**An evaluation model of 2019-nCoV prevention and control effectiveness based on immigration data in China**

Qiangsheng Huang

**Abstract**

**Background:** As of the end of February 2020, 2019-nCoV is currently well controlled in China. However, the virus is now spreading globally.

**Objective:** This study aimed to evaluate the effectiveness of outbreak prevention and control measures in a region.

**Methods:** A model is built for find the best fit for two sets of data (the number of daily new diagnosed, and the risk value of incoming immigration population). The parameters (offset and window) in the model can be used as the evaluation of effectiveness of outbreak prevention and control.

**Results:** Through study, it is found that the parameter offset and window in the model can accurately reflect the prevention effectiveness. Some related data and public news confirm this result. And this method has advantages over the method using R0 in two aspects.

**Conclusions:** If the epidemic situation is well controlled, the virus is not terrible. Now the daily new diagnosed patients in most regions of China is quickly reduced to zero or close to zero. Chinese can do a good job in the face of huge epidemic pressure. Therefore, if other countries can do well in prevention and control, the epidemic in those places can also pass quickly.

Key words: 2019-nCoV; Epidemic Prevention and Control; Big Data Analysis; Data mining; Mathematical modeling

**Introduction**

It has been 2 months since the outbreak of novel coronavirus (2019-nCoV)–infected pneumonia (NCIP) in Wuhan, Hubei, since December 2019. Chinese has been ignorant of it from the beginning. For example, the people of Wuhan held a banquet for tens of thousands of people in January 18th, which caused the spread of the epidemic [1]. Then, Chinese gradually paid attention to the investigation and control of incoming immigration population, and closed down some densely populated regions. And minimize internal people going out, wearing masks when going out, community prevention, and so on [2,3]. At present, the epidemic in the regions other than Hubei, which is the region where the virus originated, have been basically controlled.

This study uses immigration population data and daily new diagnosed data to find ways to assess the evaluation for effectiveness of epidemic prevention and control in each region of China.

**Methods**

**Abbreviations**

To simplify the description, the abbreviations are used for the following nouns in the later paper:

The provinces / municipalities of China is referred to as “region”.

The daily new diagnosed in each region is referred to as “daily new diagnosed” or just “new” (in quotes).

The daily incoming immigration population size in each region is referred to as “daily incoming population size” or just “population” (in quotes).

The risk of daily incoming immigration population in each region is referred to as “daily immigration risk” or just “risk” (in quotes).

The offset parameter in the model is referred to as “offset” (in quotes).

The window parameter in the model is referred to as “window” (in quotes).

The sum of “offset” and “window” values for a region in one day is referred to as “offset + window” (in quotes).

The “risk” data processed by “offset” and “window” parameter is referred to as “processed risk” (in quotes).

The correlation coefficient of “new” and “risk” (or “processed risk”) is referred to as “corr” (in quotes).

**Overview**

A model is built to evaluate the effectiveness of prevention and control for 2019-nCoV in all regions of China. The main theory, which will be introduced later in detail, is that both the "offset" and "window" parameters in the model are used to reveal the delayed days of "risk" convert to "new". More days means worse effectiveness. The model workflow is shown in figure 1. These is three main steps in this model for evaluate a single region in a period of days:

First, “risk” data is constructed from “population” and “new” data.

Second, “risk” data is processed into “processed risk” data by “offset” and “window” parameters.

Last, the “offset” and “window” values corresponding to the largest correlation coefficient of “new” and “processed risk” data are the output of the model.

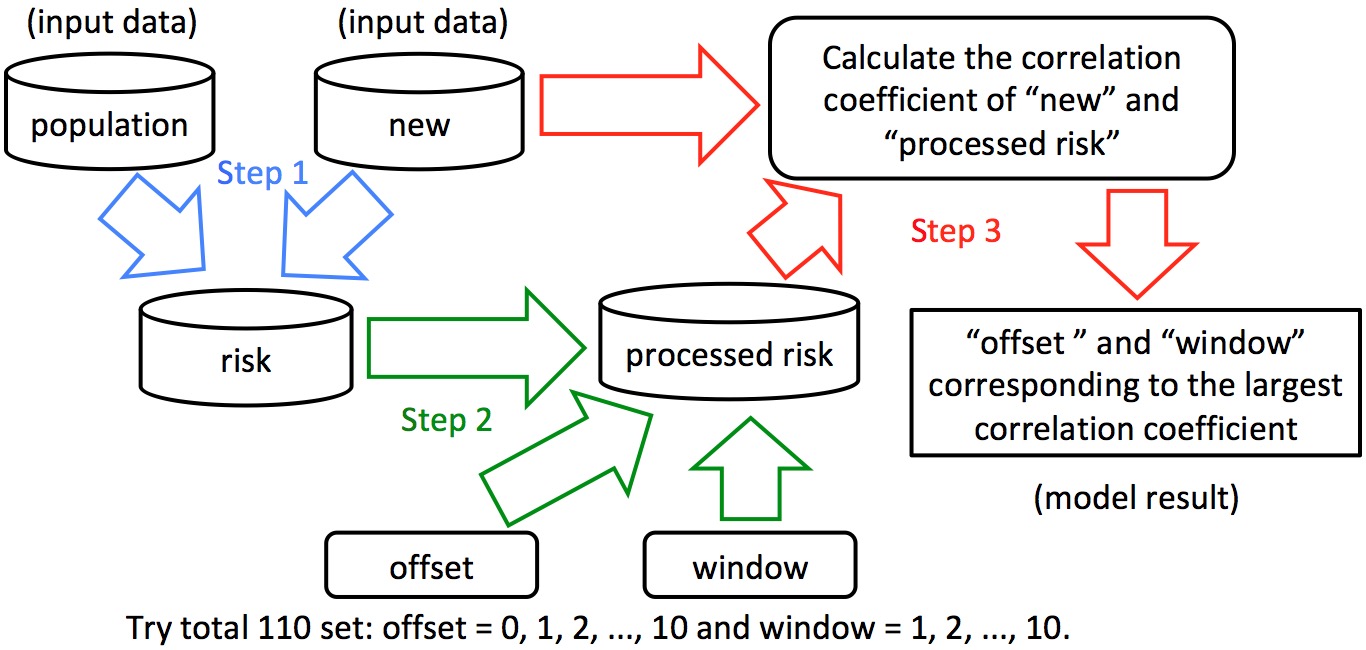


Figure 1. The model workflow

**Model input data**

***Model input Data 1: daily new diagnosed in each region data (the “new” data)***

Since January 2020, various provinces / municipalities in China have successively announced data such as the number of new confirmed diagnoses of the epidemic each day. The “new” data is crawled from Doctor Lilac Network (<https://ncov.dxy.cn/>).

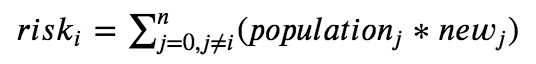
***Model input Data 2: daily incoming immigration population size in each region data (the “population” data)***

The population migration factor and its control must be the most important factor affecting the developing trend of the epidemic situation. Therefore, the daily incoming immigration population data, which are distinguished from different source regions, is crawled from Baidu Migration (<https://qianxi.baidu.com/>).

Because no immigration population data was got for Hong Kong, Macao, and Taiwan, and the traffic of Hubei has been closed from late January, these regions haven’t been analyzed. However, the immigration population moved from Hubei to other regions where are analyzed are considered.

**Construct daily immigration risk in each region data (the “risk” data)**

It should be considered that the immigration population with same size from different sources, its degree of risk is different. Therefore, process the incoming immigration population as follow, and get the “risk” data.



In the formula, all the values of “risk”, “population” and “new” are in a same day.

“riski” is the “daily immigration risk” of region i in one day. (i can be 1, 2, 3, …, n. “n” is a fix number. And in this paper “n” is 30, because this paper analyzed 30 regions.)

“populationj” is the daily incoming immigration population from source region j. (j can be 1, 2, 3, …, n, and j cannot be same as i.)

“newj” is the new diagnosed people count in source region j in recent 3 days.

**Introduction for “offset” and “window” parameter**

***The “offset” parameter, which is used to evaluate the control for incoming immigration population***

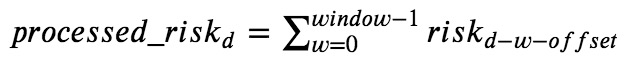
On the one hand, for the incoming immigration population, the preventive measures will be different at different times or in different regions. For example, in some areas, the temperature of immigrants, cough and other medical examinations have been strictly investigated relatively early. Such strict external people flow control will cause infected immigrants are immediately diagnosed, so the number of “daily new diagnosed” will increase significantly on the day when the “risk” is suddenly high or the day after. On the contrary, these externally infected people will be finally diagnosed relatively later. Therefore, the concept and the parameter of “offset” is introduced: the “risk” is shifted by a certain number of days, for example offset = 3, which means that the “new” of each day is compared with the “risk” of 3 days ago.

***The “window” parameter, which is used to evaluate the control for inner people***

On the other hand, the control of internal people and their own awareness of prevention will also affect the spread of the epidemic. For example, infected people who have moved inside the region (have been sick or incubated) and did not seek medical treatment or home isolation, and poor awareness such as going out without a mask will delay the external “risk” until the next few days before the “new” diagnosed people appear. In other words, the regions have the phenomenon of person-to-person transmission. On the contrary, only externally infected people are infected themselves, there will be only “offset” alone. Therefore, the concept of window is introduced into the model. For example, if the “window” is 10, it means that the total “risk” for 10 consecutive days will affect “new” on the 10th day. From reference [4], we can know the incubation period with 95% confidence interval is 4.1-7.0 days, so if the infected person who moved in 10 days ago can still affect the inside, it should be much serious from person to person.

**Process “risk” data by “offset” and “window”**

Based on the analysis and the concepts introduced above. The “risk” can be processed by “offset” and “window” as follow formula:



In the formula, all the “processed\_risk” and “risk” is for a same region.

“processed\_riskd” is the value of processed “risk” by “window” and “offset” at the date of “d”.

“riskd-w-offset” is the value of “risk” at the date of “d-w-offset”.

For example, when it is need to calculate the value of “processed\_risk” in 2020-02-10 under window = 2, offset = 3. The formula is as follow:



And if offset = 0, window = 1. “processed\_riskd” is just “riskd” without any process:



**Calculation of correlation coefficient between “new” and “processed risk”**

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 in each day.

For each region in each day, starting from January 17th, 2020 (the first day of accurate "new" data collected), the “offset” is tried from 0 to 10, and the “window” is tried from 1 to 10. There are 110 different “offsets” and “windows” sets. The 110 sets are used to process “risk” one by one, and calculate the 110 correlation coefficients with “new” and “processed risk”. Finally, the set of “offset” and “window” corresponding to the maximum correlation coefficient is just the model output for the region in the day.

**Results**

**Outputs of model**

Because the output data is a quite large (over 40 days for 30 regions), the representative result data in every 3 days from February 2nd to February 11th for 9 regions, which are compared with related news and data later, is shown in Table 1. (The reason for choosing these days is that the peak of the “new” in all regions of China is around February 2nd, and value of “new” decrease to half of the peak around February 11th.) If you need full outputs, you can find them from the attachments of this paper.

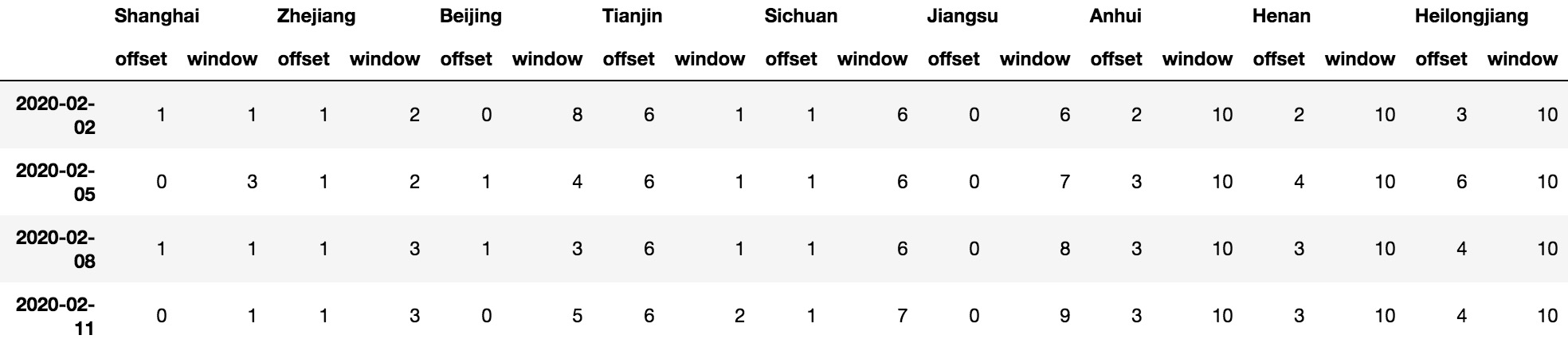


Table 1. “offset” and “window” for 9 regions in February 11th

This table can be used as an evaluation of the effectiveness of prevention and control of the epidemic for each region in each day. The regions are sorted by the values of “offset + window” in ascending order, which is the same sorting order of prevention and control effectiveness. We can see effectiveness of prevention and control in Shanghai, which owns the lowest “offset” and “window” values, has done the best job in the 9 regions. On the contrarily, the effectiveness in Heilongjiang is worst.

**Illustrate the role of “offset” and “window” parameters by line charts**

For instance, we can use the data of Jiangsu and Heilongjiang to illustrate the role of “offset” and “window” parameters by some line charts.

In Figure 2 and 3, “new” and “risk” data. (Because the value of “risk” is a very large value, drawing it with “new” will make “new” polyline chart as a very low horizontal line. Therefore, when drawing, the maximum “risk” value is set to the maximum “new” value, and the remaining “risk” values are reduced in proportion.) It was found that the correlation coefficient value of “new” and “risk” for Jiangsu and Heilongjiang are 0.684 and -0.014.

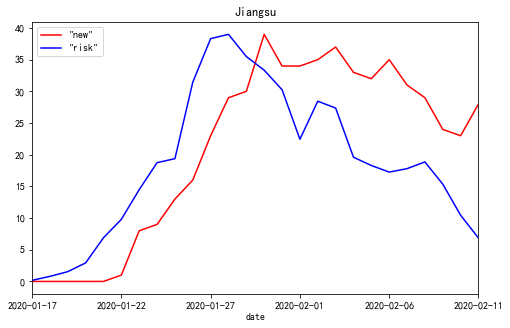


Figure 2. The polyline chart of “new” and “risk” for Jiangsu Province from 2020-01-17 to 2020-02-11

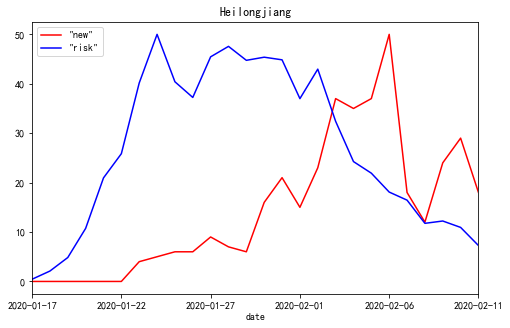


Figure 3. The polyline chart of “new” and “risk” for Heilongjiang Province from 2020-01-17 to 2020-02-11

If we use “processed risk” instead of “risk” to draw the polyline chart, the lines are much more fitted as illustrated in figure 4 and 5. And the correlation coefficient values increase to 0.979 and 0.874.

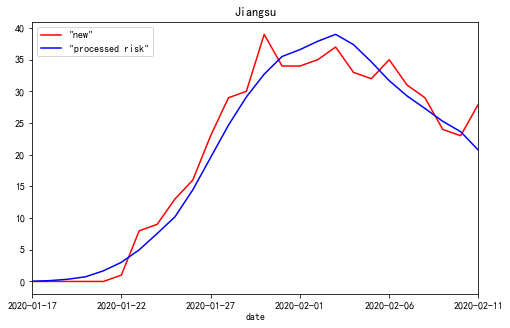


Figure 4. The polyline chart of “new” and “processed risk” for Jiangsu Province from 2020-01-17 to 2020-02-11, when offset=0, window=9

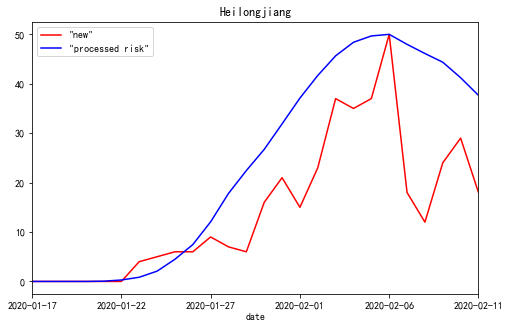


Figure 5. The polyline chart of “new” and “processed risk” for Heilongjiang Province from 2020-01-17 to 2020-02-11, when offset=4, window=10

**Discussion**

**More discussion about “offset” and “window” from the results**

We can see from figures 2, 3, 4 and 5 that “offset” and “window” can reveal the delayed days of “risk” convert to “new”. Because, if in the most ideal case that every infected person who enters from the outside into the region can be detected and isolated in the first seconds, polylines of “new” and “risk” will be fully fitted. And the value of “offset” should be 0, “window” should be 1, and the “corr” should be 1. Conversely, if external infections entering the region are not detected in time and the virus is transmitted to locals, the “risk” would be transferred to the next few days. These delayed days are evaluated by the values of “offset” and “window”.

**Related data and news confirm the correctness of the model**

Although the data released by the government which can be used to compare the effectiveness of epidemic prevention and control in different regions is relatively little, data and news of 9 regions were found which can be used for the confirmation for the model.

First, according to the data released by Doctor Lilac Network in China, the numbers of cumulative diagnosed people are divided into incoming immigrants and inner citizens in only 3 regions: Shanghai, Beijing, and Tianjin. Until February 11th, the data is show as table 2. We can focus “inner diagnosed rate” and the “offset” and “window” values in table 1. Shanghai did best control (owns lowest “offset + window” value) so the inner people diagnosed rate is lowest. Beijing is second, and the rate of Tianjin (owns highest “offset + window” value) is the highest.

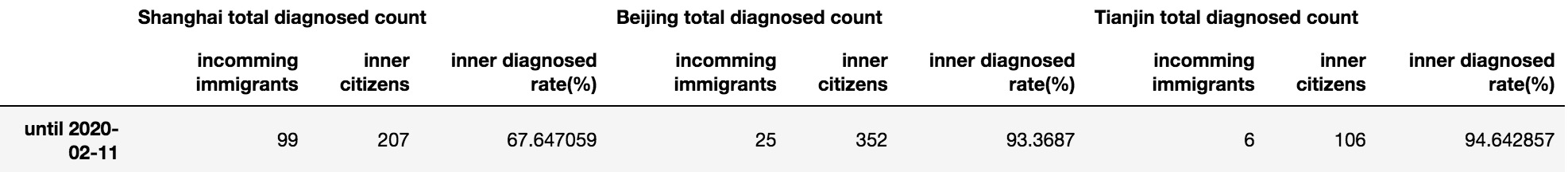


Table 2. The total diagnosed people count divided into incoming immigrants and inner citizens

Second, R0 data of Shanghai, Zhejiang, Jiangsu, Anhui, Henan, and Sichuan were found [5].

The R0 values in the date of the article published (February 10th) are shown as table 3. Compared with table 1, we can see the relative values and sorting order of R0 and “offset + window” is almost same.

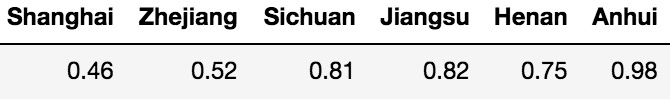


Table 3. The values of R0, “offset” and “window” in February 10th

The correctness of the model output is as confirmed by some related news as follow.

On the one hand, in late January, there were a large number of infected businessmen who were returning to Wenzhou, Zhejiang from Wuhan, Hubei [6]. On February 1th, the government of Wenzhou, Zhejiang issued 25 strict preventions and control measures in a timely manner [7,8]. On February 22, after the Wenzhou epidemic was completely under control, the Chinese government's newspaper published an article strongly affirming the achievements of Wenzhou's epidemic prevention and control [9]. Although Zhejiang (the province where Wenzhou is located) has the huge pressure, Zhejiang did a good job. And the values of “offset” and “window” are relatively low in table 1.

On the other hand, on February 8th, it is pointed out with survey data that Heilongjiang did not pay enough attention to the epidemic and had poor awareness of prevention [10]. Judging from the “offset” and “window” values of Heilongjiang in table 1, this is indeed the case.

**Conclusion**

At present, although the epidemic prevention and control in China has achieved great results, it continues. New outbreaks are gradually emerging around the world. The model in this paper may be applicable to the analysis of epidemic prevention and control effectiveness in various parts of the world. I hope this study can contribute to the prevention and control of the world epidemic. It’s like a Chinese saying, we, the people all over the world, live in a global village, and we are a family. We Chinese people hope that the epidemic will be brought under control worldwide as soon as possible and finally ended.

**References**

1. CCTV Network: "Banquet for Ten Thousand People" Turns into "Disastrous Banquet", and the history cannot be repeated. Netease News. URL: <https://news.163.com/20/0206/20/F4NTKBK60001899O.html> [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. Qun Li, M.Med., Xuhua Guan, Ph.D., Peng Wu, Ph.D., Xiaoye Wang, M.P.H., Lei Zhou, M.Med., Yeqing Tong, Ph.D., Ruiqi Ren, M.Med., Kathy S.M. Leung, Ph.D., Eric H.Y. Lau, Ph.D., Jessica Y. Wong, Ph.D., Xuesen Xing, Ph.D., 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]

5. Qingzhu Dai. When will the epidemic end? The reliable answer is here!. Alibaba Staff Blockchain Alliance. <https://mp.weixin.qq.com/s/6I1onP3jt3G76IYjy9YWlw> [accessed 2020-02-10]

6. 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]

7. 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]

8． 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]

9. Wenzhou, Zhejiang: "War" epidemic in the city highlights Wenzhou's role. People's daily. URL: <http://news.cyol.com/content/2020-02/22/content_18390200.htm> [accessed 2020-02-22]

10. 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]