Kong HSR line was basically shut down during the initial period of COVID-19 transmission from Wuhan. Furthermore, the central government has deployed a large number of medical staff and material to Hubei.

Fourthly, the high-risk regions consistently have a large number of cumulative confirmed cases of COVID-19. The cumulative number of confirmed cases exported from Wuhan may be a consequence of the infection risk. Based on this observation, in the early days of the outbreak administrators should apply the proposed method to assess the infection risk for each region, and then prioritize the allocation of emergency resources and the plan of emergency responses before the extensive spread of the pandemic.

Finally, the method proposed in this paper would be applicable for other infectious diseases, and would be applicable at the regular prevention stage of the post COVID-19 world. If a new pandemic breaks out in a region, the proposed method can quickly forecast the infection risk of other regions. Especially, for the post COVID-19 world, the proposed method can predict the regional risk under the sporadic outbreaks of the pandemic. Once the pandemic rebounds, this will help the administrators quickly prepare for long regular prevention strategies. Besides, our method and findings may be useful for the countries where the pandemic is severe and HSRN is developed, such as Japan, Germany, and Italy. Even if there is no developed HSRN in a region, the proposed method can also be extended to combine with the main transportation network of the region.

The paper has also raised a number of other issues and avenues for future research. First, since passenger flow data of HSR is not available, we use train frequency to indicate passenger flow. This may have some impact on the results we obtained here, which can be further verified if more detailed data sets are available. Second, in this study we only consider HSR and ignore other transportation models, such as aviation, highway, waterway, and shipping, which can also affect the spread of COVID-19 from Wuhan. Therefore, assessing the regional risk of COVID-19 infection from Wuhan via a multi-modal transportation may be an interesting research direction. Third, the impact of COVID-19 transmission within a region is not excluded in this research because the population and the development state of HSRN were selected as the indicators of impact of infection. Future research can investigate this aspect in a more comprehensive manner. Finally, the method proposed in this paper is only applied to the situation in mainland China. It may be applied to analyze the issue in other countries in future research for comparative analysis, which may yield further interesting results.
Declaration of competing interest

There is no conflict of interest exiting in the manuscript which is approved by all authors for publication. And we would like to declare that all or part of the work described in the manuscript is original research, which has not been published elsewhere and is not prepared for publication elsewhere.

Acknowledgements

This study was funded by the National Natural Science Foundation of China under the Grants 71871639, and 71871042, 71901047, 71421001. The authors are grateful to the Guest Editor and reviewers for their constructive comments. We would like to thank Kun Wang and Hangjun Yang from University of International Business and Economics, Jiaoe Wang from Chinese Academy of Sciences, for their kind help and useful comments. We also thank VarioFlight for data support, and Peng Zhao, an engineer of Zhengzhou high-speed train section of Zhengzhou bureau group of China Railway, for his helpful discussions.

References

ABD (Asian Development Bank), 2020. Developing Asia Growth to Fall in 2020 on COVID-19 Impact. https://www.adb.org/news/developing-asia-growth-fall-2020-covid-19-impact

Andersen, S.C., Mulberry, N., Edwards, A.M., Stockdale, J.E., Iyaniwura, S.A., Falcao, R., Cui, Q., He, L., Liu, Y., Zheng, Y., Wei, W., Yang, B., Zhou, M., 2021. The impacts of COVID-19 on household travel and activities in Australia – the early days under restrictions. Transport Pol. 96, 76–93.

Bian, Z., Zuo, F., Gao, J., Chen, Y., Venkata, S.S.C.P., Bernardes, S.D., et al., 2021. Time lag effects of COVID-19 policies on transportation systems: a comparative study of New York City and Beijing. Transport. Res. Pol. Pract. 145, 269–283.

Bogoch, I.I., Creasey, P., Brownstein, J.S., Lipkin, W.I., Nicholson, W.S., et al., 2020. The transmission dynamics of COVID-19 and implications for containment. Nat. Med. 26 (4), 506–514.

Bogoch, I.I., Creatore, M.I., Bernardes, S.D., et al., 2021. Time lag effects of COVID-19 policies on transportation systems: a comparative study of New York City and Beijing. Transport. Res. Pol. Pract. 145, 269–283.

Bogoch, I.I., Creasey, P., Brownstein, J.S., Lipkin, W.I., Nicholson, W.S., et al., 2020. The transmission dynamics of COVID-19 and implications for containment. Nat. Med. 26 (4), 506–514.

Bogoch, I.I., Creatore, M.I., Bernardes, S.D., et al., 2021. Time lag effects of COVID-19 policies on transportation systems: a comparative study of New York City and Beijing. Transport. Res. Pol. Pract. 145, 269–283.

Bogoch, I.I., Creasey, P., Brownstein, J.S., Lipkin, W.I., Nicholson, W.S., et al., 2020. The transmission dynamics of COVID-19 and implications for containment. Nat. Med. 26 (4), 506–514.

Bogoch, I.I., Creatore, M.I., Bernardes, S.D., et al., 2021. Time lag effects of COVID-19 policies on transportation systems: a comparative study of New York City and Beijing. Transport. Res. Pol. Pract. 145, 269–283.

Bogoch, I.I., Creasey, P., Brownstein, J.S., Lipkin, W.I., Nicholson, W.S., et al., 2020. The transmission dynamics of COVID-19 and implications for containment. Nat. Med. 26 (4), 506–514.

Bogoch, I.I., Creatore, M.I., Bernardes, S.D., et al., 2021. Time lag effects of COVID-19 policies on transportation systems: a comparative study of New York City and Beijing. Transport. Res. Pol. Pract. 145, 269–283.

Bogoch, I.I., Creasey, P., Brownstein, J.S., Lipkin, W.I., Nicholson, W.S., et al., 2020. The transmission dynamics of COVID-19 and implications for containment. Nat. Med. 26 (4), 506–514.

Bogoch, I.I., Creatore, M.I., Bernardes, S.D., et al., 2021. Time lag effects of COVID-19 policies on transportation systems: a comparative study of New York City and Beijing. Transport. Res. Pol. Pract. 145, 269–283.

Bogoch, I.I., Creatore, M.I., Bernardes, S.D., et al., 2021. Time lag effects of COVID-19 policies on transportation systems: a comparative study of New York City and Beijing. Transport. Res. Pol. Pract. 145, 269–283.

Bogoch, I.I., Creatore, M.I., Bernardes, S.D., et al., 2021. Time lag effects of COVID-19 policies on transportation systems: a comparative study of New York City and Beijing. Transport. Res. Pol. Pract. 145, 269–283.

Bogoch, I.I., Creatore, M.I., Bernardes, S.D., et al., 2021. Time lag effects of COVID-19 policies on transportation systems: a comparative study of New York City and Beijing. Transport. Res. Pol. Pract. 145, 269–283.
Shankar, R., Choudhary, D., Jharkharia, S., 2018. An integrated risk assessment model: a case of sustainable freight transportation systems. Transport. Res. Transport Environ. 63, 662–676.

State Council of the People’s Republic of China, 2017. Modern Integrated Transport System Development 13th Five-Year Plan. http://www.gov.cn/zhengce/content/2017-02/28/content_5171345.htm.

Sun, X., Wandelt, S., Zhang, A., 2020a. How did COVID-19 impact air transportation? A first peek through the lens of complex networks. J. Air Transport. Manag. 89, 101928.

Sun, X., Wandelt, S., Zhang, A., 2021a. On the degree of synchronization between air transport connectivity and COVID-19 cases at worldwide level. Transport Pol. 105, 115–123.

Sun, X., Wandelt, S., Zheng, C., Zhang, A., 2021b. COVID-19 pandemic and air transportation: successfully navigating the paper hurricane. J. Air Transport. Manag. (forthcoming).

Wang, J., Huang, J., Jing, Y., 2020. Competition between high-speed trains and air travel in China: from a spatial to spatiotemporal perspective. Transport. Res. Pol. Pract. 133, 62–78.

Wang, K., Xia, W., Zhang, A., 2017. Should China further expand its high-speed rail network? Consider the low-cost carrier factor. Transport. Res. Pol. Pract. 100, 105–120.

Weible, C.M., Nohrstedt, D., Cairney, P., Carter, D.P., Crow, D.A., Durnová, A.P., et al., 2020. COVID-19 and the policy sciences: initial reactions and perspectives. Pol. Sci. 53 (2), 225–241.

Weisberg, S.P., Connors, T.J., Zhu, Y., Baldwin, M.R., Lin, W.H., Wontakal, S., et al., 2021. Distinct antibody responses to SARS-CoV-2 in children and adults across the COVID-19 clinical spectrum. Nat. Immunol. 22 (1), 25–31.

Wei, Y., Wang, J., Song, W., Xiu, C., Ma, L., Pei, T., 2021. Spread of COVID-19 in China: analysis from a city-based epidemic and mobility model. Cities 110, 103010.

WHO, 2020a. Coronavirus Disease (COVID-19) Situation Reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/

WHO, 2020b. WHO coronavirus (COVID-19) dashboard. Last update: 29 March 2021, 10:52pm GMT. https://covid19.who.int/.

WHO, 2020c. Key considerations for repatriation and quarantine of travellers in relation to the outbreak of novel coronavirus 2019-nCoV. https://www.who.int/news-room/articles-detail/key-considerations-for-repatriation-and-quarantine-of-travellers-in-relation-to-the-outbreak-of-novel-coronavirus-2019-ncov.

Wuhan Bureau of Statistics, 2020. Statistical Bulletin of Wuhan’s economic and social development 2019. http://tjj.wuhan.gov.cn/details.aspx?id=4615.