**An Evaluation Model of COVID-19 Spread Control and Prevention: Effectiveness Analysis Based on Immigration Population Data in China**

**Abstract**

**Background:** Compared to the peak data in early February 2020, the spread of Coronavirus (COVID-19) has been drastically slowed down and come under control in China, as reported by the end of February 2020. Meanwhile, the outcomes of control and prevention of COVID-19 varied among different regions (i.e. provinces and municipalities) in China; moreover, COVID-19 became a global pandemic and the spread of disease accelerated among other countries outside China.

**Objective:** This study aimed to establish valid models which will evaluate the effectiveness of COVID-19 control and prevention among various regions in China. These models also targeted regions with problems in control and prevention by issuing immediate warnings.

**Methods:**We built a mathematical model: the Epidemic Risk Time Series Model, based on analyzing two sets of data: first, the daily number of COVID-19 incidence (i.e., newly-diagnosed cases); second, the daily immigration population size.

**Results:**Based on the model’s evaluation result, some regions, such as Shanghai and Zhejiang, have been successful in COVID-19 control and prevention; whereas other regions yielded poor performance, such as Heilongjiang. The evaluation result was highly correlated with R0 value, and the result was evaluated within a timely manner at the beginning of disease outbreak.

**Conclusions:**The Epidemic Risk Time Series Model is designed to evaluate the effectiveness of COVID-19 epidemic control and prevention among different regions in China, based on an analysis of immigration population data. Compared to other methods, such as R0, this model issues early warnings in a timely manner. This model can be generalized and applied to other countries regarding evaluations of COVID-19 control and prevention.

Key words: COVID-19; 2019-nCoV; epidemic control and prevention; Epidemic Risk Time Series Model; incoming immigration population; new diagnoses per day

**Introduction**

The first case of COVID-19 was diagnosed in December 2019, in Wuhan, Hubei. Despite the spread of COVID-19, few prevention actions had been reinforced at the beginning of the disease outbreak in China. For instance, a celebration banquet with tens of thousands of people was held in Wuhan on January 18th, 2020. As a result, this event accelerated the spread of COVID-19 in that region [1]. Gradually, more prevention actions were taken, including the investigation and control of incoming immigration populations from other regions; closing down some densely populated regions; and wearing face masks in public [2, 3].

Besides the traditional methods for COVID-19 prevention and control, the supplemental measures are also considered necessary, particularly to deal with people with no symptoms but who can be infectious during the incubation period [4]. Specifically, as a screening mechanism, taking temperature before people’s entrance to public areas can only detect part of the COVID-19 cases [4].

Given the recent pandemic development, limited studies have utilized the COVID-19-related data to investigate the effective of COVID-19 control and prevention [5]. Some other studies also collected media reports regarding COVID-19, examining the role that media played in this current epidemic in China [6]. Similarly, researchers used to investigate norovirus epidemics via internet surveillance and to build a model to predict the potential disease infections in China [7].

We are able to estimate the effectiveness of epidemic prevention and control from statistical data, such as the number of daily new diagnosed patients in the provinces or municipalities in China (NEW) [8, 9, 10]. However, this method does not evaluate the effectiveness of prevention and control in the regions (including provinces or municipalities) of China (REGIONS), because the newly diagnosed cases data (NEW) is not analyzed in combination with the information of immigration population (REGIONS) during the outbreak. For instance, when comparing two provinces A and B that with the same NEW during the outbreak period: the new cases in Province A were mainly immigrating from outside of the province and most of the cases were confirmed on the day of entrance; whereas Province B mainly consisted of local residents, and most of incoming cases were confirmed one week after their entrance. All confirmed cases in both of the Province A and B have been quarantined until being diagnosed. Therefore, we considered that the epidemic prevention and control in Province A was more effective than that in Province B, because the virus spread more severely in Province B, despite its lower share of immigrating residents.

The Chinese Government has been emphasizing the analysis of big data, especially immigration population data, in COVID-19 prevention and control since Mid-February [11, 12]. Immigration population data analysis was an approach in disease prevention. Particularly, the Health Code was created [13], and it was merged with various REGIONS [14, 15, 16, 17]. The Health Code is a mobile application to detect individuals’ prior travel histories, such as the epidemic zone before entering a public area. Hence, in order to detect the infected individuals prior to their entrance to the public areas, it is more effective to combine this mobile application with measuring body temperature.

Utilizing immigration population data from Baidu, there are several analysis reports on the trend of population movement during the COVID-19 [18, 19]. However, at present, very few COVID-19 control and prevention studies have used the dataset of daily incoming immigration population in each REGION (POPULATION).

In this study, we analyzed the immigration population data and to evaluate the risk of POPULATION in each REGION (RISK). The RISK presents a similar indication as to the Health Code, which evaluates the immigration risk from the relevant data source. Moreover, we built an Epidemic Risk Time Series Model to evaluate the effectiveness of COVID-19 control and prevention across different REGIONS. Using this evaluation, REGIONS with poor prevention performance can be detected as soon as possible.

**Methods**

**Overview**

In the Epidemic Risk Time Series Model, the two decision variables, OFFSET and WINDOW, are used to reveal the delayed days of RISK converting to NEW. Longer days means less effectiveness in disease control and prevention. The model workflow is shown in Figure 1. According to this model, there are three major steps to evaluate a REGION in a period of days. Specifically, first, RISK data is constructed from POPULATION and NEW data; second, RISK data is processed into PROCESSED RISK data by OFFSET and WINDOW; last, the OFFSET and WINDOW values yield the highest correlation coefficient of NEW and PROCESSED RISK data are chosen as the output of the model.

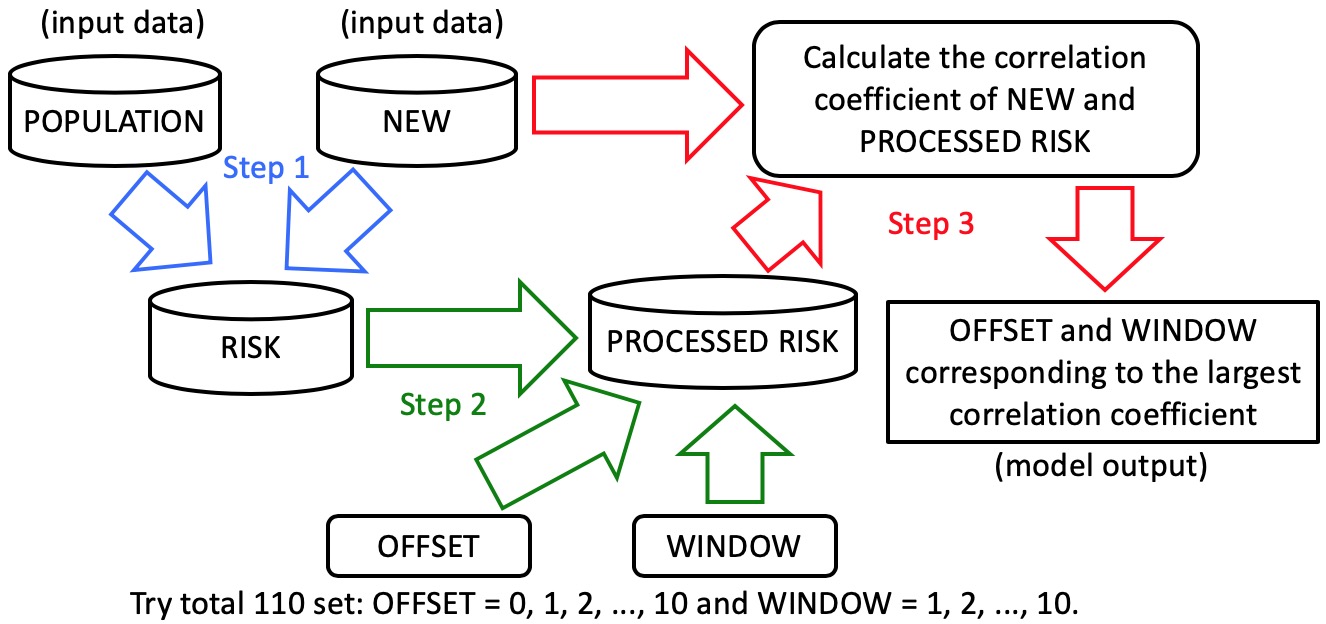


Figure 1. The Epidemic Risk Time Series Model with Three Steps

**Data source**

***Model Input Data 1: NEW.*** Since January 17th, 2020, various REGIONS have released data such as NEW. The NEW data is crawled from [20].

***Model Input Data 2: POPULATION.*** To detect ongoing trends of the COVID-19 epidemic, the daily incoming immigration population data, which are distinguished from different source REGIONS, is crawled from [21]. Since there is no data source regarding immigration population data in REGIONS such as Hong Kong, Macao, or Taiwan, and the inter-state traffic from Hubei has been shut down since late January, these REGIONS are excluded from this analysis. However, the immigration population emigrating from Hubei to other REGIONS which were included in this study. All statistical analyses were conducted using Python, version 3.7 (Copyright 2001-2020. Python Software Foundation).

**Analytical methods**

***Calculate RISK*.** The incoming immigration populations with the same size exposed to different factors were at the different levels of risks of contracting COVID-19. For example, individuals with early residence in Hubei, during the spread of the COVID-19 virus, experienced higher risks of being infected, even though they are included with the same immigration population size. Hence, POPULATION is processed as Formula 1, and the RISK data is constructed.

Formula 1. Calculation of RISK

In the Formula 1, all the values of RISK, POPULATION and ACCUMULATED NEW are in a same day. The RISKi is the daily immigration risk of REGION i in one day. i can be 1, 2, 3, …, n. “n” is a fix number. In this study “n” is 31, because this study analyzed 31 REGIONS including Hubei. The “i” in this study cannot be the number of Hubei for the reason mentioned. The POPULATIONj is the POPULATION from source REGION j. j can be 1, 2, 3, …, n, and j cannot be same as i. The ACCUMULATED\_NEWj is the sum of NEW in immigration source REGION j in recent 3 days (ACCUMULATED NEW), and its calculation is as Formula 2. The ACCUMULATED\_NEWd is the ACCUMULATED NEW on the date of “d”.

Formula 2. Calculation of ACCUMULATED NEW

***OFFSET.*** Offset is used to evaluate the control for incoming immigration population. Among the incoming immigration population, the disease control and prevention are varied at different times or in different REGIONS. Specifically, some REGIONS have taken strict screening mechanism, such as to measure temperature and to examine cough symptoms, to detect infected immigrants and to reinforce guarantee immediately. In this case, NEW increased simultaneously with the sudden increase of RISK on the same day. Whereas the infected individuals will be diagnosed and confirmed related late when they have been infected before entering the REGION. The OFFSET is the number of days that RISK is shifted. For example, when OFFSET equals 3, it means that the RISK of each day is processed as the RISK of 3 days ago.

***WINDOW.*** Window is used to evaluate the control for domestic/local people.Meanwhile, the control and spread among the local people as well as their awareness of prevention will affect the spread of the epidemic. In some REGIONS, the immigrants are strictly home-quarantined for 14 days [22]. Such rigorous measures prevented potentially infected people from spreading the virus when entering the REGION.

According this model, hypothetically, when only deals with externally infected individuals, there will only include the OFFSET. On the other hand, other conditions may contribute to the spread of COVID-19 and prolonged impact on the RISK. For instance, when the infected individual who has traveled into the REGION, being sick or incubating, did not seek immediate medical treatment; or local people with poor disease awareness, such as not wearing a face mask in public area. Therefore, the concept of window is introduced into the model. For example, when WINDOW equals to 10, it means that the total RISK of 10 consecutive days will affect the NEW on the 10th day. Moreover, the incubation period, with 95% confidence interval, is between 4.1 and 7.0 days. Hence, if the infected person who entered the REGION from 10 days ago, is still able to affect the REGION, by spreading the disease from person to person [23].

***Process RISK by OFFSET and WINDOW.*** RISK can be processed by OFFSET and WINDOW as in Formula 3.

Formula 3. Calculation of PROCESSED RISK

In Formula 3, all the PROCESSED RISK and RISK are for a same REGION. The PROCESSED\_RISKd is the value of PROCESSED RISK by OFFSET and WINDOW on the date of “d”. The RISKd-w-OFFSET is the value of RISK at the date of “d-w- OFFSET”. Specifically, if it is necessary to calculate the value of PROCESSED RISK in 02/11/2020, when OFFSET equals to 3, WINDOW is 2. The formula is as following:

Formula 4. Calculation of PROCESSED RISK (d=02/11/2020, offset=3, window=2)

When OFFSET equals to 0, WINDOW is 1. The PROCESSED\_RISKd is just RISKd without any process:

Formula 5. Calculation of PROCESSED RISK (offset=0, window=1)

***Correlation Coefficient between NEW and PROCESSED RISK, and Model Outputs.*** The final step of this model is to find a set of OFFSET and WINDOW as the best fit for NEW and PROCESSED RISK of each REGION on a daily-basis.

For each REGION on a daily basis, starting from January 17th, 2020, which is the first day of NEW data collection, the OFFSET is calculated from zero to 10, and the WINDOW is calculated from one to 10. There are 110 different OFFSET and WINDOW sets. The 110 sets are used to process RISK accordingly, to calculate the 110 correlation coefficients with NEW and PROCESSED RISK. Finally, the set of OFFSET and WINDOW corresponding to the maximum correlation coefficient (CORR) is the model output for the REGION on that day.

**Results**

**Processing POPULATION and NEW into RISK**

Based on Formula 2, ACCUMULATED NEW is processed from NEW. Particularly, the process of Hubei in the first six days is shown as Table 1. Accurate data has been released since January 17, the values before that day are set to zero. Similar calculations are performed in other 30 REGIONS on a daily basis.

|  |  |  |
| --- | --- | --- |
| Date | NEW in Hubei | ACCUMULATED NEW in Hubei |
| 01/17/2020 | 17 | 17 |
| 01/18/2020 | 59 | 76 |
| 01/19/2020 | 77 | 153 |
| 01/20/2020 | 72 | 208 |
| 01/21/2020 | 105 | 254 |
| 01/22/2020 | 69 | 246 |

Table 1. The ACCUMULATED NEW in Hubei Province

Based on Formula 1, RISK is processed from POPULATION and ACCUMULATED NEW. For example, the total POPULATION travelling into Jiangsu and Heilongjiang Province and the total ACCUMULATED NEW of source REGIONS at daily basis is compared with their RISK in Figure 2 and 3. Meanwhile, according to Formula 1, there are 30 incoming POPULATION and ACCUMULATED NEW for every targeted REGION. To avoid too many polylines plotted in the chart, the total POPULATION and ACCUMULATED NEW polylines are plotted.

Moreover, we have only analyzed the correlation among the three variables: POPULATION, ACCUMULATED NEW, RISK within the same region, so we have merged the effects and set the range of the three lines to zero and one [0, 1].

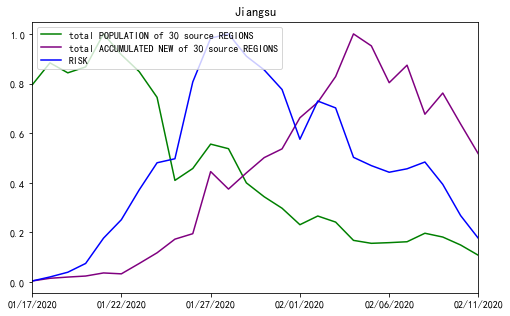


Figure 2. RISK of Jiangsu Province from January 17, 2020 to February 11, 2020

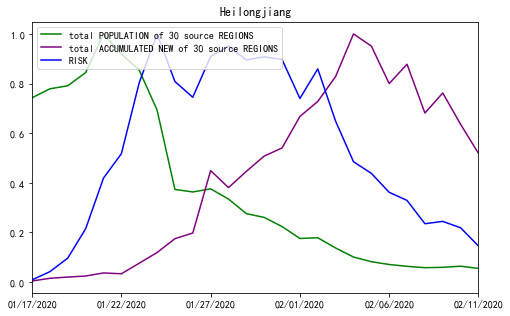


Figure 3. RISK of Heilongjiang Province from January 17, 2020 to February 11, 2020

**PROCESSED RISK and CORR**

In each REGION, 110 sets of OFFSET and WINDOW were used to generate RISK on a daily basis. Due to the large amount in data, NEW, RISK and the PROCESSED RISK processed by the model outputs in Jiangsu and Heilongjiang from January 17 to February 11 are used to illustrate the role of OFFSET and WINDOW parameters by line charts. As illustrated in Figures 4 and 5, only the absolute values of the NEW and RISK are collected from the same REGION to calculate for the relative indexes. Hence, we define the range of variable values to be between zero and one [0, 1]. The correlation coefficients between NEW and RISK of Jiangsu and Heilongjiang are 0.684 and -0.014. The value of Jiangsu is not high, and the one of Heilongjiang is nearly uncorrelated [Figure 4 and Figure 5]. If we used the PROCESSED RISK instead of RISK to draw the polyline chart, the polylines are much more fitted as illustrated in Figure 6 and 7. The correlation coefficient values increase to 0.979 and 0.874 respectively [Figure 6 and Figure 7].

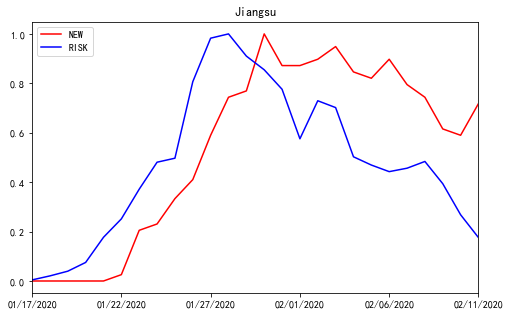


Figure 4. The polyline chart of NEW and RISK for Jiangsu Province from January 17, 2020 to February 11, 2020

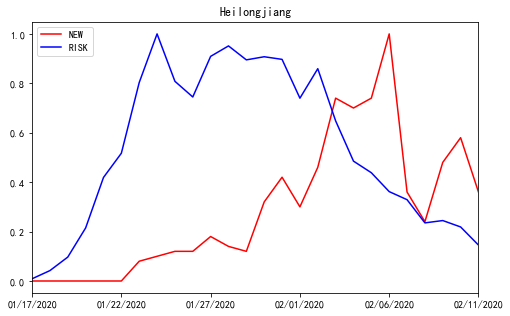


Figure 5. The polyline chart of NEW and RISK for Heilongjiang Province from January 17, 2020 to February 11, 2020

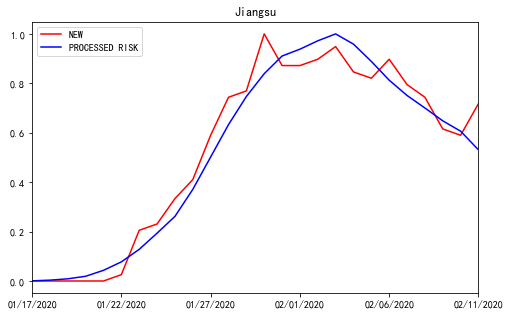


Figure 6. The polyline chart of NEW and PROCESSED RISK for Jiangsu Province from January 17, 2020 to February 11, 2020, when offset=0, and window=9

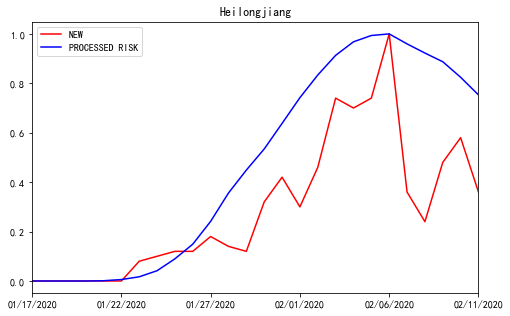


Figure 7. The polyline chart of NEW and PROCESSED RISK for Heilongjiang Province from January 17, 2020 to February 11, 2020, when offset=4, and window=10

As illustrated by Figures 4, 5, 6 and 7, the OFFSET and WINDOW reveal the delayed days before RISK converting to NEW. In theory, if all infected individuals entering into the REGION can be immediately detected and quarantined, the polylines of NEW and RISK will be fully fitted. Also under this condition, the value of OFFSET will be zero, WINDOW will be one, and the CORR will be one. On the other hand, if the infected people entering the REGION are not detected promptly, and spread the virus after entering, the RISK would impact the NEW in the next few days. The delayed days are evaluated by the values of OFFSET and WINDOW.

**Model Output**

The original size of the dataset is quite large, so we only included the sample result from every three days between January 21, 2020 and February 11, 2020 from 11 REGIONS, which were compared to the actual data released from news reports in Table 2. The study period is chosen based on the severity of the COVID-19 spread in China: from the early spread of COVID-19 from Hubei Province to other REGIONS in the country to the NEW is gradually decreasing in most of the REGIONS. The complete outputs are included as appendix to this paper.

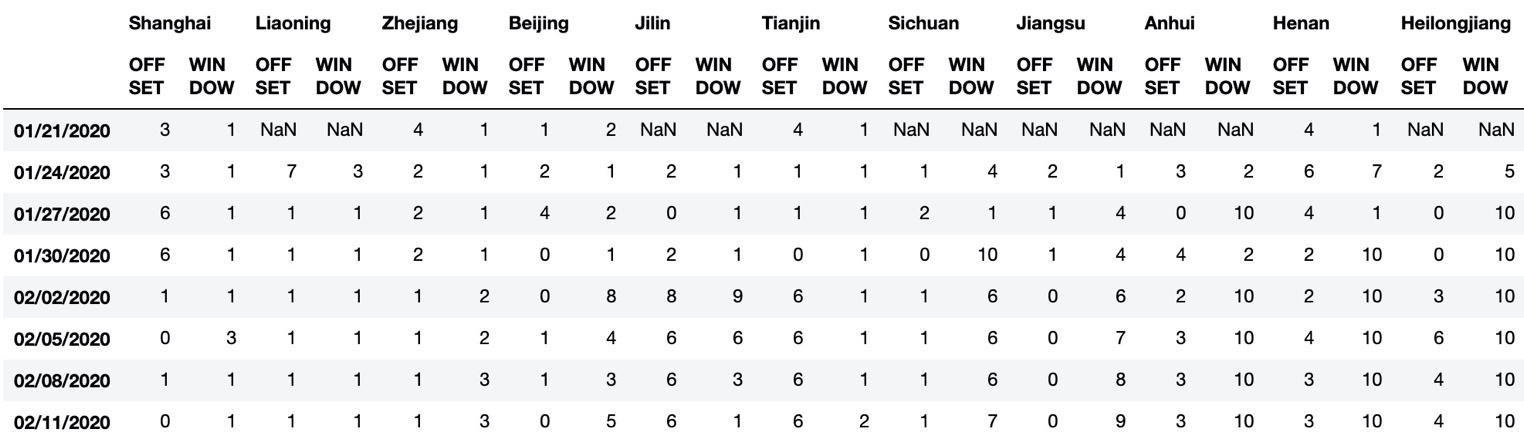


Table 2. OFFSET and WINDOW from 11 REGIONS between January 17, 2020 and February 11, 2020

Table 2 can be used to evaluate the effectiveness of the COVID-19 control and prevention from each REGION on a daily basis. The “NaN” values indicate no confirmed cases in REGIONS during the study period. REGIONS are sorted by the values of OFFSET + WINDOW in February 11, 2020 in an ascending order, which also indicates the sorting order of control and prevention effectiveness. Based on the evaluation result, Shanghai shows the lowest OFFSET and WINDOW values among 11 REGIONS, which indicates the most effectiveness in COVID-19 control and prevention. Contrarily, Heilongjiang is the least effective in COVID-19 control and prevention.

**Related News Reports Confirmed the Model Outputs**

There are limited data that has been released and can be compared to the effectiveness of disease control and prevention in different REGIONS. Regardless, we were able to collect data and news from 11 REGIONS to compare and confirm the model outputs.

First, according to the data released by Doctor Lilac Network, until February 11, 2020, the cumulative confirmed cases are grouped by incoming immigrants and local residents, which are from three REGIONS: Shanghai, Beijing, and Tianjin [Table 3].

We then compared the “cumulative confirmed cases” with the OFFSET and WINDOW values in Table 2. Shanghai, generated the lowest OFFSET + WINDOW value, performed best in COVID-19 control and prevention; and the local residents’ infection rate of Shanghai is the lowest among the REGIONS. Beijing ranked the second in performance evaluation; Tianjin demonstrated the highest OFFSET + WINDOW value, so it ranked the lowest in performance [Table 2].

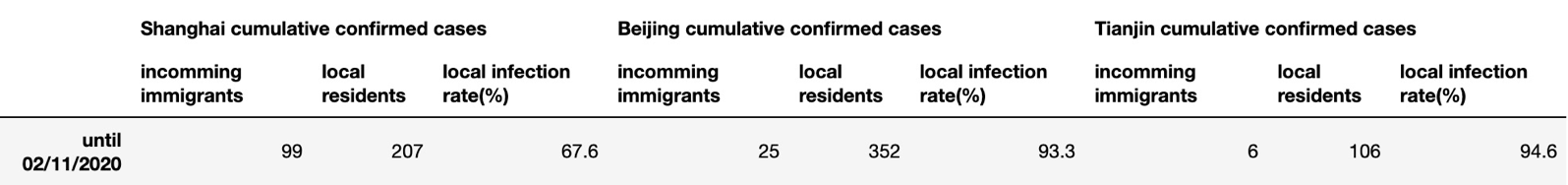


Table 3. The Cumulative Confirmed Cases by Incoming Immigrants and Local Residents

Second, R0 data of Shanghai, Zhejiang, Jiangsu, Anhui, Henan, and Sichuan were collected [24]. The R0 values on February 10, 2020, are shown as Table 4. Compared with Table 2, the relative values and ranking of R0 and OFFSET + WINDOW during the time around February 10, 2020, is nearly identical.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Shanghai | Zhejiang | Sichuan | Jiangsu | Henan | Anhui |
| 0.46 | 0.52 | 0.81 | 0.82 | 0.75 | 0.98 |

Table 4. R0 Values on February 10, 2020

The model output has been confirmed by related news as following: in late January, there were a large group of infected businessmen who were returning to Wenzhou, Zhejiang from Wuhan, Hubei [25]. On February 1, 2020, the municipal government of Wenzhou, Zhejiang issued 25 control and prevention and control measures in a timely manner [26, 27]. On February 22, 2020, after the Wenzhou epidemic was completely under control, the Chinese government’s newspaper published an article strongly affirming the achievements of Wenzhou’s epidemic control and prevention [28]. With the outbreak of COVID-19 in Wenzhou City, its province, Zhejiang performed well in COVID-19 control and prevention. Our model confirmed this evaluation result by presenting a relatively low value of OFFSET and WINDOW in Zhejiang [Table 2].

In addition, according to survey data, Heilongjiang did not pay enough attention to the epidemic and had poor awareness of prevention [29]. This is also confirmed with our study results with OFFSET and WINDOW values [Table 2]. Particularly, this online survey was conducted on January 31, 2020, targeting 10,304 residents from three provinces in Northeastern China, including Heilongjiang, Jilin, and Liaoning. This survey examined people’s feeling towards “confident”, “alert”, and “scared” during the COVID-19 outbreak. The level of feeling is ranked between zero and five, with five being the “strongest” feeling. Based on this survey, Heilongjiang demonstrated the lowest level of awareness in disease control and prevention; while Liaoning demonstrated the highest level and Jilin ranked second in awareness [Table 5]. The survey result is also confirmed by our model [Table 2].

|  |  |  |  |
| --- | --- | --- | --- |
| REGION | confident | alert | scared |
| Heilongjiang | 4.1 | 3.8 | 2.1 |
| Jilin | 3.9 | 3.9 | 2.2 |
| Liaoning | 3.7 | 3.9 | 2.3 |

Table 5. Online Survey Results regarding Awareness of COVID-19 Control and Prevention from China’s Three Northeast Provinces

Using the hypothesis-testing approach stated above [30], the data from Table 2, Table3, Table 4, and Table 5 respectively, were tested based on our hypotheses. First, the correlation coefficient between the internal infection rate of the three REGIONS in Table 3 and the corresponding “OFFSET+WINDOW” value of these three REGIONS in Table 2 on February 11 was 0.9216, and the original hypothesis H0: the correlation between the local infection rate and the “OFFSET+WINDOW” value was not statistically significant. Whereas the alternative hypothesis Ha: the correlation coefficient between the local infection rate and the “OFFSET+WINDOW” value was calculated by obtaining a t-value of: 2.374, and the p-value was between 0.15 and 0.1 (according to the t-distribution table: t0.15,1 = 1.963, t0.1,1 = 3.078).

Then, the correlation coefficient between the R0 value of the six REGIONS in Table 4, and the corresponding “OFFSET+WINDOW” value of these 6 REGIONS in Table 2 on February 11 was 0.8787. The original assumption was made that the correlation between H0: R0 value and “OFFSET+WINDOW” value was not statistically significant; while the alternative assumption was made that Ha: R0 value and “OFFSET+WINDOW” value were not correlated, and the t-value was calculated to be: 3.682, then the p-value was between 0.025 and 0.01 (according to the t-distribution table: t0.025,4 = 2.776, t0.01,4 = 3.747).

Finally, for the three REGIONS in Table 5, the correlation coefficient between the “alert + scared – confident” value and the corresponding “OFFSET+WINDOW” value for these three REGIONS in Table 2 on February 11 was -73.32. Specifically, the size of the alert and scared values were correlated to the “alert”, so it was positive; whereas the “confidence” and “alert” were inversely correlated, so it was negative. In addition, we proposed the hypothesis that the correlation coefficient between the H0: “confident + alert – scared” value and the “OFFSET+WINDOW” value was not significant. The alternative hypothesis Ha was: R0 value and the “OFFSET+WINDOW” value were not correlated, the t-value was -73.32 and the p-value was between 0.005 and 0.01 (according to the t-distribution table: t0.005, 1 = 63.66, t0.001, 4 = 127.3).

In sum, based on the three p-values, the model results were highly correlated with the three datasets, thus, confirmed the validity of the model.

**Discussion**

In this paper, the effectiveness of COVID-19 outbreak control and prevention across China is evaluated by using population movement data between regions and new daily confirmed cases. Moreover, the comparison of the model output [Table 2] through the infectious rate within local residents [Table 3], R0 value [Table 4], people's vigilance survey [Table 5] also confirmed the correctness of the Epidemic Risk Time Series Model, which is to say, the regions evaluated by the model do better in control and prevention, the R0-value was smaller, infection rate of local residents was lower, and the people's vigilance toward the COVID-19 was stronger.

**Early Warning by the Epidemic Risk Time Series Model in Epidemic Control and Prevention**

According to Figure 5, the peak of NEW in Heilongjiang happened in February 6, 2020. The peak day of RISK in Heilongjiang is January 24th, which is 13 days prior to the peak day of NEW. Based on Table 3, the values of OFFSET and WINDOW in Heilongjiang were rising gradually from the first day. Therefore, the current daily incidence (newly-diagnosed cases) could be lower in Heilongjiang, if the control and prevention measures were stricter in Heilongjiang since the end of January, 2020.

Based on our model, the warning threshold should be triggered as “problematic” when the value of OFFSET + WINDOW is greater than or equal to 5 [Table 3]; when the combined value of OFFSET + WINDOW is greater than or equal to 10, the situation is considered as “serious”. The warning level may be affected by factors such as the incubation period. Hence, when this model is used to evaluate the effectiveness of control and prevention for other epidemics, the warning values should be modified accordingly.

**The Epidemic Risk Time Series Model vs. R0 Method**

Compared to the evaluation method by R0 [24], the Epidemic Risk Time Series Model was able to detect the “warning threshold” more promptly. For example, the first confirmed case in Heilongjiang was diagnosed on January 23, 2020. According to our model, the OFFSET + WINDOW values of Heilongjiang on that day were 6 and 7; the value of OFFSET + WINDOW had been increasing gradually since then [Table 3]. On the other hand, the method by R0 can only be used at least 5 days after the first confirmed cases in that REGION, which is the average incubation period [23].

**The Formula for Calculating RISK**

In Formula 2, the “recent 3 days” in ACCUMULATED NEWj is derived from the following considerations. Through testing, it is found that the fewer days used in calculation, the greater the CORR value will be generated from the later step of the model. The number of diagnoses after a long-term incubation period is difficult to reflect the current RISK from its original REGIONS. The NEW value may vary greatly on a daily basis. Moreover, the days of suspected cases converting into confirmed cases may vary by days. Therefore, “recent 3 days” is used to calculate RISK in this model.

On the other hand, Formula 1 needs to divide the total population of the source REGION before calculating the cumulative cases. The values of ACCUMULATED NEWj by the two source REGIONS are equal. Particularly, the people from a REGION with smaller population size has greater probability than the infected patients traveling to the destination REGION. The CORR values remain constant, whereas the values of OFFSET + WINDOW are increased to fit the similar CORR. Compared to the local residents, immigrating individuals are more likely to be infected with the virus. Hence, Formula 1 was used in calculating RISK.

**Conclusion**

In this study, a mathematical model is built, using the number of daily confirmed cases and daily immigration population size data. Moreover, the effectiveness of epidemic control and prevention are to be evaluated by OFFSET + WINDOW, which are the outputs of the model. The results show that OFFSET + WINDOW values may change daily with the control effective control and prevention. For the REGIONS with poor performance, warning systems have been triggered by the OFFSET + WINDOW values, two weeks prior to its peak days of cases. The Epidemic Risk Time Series Model will be more prompt, compared to the method by R0.

Although the POPULATION data may have different statistical units in other countries, we utilized the relative values of the POPULATION to calculate the correlation coefficient. Therefore, the model does not only apply to Chinese-specific data. Theoretically, the method of this study can be generalized to other countries to evaluate the effectiveness of COVID-19 control and prevention.

**References**

1. CCTV Network: “Banquet for ten thousand people” turns into “disastrous banquet”, and the history should not be repeated. CCTV Network. URL: <http://news.cctv.com/2020/02/06/ARTI7I0Kl8rF7E1ruR41iR5v200206.shtml> [accessed 2020-02-06]

2. Notice of the general office of the state council on extending the Spring Festival Holiday of 2020. Chinese Government. <http://www.gov.cn/zhengce/content/2020-01/27/content_5472352.htm>

3. Summary of closure information for tourist attractions and some venues across the country. Happy Chaobai River Community in Sohu.com. URL: <https://www.sohu.com/a/368851082_717402> [accessed 2020-01-25]

4. Lanjuan Li, Academician of Chinese academy of engineering: attention should be paid to infectivity of COVID-19 during incubation period. China News. URL: <http://www.chinanews.com/sh/2020/01-22/9067646.shtml> [accessed 2020-01-22]

5. Mengchun Gong, MD; Li Liu, MD; Xin Sun, MD; Yue Yang, Shuang Wang, Hong Zhu. Cloud-Based System for Effective Surveillance and Control ofCOVID-19: Useful Experiences From Hubei, China. J Med Internet Res 2020;22(4):e18948

J Med Internet Res 2020 | vol. 22 | iss. 4 | e18948 | p. 1

6. Qian Liu, Zequan Zheng, Jiabin Zheng, Qiuyi Chen, Guan Liu, SihanChen, Bojia Chu, Hongyu Zhu, Babatunde Akinwunmi, Jian Huang, Casper J P Zhang, Wai-Kit Ming. Health Communication Through News Media During the EarlyStage of the COVID-19 Outbreak in China: Digital Topic Modeling Approach. J Med Internet Res 2020;22(4):e19118

7. Kui Liu, Sichao Huang, Zi-Ping Miao, Bin Chen, Tao Jiang, Gaofeng Cai1, Zhenggang Jiang, Yongdi Chen, Zhengting Wang, Hua Gu, ChengliangChai, Jianmin Jiang. Identifying Potential Norovirus Epidemics in China via Internet Surveillance. J Med Internet Res 2017;19(8):e282

8. Epidemic situation in real time. Doctor Lilac Network. URL: <https://ncov.dxy.cn/ncovh5/view/pneumonia> [accessed 2020-01-20]

9. Epidemic real-time big data report. Baidu. URL: <https://voice.baidu.com/act/newpneumonia/newpneumonia/> [accessed 2020-01-20]

10. Real-time outbreak tracking. Tencent. URL: <https://news.qq.com/zt2020/page/feiyan.htm#/> [accessed 2020-01-20]

11. Using big data to prevent and control epidemics. Cyberspace Administration of China and Office of the Central Cyberspace Affairs Commission. URL: <http://www.cac.gov.cn/2020-02/17/c_1583484435541835.htm> [accessed 2020-02-17]

12. Ministry of industry and information technology: strengthen big data analysis to support epidemic prevention and control. Economic Daily. URL: <https://baijiahao.baidu.com/s?id=1658594919216396522> [accessed 2020-02-15]

13. Zhejiang takes the lead in achieving full coverage of “Health Code”. Chinese Government. URL: <http://www.gov.cn/xinwen/2020-02/18/content_5480261.htm> [accessed 2020-02-18]

14. Chongqing “Health Code” goes online and people go back to work. Chinese Government. URL: <http://www.gov.cn/xinwen/2020-02/19/content_5481017.htm> [accessed 2020-02-19]

15. Tianjin launches “Health Code” to help prevent epidemic resumption. Chinese Government. URL: <http://www.gov.cn/xinwen/2020-02/29/content_5485080.htm>

<http://www.gov.cn/xinwen/2020-03/01/content_5485215.htm> [accessed 2020-02-29]

16. Guidance for prevention and control work in the commerce and trade industry is published, and by using the “Health Code in Shanghai” as a safety assessment tool for employees and customers. Shanghai Municipal Government. URL: <http://www.shanghai.gov.cn/nw2/nw2314/nw32419/nw48516/nw48539/u21aw1431352.html> [accessed 2020-03-03]

17. Jiangsu launches “Health Code”, and the Yangtze River Delta takes the lead in achieving full coverage. URL: <http://www.gov.cn/xinwen/2020-03/05/content_5487067.htm> [accessed 2020-03-05]

18. Li Ying. Vice President of Baidu: Spring Festival return journey is lengthened under epidemic prevention and control. CCTV News. URL: <http://www.cctime.com/html/2020-1-31/1497804.htm> [accessed 2020-01-31]

19. Guang Shi. Epidemic prevention and control from the perspective of big data. State Council Development Research Center. URL: <https://xw.qq.com/cmsid/20200216A0LJVV00?f=newdc> [accessed 2020-02-16]

20. Real-Time Epidemic Data. Doctor Lilac Network. URL: <https://ncov.dxy.cn/> [accessed 2020-02-29]

21. China's Population Migration Data. Baidu Migration. URL: <https://qianxi.baidu.com/> [accessed 2020-02-29]

22. Huiyun Xu. People coming to Shanghai from severely affected regions need to be isolated at home for 14 days strictly. First Financial Network. URL: <https://www.yicai.com/news/100481608.html?clicktime=1580525715>

23. Qun Li, M.Med., Xuhua Guan, Peng Wu, Xiaoye Wang, M.P.H., Lei Zhou, M.Med., Yeqing Tong, Ruiqi Ren, M.Med., Kathy S.M. Leung, Eric H.Y. Lau, Jessica Y. Wong, Xuesen Xing, Nijuan Xiang, M.Med., et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. The New England Journal of Medicine, January 29th, 2020 DOI: 10.1056/NEJMoa2001316. URL: <https://www.nejm.org/doi/full/10.1056/NEJMoa2001316?query=featured_home> [accessed 2020-01-29]

24. Qingzhu Dai. When the Epidemic Ends? The Reliable Answer Is Here!. Alibaba Staff Blockchain Alliance. <https://mp.weixin.qq.com/s/6I1onP3jt3G76IYjy9YWlw> [accessed 2020-02-10]

25. Outside Hubei, the City Has the Most Confirmed Diagnoses, with 227 Cases! Wuhan is Their “Second Hometown” Jiang You, China Fund News URL: <http://sa.sogou.com/sgsearch/sgs_tc_news.php?req=36lX5Aw6IoHk0uKWsTZiH9M3ZaYRYhjRcv0vr9jcX6Q=> [accessed 2020-01-31]

26. Wenzhou Launches Comprehensive Prevention and Control Fight. Wenzhou Government. <http://www.wenzhou.gov.cn/art/2020/1/29/art_1217829_41860286.html> [accessed 2020-01-29]

27． Wenzhou Comprehensively Implements "25" Emergency Measures for Epidemic Prevention and Control. Wenzhou Government. <http://www.wenzhou.gov.cn/art/2020/2/1/art_1217828_41865258.html> [accessed 2020-02-01]

28. Wenzhou, Zhejiang: “Fighting” with Epidemic Highlights Wenzhou's Role. People's Daily. URL: <http://news.cyol.com/content/2020-02/22/content_18390200.htm> [accessed 2020-02-22]

29. Why Heilongjiang Has the Highest Number of Confirmed Diagnoses in Northeast of China. Doctor Lilac Network. URL: <https://mama.dxy.com/outbreak/daily-of-nationwide-new?index=20200208&locationIds=999&share=true&entry=Previous_Review> [accessed 2020-02-08]

30. Hypothesis Test for the Population Correlation Coefficient. URL: <https://online.stat.psu.edu/stat501/lesson/1/1.9> [accessed 2020-02-29]

**Abbreviations**

REGION: province or municipality in China

REGIONS: the plural of REGION

NEW: The daily new cases diagnosed in the REGION

ACCUMULATED NEW: The sum of NEW values in the recent 3 days

POPULATION: The daily incoming immigration population in each REGION

RISK: The risk of POPULATION in each REGION

OFFSET: The offset parameter in the model

WINDOW: The window parameter in the model

OFFSET + WINDOW: The sum of OFFSET and WINDOW values for a region in one day

PROCESSED RISK: RISK processed by OFFSET and WINDOW

CORR: The correlation coefficient of NEW and RISK or PROCESSED RISK