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.

Science of the Total Environment 727 (2020) 138394

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Real-time estimation and prediction of mortality caused by COVID-19 with patient information based algorithm

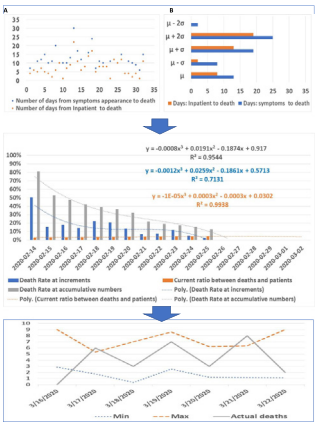
Lishi Wang [a,b,1], Jing Li [b,1], Sumin Guo [c], Ning Xie [d], Lan Yao [e,f,k], Yanhong Cao [f,k], Sara W. Day [g], Scott C. Howard [g], J. Carolyn Graff [g], Tianshu Gu [h], Jiafu Ji [i], Weikuan Gu [b,j,*], Dianjun Sun [f,**]

[a] Department of Basic Medicine, Inner Mongolia Medical University, Inner Mongolia 010110, PR China
[b] Department of Orthopedic Surgery and BME-Campbell Clinic, University of Tennessee Health Science Center, Memphis, TN 38163, USA
[c] Department of Oncology, Hebei Chest Hospital, Lung Cancer Control and Prevention Center of Hebei Province, Shijiazhuang, Hebei 050041, PR China
[d] College of Business, University of Louisville, Louisville, KY 40292, USA
[e] Health Outcomes and Policy Research, College of Graduate Health Sciences, University of Tennessee Health Science Center, Memphis, TN 38103, USA
[f] Center for Endemic Disease Control, Chinese Center for Disease Control and Prevention, Harbin Medical University, 157 Baojian Road, Harbin, Heilongjiang 150081, PR China
[g] College of Nursing, University of Tennessee Health Science Center, Memphis, TN 38105, USA
[h] Department of Neurology, Beijing Tiantan Hospital, Capital Medical University, Beijing 100050, PR China
[i] Beijing Cancer Hospital and Key Laboratory of Carcinogenesis and Translational Research, Department of Gastrointestinal Surgery, Peking University Cancer Hospital and Institute, Beijing 100142, PR China
[j] Research Service, Memphis VA Medical Center, 1030 Jefferson Avenue, Memphis, TN 38104, USA
[k] Key Laboratory of Etiologic Epidemiology, Education Bureau of Heilongjiang Province & Ministry of Health (23618104), Harbin, Heilongjiang 150081, PR China

HIGHLIGHTS

- The mortality rate determines whether a highly infectious disease becomes a public concern.
- Summarizing information after the fact does not contribute to real-time readiness to deal with the disease.
- The Patient Information Based Algorithm (PIBA) estimates the death rate of a disease in real-time.
- PIBA can be used to estimate the death rate of a new infectious disease in real time and to predict future deaths.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 28 March 2020
Received in revised form 30 March 2020
Accepted 31 March 2020
Available online 8 April 2020

ABSTRACT

The global COVID-19 outbreak is worrisome both for its high rate of spread, and the high case fatality rate reported by early studies and now in Italy. We report a new methodology, the Patient Information Based Algorithm (PIBA), for estimating the death rate of a disease in real-time using publicly available data collected during an outbreak. PIBA estimated the death rate based on data of the patients in Wuhan and then in other cities throughout China. The estimated days from hospital admission to death was 13 (standard deviation (SD), 6 days). The death

* Correspondence to: W. Gu, 956 Court Avenue, Memphis, TN 38163, USA.
** Correspondence to: D. Sun, 157 Baojian Road, Harbin, Heilongjiang 150081, PR China.
E-mail addresses: wgu@uthsc.edu (W. Gu), hrbmusdj@163.com (D. Sun).
[1] These two authors contributed to this manuscript equally.

https://doi.org/10.1016/j.scitotenv.2020.138394
0048-9697/© 2020 Elsevier B.V. All rights reserved.

Editor: Jay Gan

Keywords:
Coronavirus
COVID-19
Death rate
Inpatient
Normal distribution
Prediction

rates based on PIBA were used to predict the daily numbers of deaths since the week of February 25, 2020, in China overall, Hubei province, Wuhan city, and the rest of the country except Hubei province. The death rate of COVID-19 ranges from 0.75% to 3% and may decrease in the future. The results showed that the real death numbers had fallen into the predicted ranges. In addition, using the preliminary data from China, the PIBA method was successfully used to estimate the death rate and predict the death numbers of the Korean population. In conclusion, PIBA can be used to efficiently estimate the death rate of a new infectious disease in real-time and to predict future deaths. The spread of 2019-nCoV and its case fatality rate may vary in regions with different climates and temperatures from Hubei and Wuhan. PIBA model can be built based on known information of early patients in different countries.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

The mortality rate is the most important factor that determines whether a highly infectious disease becomes a public concern and carries risks causing a pandemic. Different virus epidemics take place throughout the world every year, but only a few rise to the level of public concern (Schlagenhauf and Ashra, 2003; Viboud and Simonsen, 2012; WHO Ebola Response Team, 2014). Severe acute respiratory syndrome (SARS), swine influenza A H1N1 virus (H1N1), and Zaire ebolavirus (Ebola) brought on the public's attention because they caused many severe infections and thousands of deaths (Dawood et al., 2012; Nicholls et al., 2003; WHO Ebola Response Team, 2014). Similarly, the disease COVID-19 caused by a coronavirus (2019-nCoV) brought world-wide attention and caused public panic because many deaths had been reported without being put in the context of the many mild infections and its potentially low case fatality rate (Chan et al., 2020; Huang et al., 2020; Wang et al., 2020; Wu et al., 2020). For example, despite being a common infection, influenza rarely causes public concern because even though it is common, it leads to death in only 0.1% of cases. A variety of reports indicate that 2019-nCoV is highly infectious through multiple routes (Chen et al., 2020; Huang et al., 2020; Wu et al., 2020). While the high infection rate is certain, the mortality rate of COVID-19 has not been definitively determined. It is reasonable to suspect that the deaths of six of the first 41 patients (15%) in Wuhan (Huang et al., 2020) in the earliest reports by Chinese scholars were inaccurate. When the initial mortality rates were reported, only patients who were critically ill were included. Patients with mild symptoms, as well as those with asymptomatic infections, were not analyzed (Chen et al., 2020; Huang et al., 2020; Wu et al., 2020). Case-fatality rates reported by Huang et al. (2020) analyzed a skewed patient sample since it included only a small number of patients who had been transferred from other hospitals due to their critical condition. Therefore Huang et al.'s sample was skewed towards a concentration of severely ill patients, while the general patient population includes more patients with COVID-19 who are asymptomatic or only have mild symptoms and who have not been hospitalized. Chen et al. (2020) reported an 11% death rate, again based on patients with severe conditions.

We have estimated the mortality rate using a Patient Information Based Algorithm (PIBA). The PIBA uses patient data in real-time to build a model that estimates and predicts death rates for the near future. PIBA uses data of patients identified early in the disease process to calculate the average number of days from hospitalization to death for those hospitalized. Another feature is to take into account variations based on mathematical models. The PIBA calculation method does not divide the total number of patients on a day by the number of deaths on the same day. Instead, the PIBA calculation method divides the number of deaths on that day by the number of possible patients of a day or days when the patients have just begun to develop the disease.

Classic death rate calculation method:

$$\text{Death rate} = \frac{\text{Total number of deaths from the disease}}{\text{Total number of infected patients}}$$

PIBA method:

$$\text{Death rate} = \frac{\text{Number of deaths (in a given day)}}{\text{Average number of patients in previous days determined by the alogorithmn}}$$

Thus, PIBA comprehensively and reasonably estimates the mortality rate based on the actual number of deaths and estimates the number of patients on a specific day. As time goes on, large amounts of data from northern and southern China have been accumulated through continuous reporting, all of which are used by PIBA, which then becomes more accurate as data accumulates. We conclude that it is time to utilize the accumulated data to estimate the case fatality rate of COVID-19 infection. Based on national data from the China National Health Center, the COVID-19 death rate is much lower than that reported in Huang et al. (2020). Holistic data covering all of Wuhan, the epicenter city of COVID-19, also indicates a death rate lower than that reported by Huang et al. These data sources cover a larger patient sample, and include patients displaying symptoms with varying levels of severity. Therefore, the updated estimation of the death rate should reference these larger scale and more representative data.

Our study contributes to knowledge on COVID-19 death rate by building on Huang et al.'s (2020) estimation and available data from official websites and addressing the limitations with a larger and more representative sample.

2. Methodological approach

2.1. Steps for estimating and predicting mortality using PIBA

1) To collect data from the patient's initial admission to death. Strive to collect data for a certain number of patients.
2) To calculate the average number of days (μ) from hospital admission to the death and the number of days between one standard deviation ($\mu \pm \sigma$) interval and two standard deviations ($\mu \pm 2\sigma$).
3) To use these parameters ($\mu, \mu \pm \sigma, \mu \pm 2\sigma$) to calculate the daily mortality during the epidemic.
4) To predict the mortality of infectious diseases in the future based on the calculated known mortality combined with the number of patients in a region. The predicted numbers are compared with real mortality to test and correct model data.
5) To conduct following-up modification of the PIBA model according to different nationalities and regions. In particular, the initial patient data collected may vary significantly from country to country, one ethnic group to the other, and region to region.

2.2. Estimation of the number of days between hospitalization and death using PIBA

The calculation based on the number of deaths and the number of patients on the same day does not reflect the real death rate because most patients with COVID-19 do not die on the same day that they entered the hospital (Chan et al., 2020; Huang et al., 2020; Wang et al., 2020). With the PIBA method, we recognize that the patient population

size was inaccurate in the early days but trust the published information of patients who died right after COVID-19 outbreaks. The estimation is built upon data from patients with a normal distribution model.

Based on information about patients in Wuhan who died during the period between Dec 16, 2019, to Jan 2, 2020 (Huang et al., 2020), two parameters were used to estimate days from onset of symptoms to death and days from admission to the intensive care unit (ICU) to death. These two parameters are adopted in the estimation and prediction of COVID-19 death rate. Each parameter has five values including the mean, μ, one standard deviation from the mean, $\mu \pm \sigma$, and two standard deviations from the mean, $\mu \pm 2\sigma$.

2.3. Data collection of patient information

We collected data from COVID-19 patients in China from three public websites. The data from the whole country are collected and made available on the official website of the Health Emergency Office of the National Health Commission of the People's Republic of China at http://www.nhc.gov.cn/yjb/new_index.shtml. The data from Hubei Province and Wuhan are from the Health Commission of Hubei Province at http://wjw.hubei.gov.cn/fbjd/dtyw/. These data include the number of patients with COVID-19 who were confirmed as having the disease, who died from the disease, whose condition was severe, and who were admitted to the hospital or ICU. Other collected data included daily new cases, new deaths, people who were in close contact with an infection source, and accumulated number of patients. We paid particular attention to data from Wuhan, plus two additional cities in Hubei Province, Xiaogan, and Huanggang, in which the number of patients was higher than in other cities in Hubei Province.

Information from a northern province, Heilongjiang Province, was collected from the official website of Outbreak Information of the Health Commission of Heilongjiang province at http://wsjkw.hlj.gov.cn/index.php/Home/Zwgk/all/typeid/42. Data of Heilongjiang Province and Harbin city were included because the province is located in the northern high-altitude zone. These data are used to assess whether the COVID-19 is more, less, or equally likely to spread to an area with a cold climate. Collected information included numbers of patients and numbers of deaths from each city and in the whole province.

For any missing data in any day, a formula was used to estimate the data in that day: $Ni = \{(N(i+j) + (N(i-j)) / (j+1)\} + (N(i-j)$, where Ni = the estimated value of the missing data of the day i. j is the number of days of missing data, usually is 1; in the rare case, data of two consecutive days may be missing. If the data of two days are missing, the first day will be considered as the day i, the second day $N(i+1)$ will be calculated as $N(i+1) = Ni + \{(N(i+j) + (N(i-j)) / (j+1)\}$.

2.4. PIBA formula building for death rate estimation

Based on the days between confirmation of COVID-19 and the days of death in the hospital, calculated from Wuhan, as mentioned in method 1 and information from the whole country and Hubei Province, we tested the number of days from diagnosis to death, that most likely reflects the actual death rate. The estimated days are used to estimate the death rate using data from Hubei province and Wuhan city with the five values from above (μ, $\mu \pm \sigma$, $\mu \pm 2\sigma$). In consideration of the contribution of a variety of sources for the estimation, we fractured the data from (μ, $\mu \pm \sigma$ and $\mu \pm 2\sigma$) into the PIBA and built the testing model as follows.

1) $M_i = (D_i - D_{i-1}) / (P_{i-n} - P_{i-n-1})$ (death rate at increments)
2) $M_i = D_i / P_{i-n}$ (death rate at accumulative numbers)

where M_i = mortality rate, D_i = the cumulative numbers of deaths on day i, P_i = the cumulative numbers of patients on day $_i$, $_i$ = the current day for calculating the death rate, n = the number of days from severe infection to death.

When we considered these five partial values in normal distribution as a good indicator with a width of one standard deviation, each one of the five death rates calculated above on each day would have its own weight as the possible normal distribution ($\mu = 38.2\%$, $\mu - \sigma = \mu + \sigma = 24.2\%$, and $\mu + 2\sigma = \mu - 2\sigma = 6.7\%$).

From here, we could give the death rate for every single day just a single value that results from the weighted average of all five cohorts of patients, as defined by time from severe illness to death. The equation is as follows:

$$D = M\mu \times W\mu + M\mu{-}\sigma \times W\mu{-}\sigma + M\mu + \sigma \times W\mu + \sigma + M\mu + 2\sigma \times W\mu + 2\sigma + M\mu{-}2\sigma \times W\mu{-}2\sigma$$

where D = death rate, $M\mu$ = mortality rate with μ days, $W\mu$ = weight with μ days gap, μ = Mean in normal distribution, σ = Standard deviation.

2.5. Confirmation of the best estimation of the days to calculate the death rate in the other cities

The same formula was then used to estimate the death rate from the other two cities in Hubei province, namely Xiaogan and Huanggang. The PIBA model was developed using data from Hubei province, including

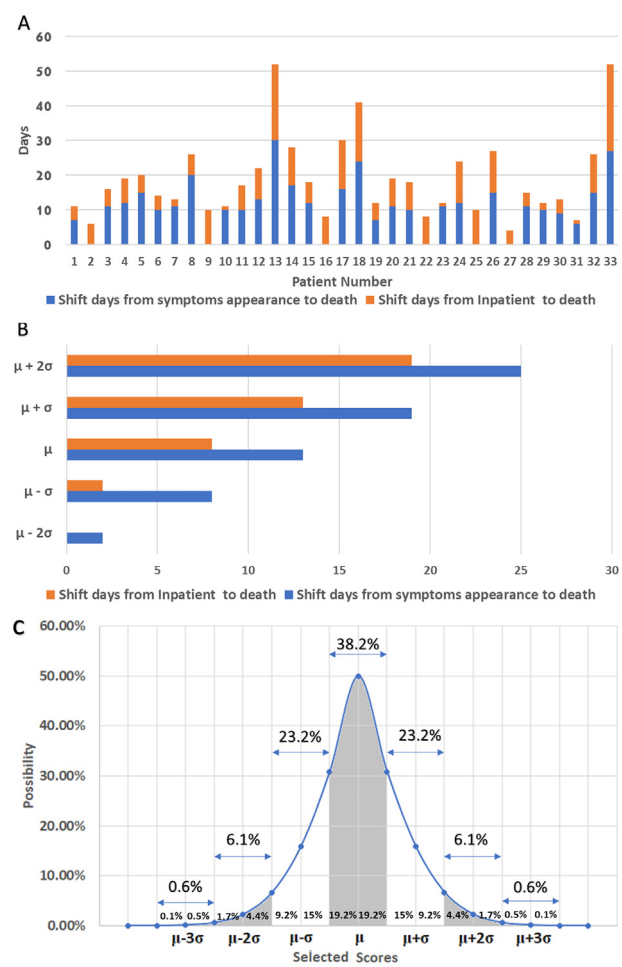


Fig. 1. Duration distribution of 33 death cases in Wuhan. A. Distribution of days between disease symptoms and death and between time of ICU admission and death. Vertical axis: days, Horizontal axis: cases. B. Estimated days from first symptoms to death and days from ICU admission to death. C. Lagging days (days from first symptoms to the day of death), μ, $\mu \pm \sigma$ and $\mu \pm 2\sigma$ and their weight (in percentages) used for the estimation of death rate in the broader patient population. Note: Among these values above, the lagging day $\mu - 2\sigma$ from symptom confirmation to death in panel B that equals to -3 has been set to 0.

Wuhan, Xiaogan, and Huanggang, and was further validated using data from Heilongjiang province and Harbin city. PIBA was then used to predict trends in new number of deaths.

2.6. Prediction of the number of deaths per day in the coming weeks using PIBA

In order to further test the validity of our PIBA method in predicting actual mortality, we used a combination of the curve trend data and the overall mortality rate of the country, Hubei, Wuhan, and the rest of the country (China overall except Hubei). Based on our prediction of the days from actual hospitalization to death, we separately predicted the number of deaths in each day of the coming week. That is, from the comprehensive information of the number of new patients on the seventh day, the 13th day, and the 19th day before the targeted prediction day, we obtained three numbers of deaths for each of the predicted days. Then from three of these numbers, the lower and upper values of the number of deaths on that day are used as the minimum and the maximum number of predicted deaths on that day, respectively. Also, the same formula was used to predict the death number of a week in South Korea.

3. Results

3.1. Estimated days from diagnosis to death and from ICU intake to death

Using information published by Wuhan, we calculated the days between ICU admission and death. We obtained the actual data from 33 patients who died in the hospital in Wuhan. The days from onset of symptoms to deaths ranged from 6 to 30 (see Fig. 1A). From ICU intake to death, the shortest number is one day, and the longest is 22 days. We derived two parameters, each from the 33 death cases, i.e., the days from onset of symptoms to death and the days from inpatient admission to death. Since there are six patients out of these 33 death cases who have the same date of symptoms' appearance and inpatient, there were 33 values in the dataset related to inpatient and 27 values in another dataset related to symptoms' appearance (Fig. 1A).

The results indicated that the average time from onset of the symptoms to death is 13 days ($M = 13$, $S.D. = 6$) (see Fig. 1B). Accordingly, the lagging days from the day of death and their weight in the calculation of death rate were derived based on the new inpatient days (Fig. 1C). The prediction of death rate is based on data from Wuhan city in which patients diagnosed with COVID-19 had been confirmed since January 19, 2020 and where deaths had occurred, which were among the first confirmed cases of coronavirus.

3.2. Estimated death rate for the whole country and Hubei province using PIBA formula

According to our five estimation parameters, from illness (i.e., symptom appearance) to death, the maximum number of days is 25 days. The earliest reported data in Wuhan was published on January 19, 2020. Based on these data, we were able to calculate the mortality rate from February 8, 2020, to the present. However, on February 12, the National Health Committee revised the data again (see Appendix

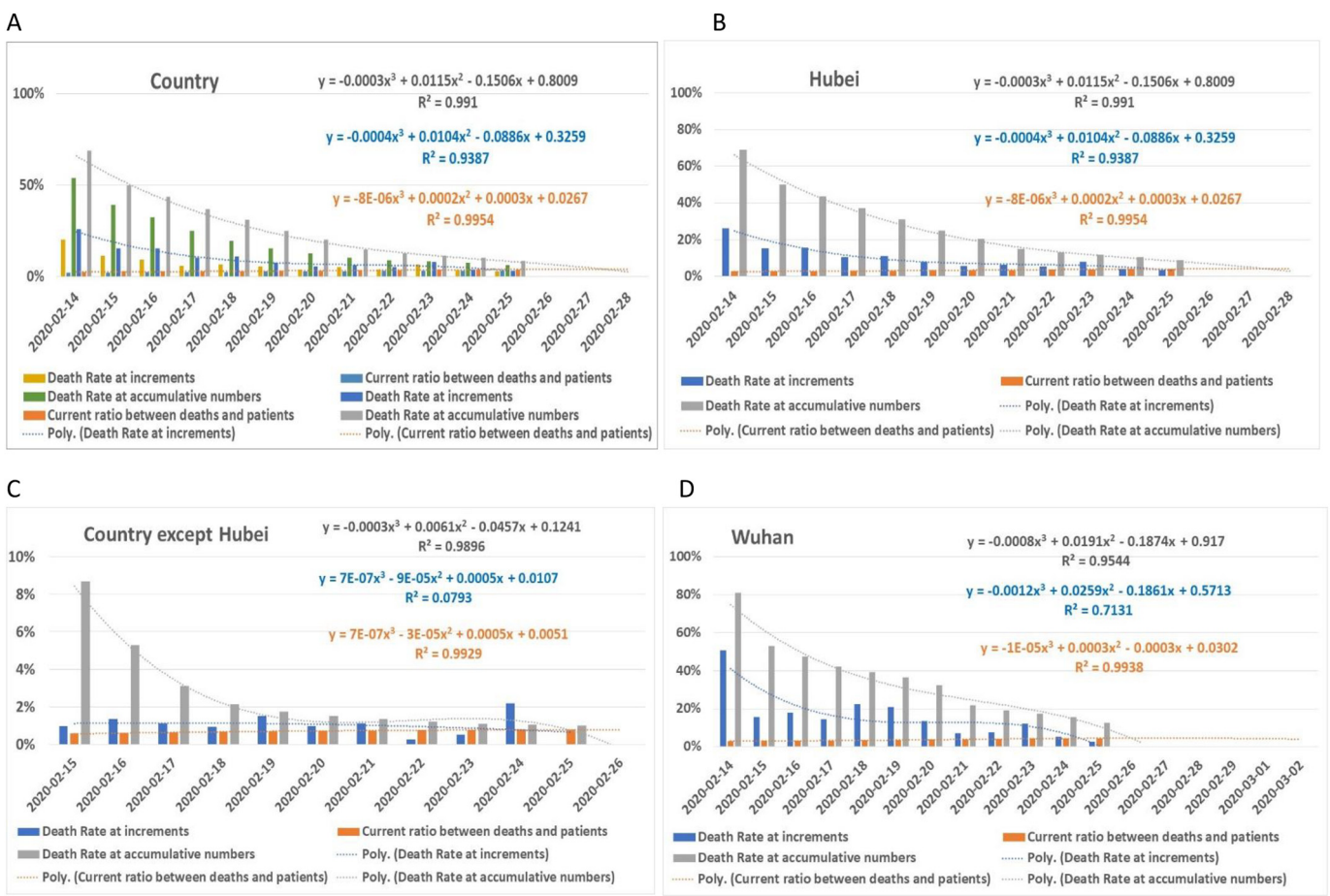


Fig. 2. The estimated death rate in Mainland China and Hubei Province. The blue curve represents the mortality calculated by the actual increase in deaths per lagging day divided by the increase in actual patients on the previous corresponding day. The gray curve represents the total number of deaths per lagging day, divided by the total number of identified actual patients on the corresponding previous day. The orange curve shows the number of deaths per day divided by the total number of patients the same day. The number on the vertical bar represents the death rate; the number on the horizontal bar shows the date. A: Overall death rate in Mainland China. B: Death rate in Hubei province. C. The death rate in the rest of the country except Hubei province. D: Death rate in Wuhan city.

Table 1). Because of this amendment, the number of confirmed cases appeared to have changed significantly in only one day. We chose the calculation results from February 14 up to February 25 (Appendix Table 2), considering that the death rates on February 12 and February 13 are likely distorted by this sharp rise within a short term. Fig. 2A through D provide information about the overall death rates in mainland China (hereafter referred to as country), Hubei, Wuhan, and rest of country (excluding Hubei) (Appendix Table 3).

We noticed that the death rate at increments based on PIBA in the whole country (in blue) in Fig. 2A is below 10%, with most values between 2.7% and 6% in the last five days. The death rate in Hubei province is similar to that of the whole country because 90% of the patients in the whole country were from Hubei province (see Appendix Table 1) (Fig. 2B). In Wuhan, the accumulated death rate was still high, as much as 20% (Fig. 2D). When we used the data from the rest of the country to test our PIBA formula, as expected, the curve is different from the curves from Hubei and Wuhan. Unlike in Hubei and Wuhan, the death rate of the rest of the country is much lower and stable, mostly lower than 1% (Fig. 2C). The predicted death rate will remain between 1% and 2% for the near future.

3.3. Estimation of death rate in other cities in Hubei province and in Heilongjiang province

Xiaogan and Huanggang are the two cities in Hubei province. The number of patients with COVID-19 in these two cities is higher than in other cities in Hubei except Wuhan. They also are the cities with the largest number of patients with COVID-19 in China. We, therefore, tested the PIBA formula using data from these two cities. Currently, the death rate based on the increment data is around 3%, lower than that in Wuhan but higher than that in the rest of the country. However, according to PIBA, the rate of deaths may decrease in the near future.

Heilongjiang province, including its capital city, Harbin, is the province outside of Hubei with the largest number of diagnosed patients. Harbin city is located in the northeast of China and is in the coldest area in China. No patients from Harbin city or the Heilongjiang province were reported during the SARS epidemic period. We used the PIBA formula to estimate the death rate in both the Heilongjiang province (Fig. 3C) and Harbin city (Fig. 3D). The death rate of Harbin decreased sharply in the past several days, into 0%. The low rate of less than 1% will possibly remain for the future.

3.4. Test the PIBA by predicting the daily death rate in the coming week

Based on the PIBA and the death rate of accumulated numbers, the expected final death rate of the whole country, Hubei, Wuhan, and rest of the country except Hubei, is predicted as follows (see Table 1). The predicted values are from the intersection points between the incremental estimation and net values estimation.

We used the predicted death rate to calculate the potential number of deaths per day in the coming week. Because our initial estimation on the lagging days between inpatient and death was only based on 33

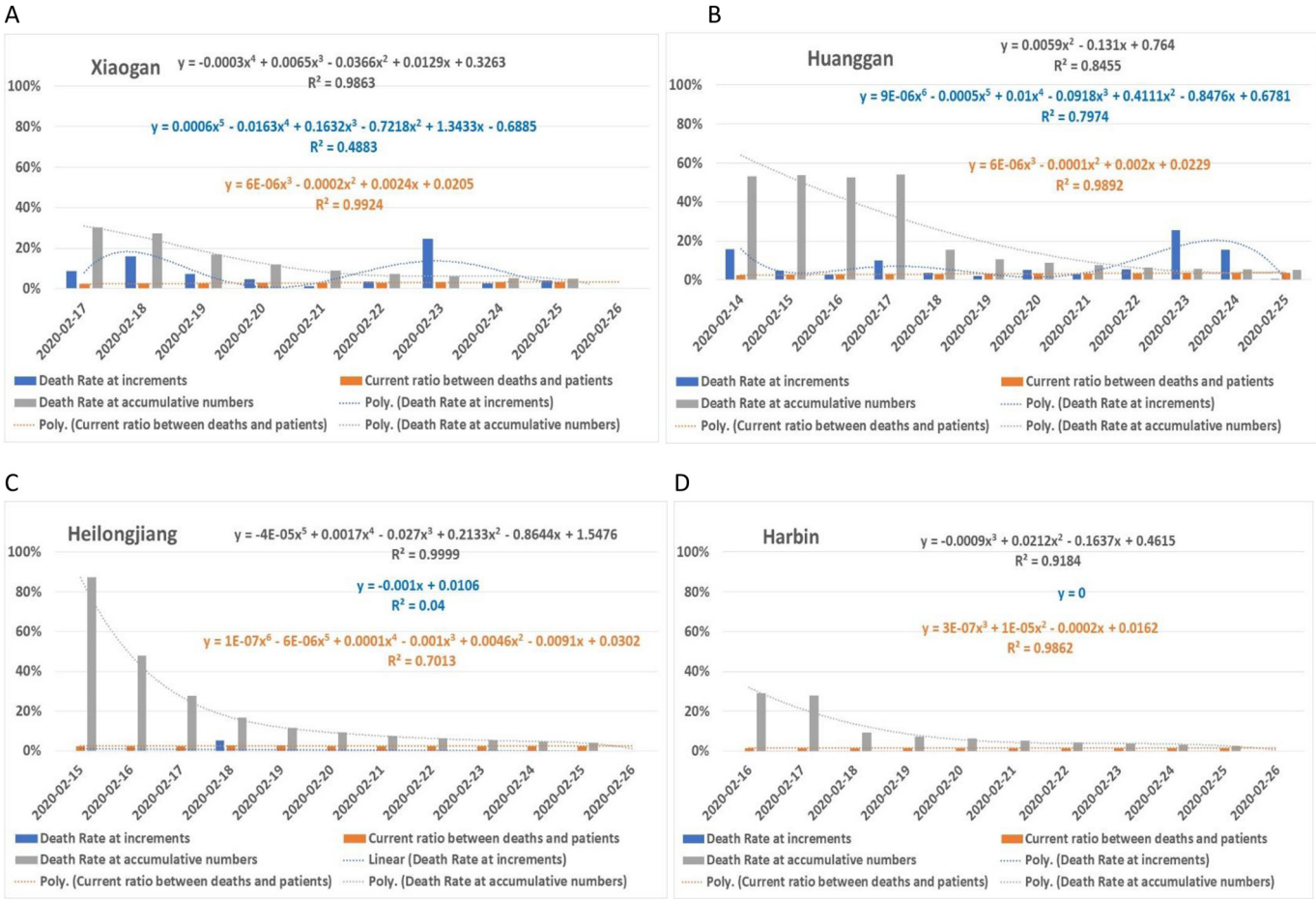


Fig. 3. Death rate estimations of four places. The blue curve represents the mortality calculated by the actual increase in deaths per lagging day divided by the increase in actual patients on the previous corresponding day. The gray curve represents the total number of deaths per lagging day, divided by the total number of identified actual patients on the corresponding previous day. The orange curve shows the number of deaths per day divided by the total number of patients the same day. Numbers on the vertical axis represent the death rate; on the horizontal axis is the date. A. The death rate of Xiaogan city in Hubei province B. Death rate of Huanggang city in Hubei province. C. The death rate in Heilongjiang province. D. The death rate in Harbin city.

Table 1
Prediction of death rate related to COVID-19.

Prediction of death rate	Intersect1	Intersect2	Max.
Country	3.23%	3.28%	3.28%
Hubei	4.72%	5.00%	5.00%
Wuhan	5.20%	5.20%	5.20%
Rest of country	0.71%	0.78%	0.78%

patients, we, therefore, used the days of average 13 days plus (19 days) and minus one standard deviation (7 days) as the range of number of deaths on a given day in the coming week (see Appendix Table 4. Predicted number of deaths in the days of the coming week after February 25, 2020). As shown in Fig. 4, the actual number of deaths in the past four days fell into the predicted range. In the country (Fig. 4A), Hubei (Fig. 4B), and Wuhan (Fig. 4C), the numbers of actual death were near the predicted minimum numbers. While, for the rest of the regions of the country except Hubei, the actual death data fluctuates between the predicted maximum and minimum values (Fig. 4C).

Due to the number of newly infected patients dropping in the last few days, the total number of patients tends to be constant or even less in the coming days if unexpected events do not occur. The peaks in these figures reflect sudden changes in numbers of patients (see Fig. 4). We believe that the intersecting point between the trendlines could reasonably be considered one of the rates in its range of the death rate of patients infected in the future.

3.5. Testing the PIBA using data from South Korea

As shown in the data above, the incidence in mainland China's provinces and cities was basically zero in late middle March. Because of this, we were not able to prove the feasibility of this method in more regions in mainland China. However, because the environment, medical conditions, and population races in different countries are different, to test the usefulness of the PIBA model in other countries, we need to get the basic information of the initial population. This information includes the specific number of days from onset to death of a reasonable number of patients in different regions of different countries. At present, we could not access these data accurately. The only thing we can do is to test Asian countries such as South Korea and Japan based on their ethnic similarities with populations in China. Taking all aspects into consideration, we believe that South Korea's data are more reliable. Therefore,

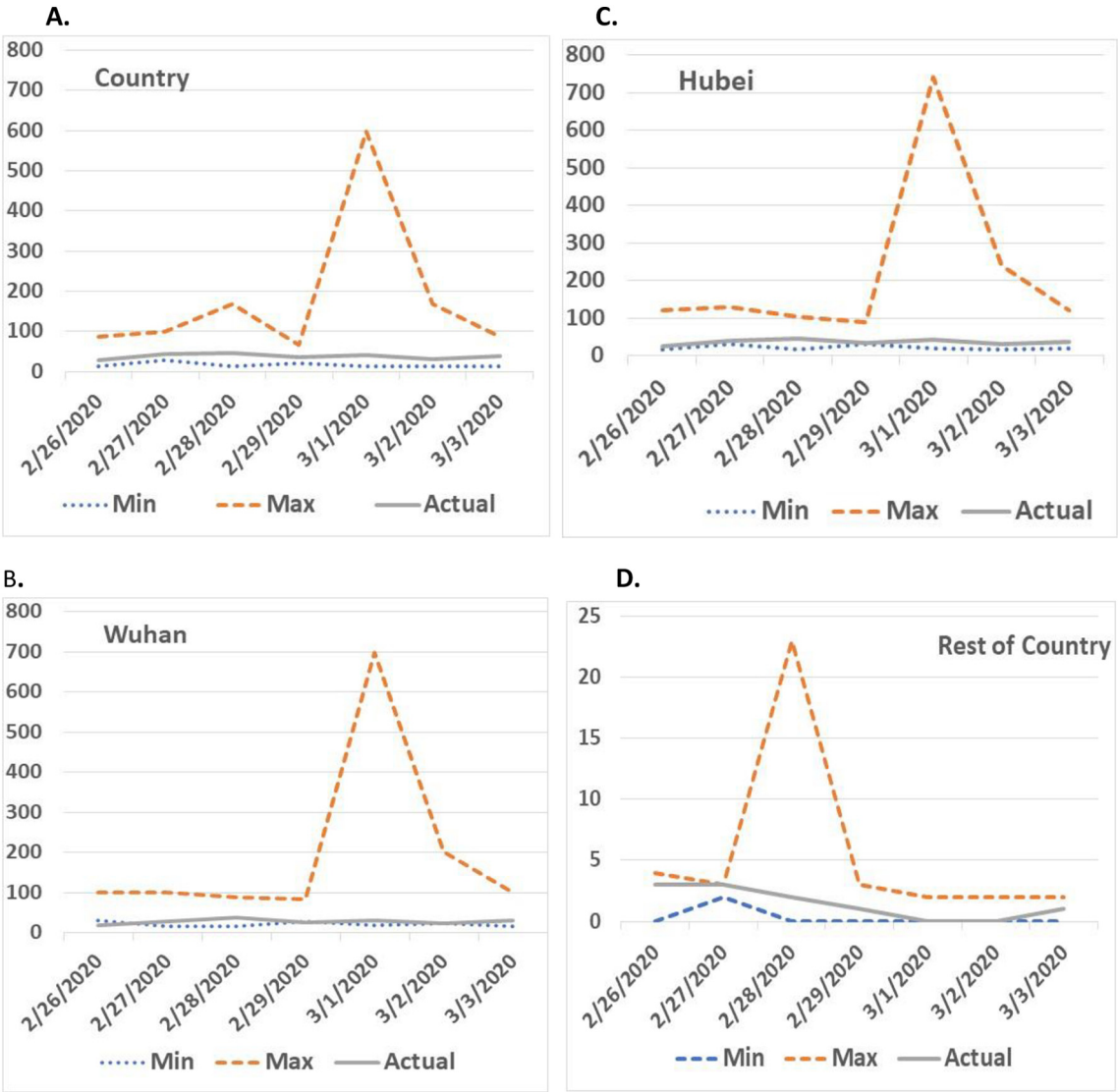


Fig. 4. Comparison between the predicted number of deaths based on PIBA and the actual number of deaths. The blue color represents the estimated minimum number of deaths line. The orange color represents the estimated maximum number of deaths line. The gray line represents the actual number of deaths. Panels A, B, C, and D showed these death numbers in the country, Hubei, Wuhan and the rest of country except Hubei.

we further tested our model using the affected population in South Korea. As shown in Fig. 5, the trend of deaths in South Korea in recent days is consistent with our prediction.

4. Discussion

First, PIBA is capable of accurately estimating the disease mortality and the number of future deaths. This real-time accurate prediction and estimation of disease mortality provide the public, government, and society with more accurate disease information. Based on currently available data that includes patients with varying degrees of severity, the estimated prediction of the mortality rate of COVID-19 is less than 3%, and less than the prior prediction based on limited available data. This finding may ease public concern and panic. Updated scientific findings will be widely disseminated to broaden public awareness and contribute to helping fight COVID-19. The medical, clinical, and research community should strive to publish scientifically rigorous findings related to urgent public health issues. Publishing findings based on the availability of limited data contributes to unnecessary public concern and government action. In this particular case, the first report on the estimation of coronavirus death rate is an applaudable effort. However, it also had the limitations of a skewed dataset that focused on patients who were transferred from local hospitals because of their critical condition while excluding patients with less severe symptoms who remained at local hospitals. As soon as more data are available, we should provide updated reports and introduce improved estimation and prediction algorithms.

This study indicates that as the number of transmissions of 2019nCOV increases among the human population, its lethality will gradually decrease. Indeed, the reasons are not necessarily all because of their reduced toxicity. There may also be improvements in treatments and implementation of early detection methods. Therefore, a real-time estimate of death rate using patient information such as the PIBA method would demonstrate an appreciation of the importance of public and societal awareness.

A critical issue to consider is that if the mortality rate of the COVID-19 in a certain area is relatively high, the COVID-19 in the area is still spreading and endemic. One of the most obvious questions is why the

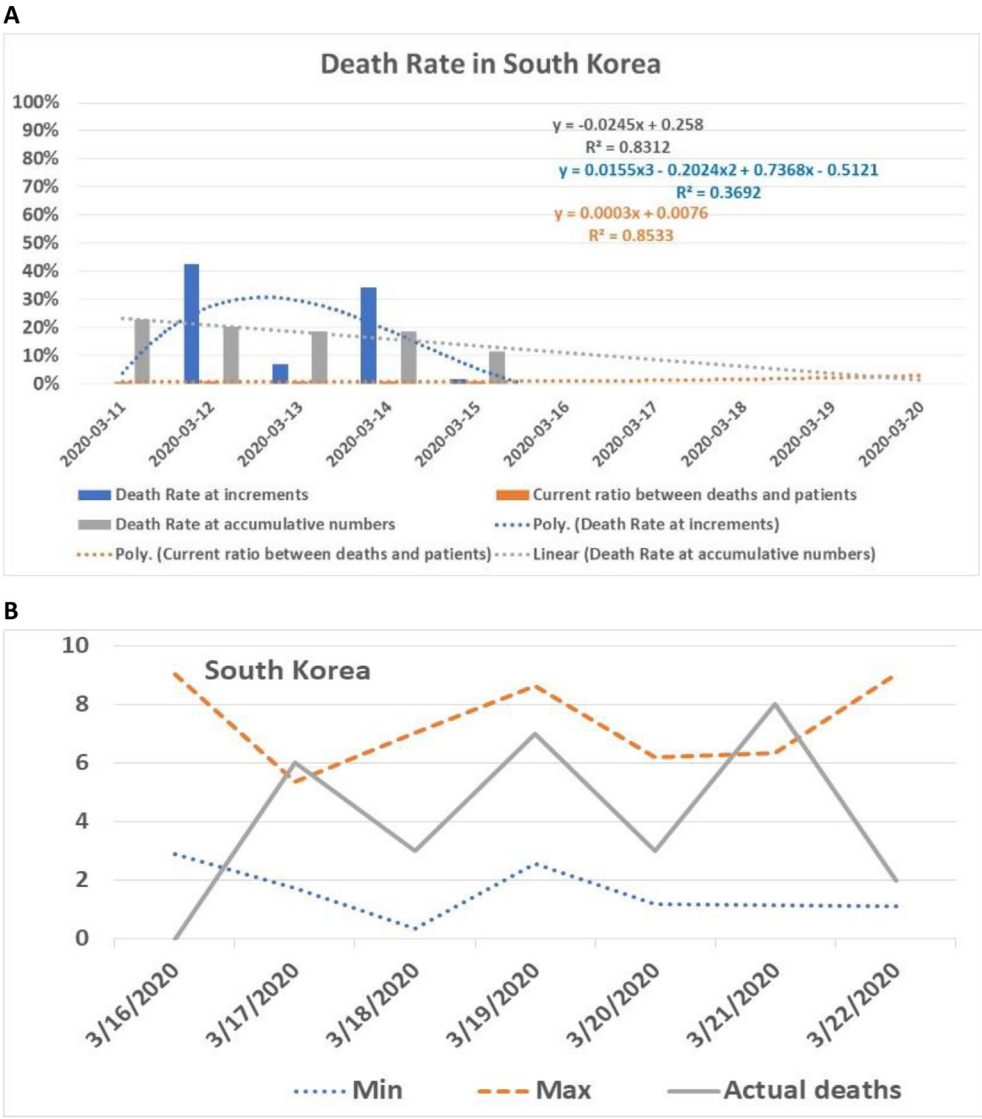


Fig. 5. Test PIBA model using COVID-19 population from South Korea. A. Estimation of death rate in the Korean population using the PIBA method. The blue curve represents the mortality calculated by the actual increase in deaths per lagging day divided by the increase in actual patients on the previous corresponding day. The gray curve represents the total number of deaths per lagging day, divided by the total number of identified actual patients on the corresponding previous day. The orange curve shows the number of deaths per day divided by the total number of patients the same day. The number on the vertical bar represents the death rate, number on the horizontal bar shows the date. B. Comparison between the predicted number of deaths based on PIBA and the actual number of deaths. The blue color represents the estimated minimum number of deaths line. The orange color represents the estimated maximum number of deaths line. The gray line represents the actual deaths.

mortality rate in Wuhan is considerably higher than in other places. Based on our assessment, Wuhan's medical equipment and rescue measures are comparable with other areas in China, and the pathogenicity of the virus is similar. We conclude that there is a large proportion of patients in Wuhan who have mild illness and not been hospitalized at all. Due to the uncertainty of the movement of infected people in the early stages of the onset, these mildly ill people move around in Wuhan unidentified. This problem reminds other parts of the world that if the fatality rate of the COVID-19 is found to be high, a large number of infected people have not been able to be identified or diagnosed. Therefore, the work of controlling and isolating this infected group has not been completed, and the disease is still spreading and circulating in the area.

The data on Heilongjiang Province and Harbin show that, unlike some experts' predictions (cf. https://news.ifeng.com/c/7uHMHXcFHmq), it will occur more intensely in the high-altitude regions with a cold climate, and the mortality rate will be higher. With the development of the generations of 2019-nCoV, its toxicity will gradually weaken, and we expect that the mortality rate in the cold northern regions will not increase, nor will it exceed that in Wuhan or Hubei Province.

Our research has limitations, mainly due to available data. First, the estimation of number of patients from the date of hospital admission or ICU intake to the date of death is based on data from official public websites. Information from 33 individuals was estimated. If the information had been available regarding more patients, the initial estimate would have been more accurate. The second aspect is the accuracy of the number of patients diagnosed and the number of hospitalizations per day. Due to the back and forth revision and correction of the data as announced by the official sources, we are not confident that all the data are error-free; however, we feel that these data as a whole are reliable. The third limitation of the PIBA method is that it depends on accurate patient information at the beginning of the epidemic. Depending on different situations from different countries or regions, this information may or may not be available, or the information may not be accurate.

5. Conclusion

The PIBA model accurately predicted a case fatality of 1.6% for symptomatic patients in China at a very early stage in the Covid-19 pandemic. The model can be generalized to predict case fatality for any infection (including asymptomat), to predict the rate of severe disease, and to predict the death rate for patients who develop severe disease. These early, accurate predictions inform the public, society, and governments to estimate the extent of the disease's harm and to develop suitable strategies.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.138394.

CRediT authorship contribution statement

Lishi Wang: Conceptualization, Formal analysis, Methodology, Writing - original draft, Writing - review & editing,**Jing Li:**Conceptualization, Methodology, Validation, Writing - original draft, Writing - review & editing.**Sumin Guo:**Conceptualization, Validation, Writing - review & editing.**Ning Xie:**Investigation, Writing - review & editing.**Lan Yao:**Investigation, Methodology, Writing - review & editing.**Yanhong Cao:**Investigation, Writing - review & editing.**Sara W. Day:**Conceptualization, Validation, Visualization, Writing - review & editing.**Scott C. Howard:** Conceptualization, Methodology, Validation, Visualization, Writing - review & editing.**J. Carolyn Graff:**Conceptualization, Validation, Visualization, Writing - review & editing.**Tianshu Gu:**Investigation, Writing - review & editing,**Jiafu Ji:**Conceptualization, Resources, Supervision, Validation, Writing - review & editing.**Weikuan Gu:**Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing - original draft, Writing - review & editing.**Dianjun Sun:**Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially supported by funding from merit grant I01 BX000671 to WG from the Department of Veterans Affairs and the Veterans Administration Medical Center in Memphis, TN, USA and grant 90DDUC0058 to CG from U.S. Department of Health and Human Services, Administration for Community Living.

Author contributions

Conceived and designed the experiments: LW, WG, CG, SW, DS, JJ. Performed data searching and collection: LW, JL, LY, TG, SG, NX, WG. Analyzed the data: JL NX, TG, LY, WG. Contributed analysis tools: JL, LY, WG, DS. Wrote the manuscript: LW, JL, SG, NX, YC, SH, SD, CG, LY, WG. Revise and approve and manuscript: All authors.

Human subjects

All the data of patients in this study are from official public websites.

References

Chan, J.F.W., Yuan, S., Kok, K.H., To, K.K.W., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C.C.Y., Poon, R.W.S., Tsoi, H.W., Lo, S.K.F., Chan, K.H., Poon, V.K.M., Chan, W.M., Ip, J.D., Cai, J.P., Cheng, V.C.C., Chen, H., Hui, C.K.M., Yuen, K.Y., 2020. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 395, 514–523. https://doi.org/10.1016/S0140-6736(20)30154-9.

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., Xia, J., Yu, T., Zhang, X., Zhang, L., 2020. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 395, 507–513. https://doi.org/10.1016/S0140-6736(20)30211-7.

Dawood, F.S., Iuliano, A.D., Reed, C., Meltzer, M.I., Shay, D.K., Cheng, P.Y., Bandaranayake, D., Breiman, R.F., Brooks, W.A., Buchy, P., Feikin, D.R., Fowler, K.B., Gordon, A., Hien, N.T., Horby, P., Huang, Q.S., Katz, M.A., Krishnan, A., Lal, R., Montgomery, J.M., Molbak, K., Pebody, R., Presanis, A.M., Razuri, H., Steens, A., Tinoco, Y.O., Wallinga, J., Yu, H., Vong, S., Bresee, J., Widdowson, M.A., 2012. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. Lancet Infect. Dis. 12, 687–695. https://doi.org/10.1016/S1473-3099(12)70121-4.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., Xiao, Y., Gao, H., Guo, L., Xie, J., Wang, G., Jiang, R., Gao, Z., Jin, Q., Wang, J., Cao, B., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395, 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5.

Nicholls, J.M., Poon, L.L., Lee, K.C., Ng, W.F., Lai, S.T., Leung, C.Y., Chu, C.M., Hui, P.K., Mak, K.L., Lim, W., Yan, K.W., Chan, K.H., Tsang, N.C., Guan, Y., Yuen, K.Y., Peiris, J.S.M., 2003. Lung pathology of fatal severe acute respiratory syndrome. Lancet 361, 1773–1778. https://doi.org/10.1016/S0140-6736(03)13413-7.

Schlagenhauf, P., Ashra, H., 2003. Severe acute respiratory syndrome spreads worldwide. Lancet 361, 1017. https://doi.org/10.1016/S0140-6736(03)12843-7.

Viboud, C., Simonsen, L., 2012. Global mortality of 2009 pandemic influenza A H1N1. Lancet Infect. Dis. 12, 651–653. https://doi.org/10.1016/S1473-3099(12)70152-4.

Wang, C., Horby, P.W., Hayden, F.G., Gao, G.F., 2020. A novel coronavirus outbreak of global health concern. Lancet 395, 470–473. https://doi.org/10.1016/S0140-6736(20)30185-9.

WHO Ebola Response Team, Aylward, B., Barboza, P., Bawo, L., et al., 2014. Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections. N. Engl. J. Med. 371, 1481–1495. https://doi.org/10.1056/NEJMoa1411100.

Wu, J.T., Leung, K., Leung, G.M., 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 395, 689–697. https://doi.org/10.1016/S0140-6736(20)30260-9.