Higher Mortality and Infection Rate in a Region or Country Suggest a Longer Time of COVID-19 Disease Pandemic

Tianshu Gu

University of Tennessee Health Science Center

Lan Yao

University of Tennessee Health Science Center

Tong Sun

The first hospital of Qiqihar city

Sara W. Day

University of Tennessee Health Science Center

Scott C. Howard

University of Tennessee Health Science Center

J. Carolyn Graff

University of Tennessee Health Science Center

Yan Jiao

University of Tennessee Health Science Center

Hong Chen

The first hospital of Qiqihar city

Dianjun Sun

Harbin Medical University

Lotfi Aleya

Universite Bourgogne Franche-Comte

Weikuan Gu (✉ wgu@uthsc.edu)

University of Tennessee Health Science Center College of Medicine Memphis <https://orcid.org/0000-0003-1112-8088>

Yongjun Wang

Beijing Tiantan Hospital

Research article

Keywords: COVID-19, Coronavirus, Death rate, Disease duration, Patient

Posted Date: June 13th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-34706/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

In view of the fact that the 2019-nCoV has spread to most countries in the world, it is necessary to make scientific and well-founded predictions of the current pandemic situation caused by the virus worldwide, which are conducive to public, social and government responses that mitigate and appropriately address the pandemic. We collected data from provinces with more than 200 cases in China and from eight other countries. Our analyses showed that the disease duration has no correlation with the number of patients, with $r = 0.184$. The number of deaths was not correlated to the disease duration, with $r = 0.242$. However, a positive correlation between the days of disease duration and infection rate, with a $r = 0.626$. Furthermore, there is a strong positive correlation between the disease duration and total death rate, with a $r = 0.707$. Using death rate of first 25 days, we obtained a positive relationship with a r value of 0.597. Based on the data from first 25 days, the minimum and maximum days of COVID-19 pandemic duration of eight countries was estimated between days of 37 and 114 days.

Background

At present, a new round of disease caused by the new coronavirus has spread throughout the world [1]. When new cases and new deaths are announced each day, people in many countries are asking if this wave of the disease will occur in their country [2, 3]. If it does occur in their country, they are asking on what scale it will occur and they are worried about how to deal with the disease. Even leaders of many countries have spoken in succession on the occurrence and development of the disease and the response strategies adopted. How to evaluate the trend of the outbreak and development of this round of disease in a certain region and country, how to evaluate whether a certain area or a country has been effective in preventing and controlling this round of disease, and whether it can suppress the large-scale infection of the disease crowd are critical areas to address. The publicly reported data on this period of the disease may give us some hints.

Currently, there are two major categories of publicly reported data. The first is China's data, including national and provincial data. This batch of data is reported on a daily basis and includes the number of people affected and the number of deaths [4, 5]. The other is reporting from other countries in the world and includes daily reports of new outbreaks and deaths [6]. Comparing the two types of data, China's data now seems to be more comprehensive and complete. The China's data also covers a longer period of time. So, the question is, what information, insights, and inspiration can be gain from the data reported in China? What valuable hints and references can these data provide to other countries in the world in response to this pandemic? Our study attempts to assess the situation of other countries in the world in response to the disease and to predict the future development (consider using the word, developmental instead of development) trend based on the data that China has published up to this point.

We begin with the data of morbidity and mortality, to explain what different morbidity and mortality rates mean [7]. At present, data both from the reports of different provinces and cities in China and the reports from different countries in the world indicated that infection and lethal rates vary considerably in different

places. Does this mean that changes in the virus itself have caused different infectivity and mortality rates or something else? Our research team believes that the differences in infection rates and lethality in different places are not due to changes in the virus, but to differences in the detection and follow-up of patient groups, statistics, and surveys of infected persons. This information can be used as a scoring standard for the work of a certain region or a country on the prevention and control of this disease.

Methods

Data collection from different regions and countries

We collected data on COVID-19 patients in China from official and publicly available websites. Table 1 listed the web pages of data collection from a variety of regions and cities of China (Supplemental Table 1).

Data from Japan is from <https://newsdigest.jp/pages/coronavirus/>. The data from other countries are collected from the website of Worldometer at <https://www.worldometers.info/coronavirus/>.

Calculation of initial infection rate

Since many provinces and cities in China provide information about the number of people who have been in close contact with the carrier and the cumulative number of confirmed cases, we are able to calculate the infection rate of the of 2019-nCoV in a variety of places. In order to give a reasonable estimation of the initial infection rate, we divided the average number of cases in the first 14 days by the number of persons known to be close contacts to the infected persons. The reason for this is because currently the 14-day period is generally recognized as the incubation period from 2019-nCoV infection to disease onset. In addition, the average number of days from hospitalization to death for the first 41 patients reported in Wuhan was 13 days [7]. Altogether, the 14-day average infection incidence is representative of the initial incidence generally.

The formula is: $I_r = \sum(i_1 \sim i_{14}) / \sum(t_1 \sim t_{14})$, where I_r is the initial infection rate, $i_1 \sim i_{14}$ = infected person from day 1 to day 14, and $t_1 \sim t_{14}$ = total persons in close contact with the carrier or infected person.

Initial death rate estimation

Using the data collected, we have calculated the mortality rate based on the number of confirmed patients and the number of deaths. Based on the previous data that the average number of days from inpatient to death is 13 days and that the maximum days of $\mu + 2\sigma$ is 25 days [7], we used the average death rate of the first 25 days. The method is the direct calculation of death rate by the total number of deaths divided by the total number of patients on the same day. Thus, $D_{rr} = \sum(d_1 \sim d_{25}) / \sum(t_1 \sim t_{25})$, where D_{rr} is the initial death rate, $d_1 \sim d_{25}$ = number of deaths from day 1 to day 25. $T_1 \sim T_{25}$ = total number of inpatients from day 1 to day 25. However, mortality in the other countries are calculated only based on the number of diagnoses and deaths.

Disease patterns and ending data estimation

For the development patterns and the peak period of the COVID-19 in various provinces and cities, we made a case distribution graph according to the number of new cases, and determined the peak date of the pandemic in a certain area according to the peak segment of the graph.

The criteria for the end of an pandemic in a region are based on the absence of any new cases in a region for a consecutive 14 days. Thus, if there are no new cases in a certain area for 14 consecutive days after a certain day, then that day will be the end date of the pandemic in that area. Regions in which new cases have not been found in more than 10 days but less than 14 days are categorized as the regions of potential ending of COVID-19 pandemics. In the calculation of new cases, in this regard, we exclude special cases or outliers. These excluded cases are imported cases and outbreaks of infectious cases in special situations, such as infections of prison populations. In our statistical method, these are not included as new cases in the regions, such as Beijing, Guangdong, and Shanghai.

Statistical analyses

Using the data from different regions of China, we calculated the correlation between the death rate and the infection rate. We compared the total number of close contact persons and the death rate. We then compared the current death rate of different countries other than China. We used the standard criteria to categorize the strength of the correlation [8]. In general, we consider the R value equal or more than 0.7 or -0.7 as significant, meaning strong positive or negative correlation. When the R value is between 0.35 and 0.69 or -0.35 and -0.69, we regard the existence of a correlation but the correlation is not strong. When R values fell between 0 and 0.35 or 0 and -0.35, we regarded these data as no correlation. The data sets of death rate and infection rate were analyzed by one-way ANOVA. $P < 0.05$ was considered statistically significant.

Results

Basic patterns and information of different regions

We collected disease information from the total country of China and 18 provincial regions up to the March 11, 2020 (Supplemental able Table 2). The infection rate, death rate based on PIBA, and daily numbers of patients and deaths are calculated from 15 of these regions (Supplemental Table 3). The data from three regions, Guangdong, Shanghai and Sichuan were not usable for the analyses because of missing data on the total number of persons who were in close contact with the disease careers. In addition, these calculations did not include the patients who came to China after they were infected outside of China. Based on this information, we divided these regions into four groups (Figure 1A):

1). Hubei, a single province as a group because of its large patient population and disease pandemic situation. The initial date of disease has been estimated at early December 2019 or earlier. From the

December 1, 2019 through March 11, 2020, the COVID-19 has been an pandemic in Hubei for 95 days (Fig. 1C). More importantly, currently new patients are still identified every day.

2). The three regions in which the pandemic of COVID-19 has not been ended include Heilongjiang, Beijing and Shandong (Fig. 1D).

3). The four regions that the disease is potentially ending because no new patients were last found in more than 10 days, but not yet to 14 days (Fig. 1D).

4). There are 7 regions that in which the COVID-19 has stopped its status as an pandemic (Fig. 1F). The peak period of disease pandemics of these four groups are similar, with a period of 3 to 4 weeks, approximately from January 20, 2020 to the middle of February 2020 (Fig. 1C-1F). The total number of patients from each region were logitized to its 10^{th} [9] so that Hubei's data can be listed together with others. However, because of its extreme large number of patients, its number in the analysis is over-weighted and was not used in the correlation analysis. The correlation analysis indicated that the disease duration has no correlation with the number of patients, with $r = 0.184$ and $P = 1.7987\text{E-}09$. In addition, the number of deaths was not correlated to the disease duration, with $r = 0.242$ and $P = 7.43019\text{E-}09$.

Infection rate in different groups of pandemic duration of COVID-19.

Based on the collected data, we first analyzed the infection rate. Due to the difference in the availability of data in these provinces, we are counting from February 7, 2020. From this date, every province and city posted available data that we can use to calculate the daily infection rate and mortality. Our results show that the infection rates of these four groups are very different from each other (Fig. 2A). Hubei Province has the highest infection rate (Fig. 2B), followed by provinces where the pandemic is continuing (Fig. 2C). The lowest infection rates are in the groups that the pandemic has potentially ended or ended (Figure 2D-2E). T-test showed that the infection rate of Hubei is significantly different from other three groups, with P values of $1.126\text{E-}27$, $7.51\text{E-}32$, $7.43\text{E-}23$, to not ending, potentially ending and ended groups, respectively. The infection rate of the not ending group is also significantly different from the potentially and ended groups, with P values of $1.69\text{E-}31$, and $8.62\text{E-}26$, respectively.

Daily death rate in different disease development groups

In view of the differences in the daily infection rates among these four groups, we further observed whether the cumulative daily mortality rate was different among these four groups. Daily mortality is simply the number of deaths per day divided by the total number of patients. Similarly, the calculation was started on Feb 7, 2020.

As we predicted, these four groups also have different mortality rates (Fig. 3A). Hubei has the highest mortality rate (Fig. 3B), followed by the group that the COVID-19 pandemic has not ended yet (Fig. 3C). The next group is likely to end in the near future (Fig. 3D). Finally, it is the group whose pandemic has ended (Fig. 3E). The current average mortality rates for these four groups are 0.0392, 0.0178, 0.0075 and

0.0069, respectively. The P values between Hubei and not ending, potentially ending and ended groups are 1.733E-23, 1.61E-24, 7.77E-22, respectively. The not ending group was also different from the potential and ending groups, with P values of 6.73E-25 and 3.48E-25, respectively.

We noticed that the death rate in Hubei province has been dropped from 4% at the beginning to 2% at the end of pandemic. However, the death rate of rest of the country is much lower and stable, mostly lower than 2% (Fig. 3A). Death rate in regions with shortest pandemic period were below 1%.

Initial infection and death rate in different groups

We next compared the infection and death rate of different groups at the beginning of the COVID-19 pandemic. Based on the PIBA method, the maximum days of a patient from inpatients to the death $\mu + 2\sigma$ is 25 days. We therefore used the data from 25 days at the beginning to calculate the initial infection and death rate in different groups. We used the average death rate of the first 25 days. Thus, $Dir = \sum(d1 \sim d25) / \sum(t1 \sim t25)$, where Dir is the initial death rate, $d1 \sim d25$ = number of deaths from day 1 to day 25. $T1 \sim t25$, total number of inpatients from day 1 to day 25.

Our analysis indicated that the initial infection and death rate in Wuhan and the group of not ending of disease are higher than that of other two groups (Figure 4). These data confirm the analysis on the infection and death rate of different groups in which the Wuhan and not ended groups have higher infection and death rates when the potentially ended and ended groups.

Prediction of disease development in different countries.

Based on the differences in infection rate and death rate of different groups, we further investigated the relationship between the days of disease durations and the infection and death rates (Figure 5). Our analysis showed that there is a positive correlation between the days of disease duration and infection rate, with a $R = 0.626$ (Fig. 5A). Furthermore, there is a strong positive correlation between the disease duration and total death rate, with a $R = 0.707$ (Fig. 5B). These data together with the data above clearly demonstrate that the high infection rate and death rate predict a long pandemic period of COVID-19.

However, we realize that, due to the differences in sizes of tested populations and availability and variations in the testing technologies, the mathematic equation for the relationship between infection rate and pandemic duration from China may not be equivalent to other countries.

As to the relation between the death rate and pandemic duration, it may be possible to apply this approach to other countries or regions, not only because of their strong correlation, but also because of the similar criteria and credibility of the death rate among different countries. Accordingly, we attempted to predict the pandemic durations of some countries using their death rates. As the pandemic of disease of COVID-19 is still ongoing in many countries, we question whether the initial data of death rate can be a predictor for the pandemic duration. As mentioned above, based on PIBA method we collected the death rate of $(\mu + 2\sigma)$ 25 days of these regions in China and examined the relationship between the death rate and pandemic duration. We obtained a positive relationship with a R value of 0.597. A line regress(?)

formula is obtained from their data (Fig. 5C). We then obtained the death rate of first 25 days of 8 countries. We calculated the pandemic days of these countries using the regression formula (Fig. 5D). Because China took an extreme measure on the social isolation, while many other countries used the soft measure of social distance, we assume there could be a longer time of pandemic period in other countries. Therefore, we also provide secondary days which are derived by using calculated days to multiple 1.5, as the potential maximum days of pandemic duration. Our calculation suggested that the pandemic days of these countries ranges from 37 to 76 days (Fig. 5D) based on direct calculation and from 56 to 114 days based on the multiplication of 1.5.

Discussion

First, our results show that the high early mortality rate is directly related to the long duration of the CONID-19 pandemic, that is, the higher the mortality rate, the longer the pandemic time. The number of days of appearances of new cases are more in places with high mortality than in areas with low mortality. In places where mortality rates are low at the beginning of the pandemic, the outbreak develops slowly and ends early. New cases start out less frequently in places with lower mortality rates. This result seems to suggest that where the early mortality rate is high, because the actual total number of cases is not completely counted. In other words, some patients are not counted into the total number of patients in places with high mortality rate. This study indicates that as the number of transmissions of 2019nCOV increases among the human population, its lethality will gradually decrease. Certainly, 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 estimation of death rate using patient information such as the PIBA method would contribute to awareness on the part of the public and society [7]. That is to say, a relatively large number of patients have not been found in places with a relatively high initial mortality rate, and this group of people has not been diagnosed and observed early. These people continue to be out of sight of the government and the public. As a result, the disease continues to spread.

Secondly, the positive correlation between mortality and disease development has also been accompanied by a positive correlation between high transmission rates and disease development. In areas with a high initial transmission rate, the number of patients with disease onset increases rapidly and on a larger scale. However, in areas with low infection rates, the number of patients with late onset increases slowly and is relatively small. Similar to the case of mortality, the so-called high initial infection rate may be due to the fact that these monitored populations are basically those who are clearly at risk of becoming ill. Those who did not have a clear initial infection, and who did not have a clear possibility of developing the disease, were not included in the initial test population. However, in the place where the initial infection rate is relatively low, the development of the pandemic disease is slower, and the end is faster. This result basically explains the same situation as the result of mortality. That is, the level of infection, in fact, is about the same in different places. In the initial statistics, it was shown that infectious places did not actually include all infected people. That is to say, there is a large number of people who have had close contact with the source of the disease, but they have not been included in the surveillance

population. This has also caused individual cases to continue to spread and cause the disease in these areas to continue to develop. The situation is more serious, and the pandemic end date is very late.

From the above-mentioned mortality and situation in different parts of China, we have evaluated the development of COVID-19 pandemics in different countries. According to the reported situation of the virus in different regions of China, there is currently no reason to believe that mutations in viruses in countries with relatively high lethality have caused increased toxicity. In other words, at present, in countries that report relatively high mortality rates, the real situation is that these countries have not counted the number of all patients with COVID-19, or they have not detected all infected people. Because there are a significant number of patients in these countries with minor infections are asymptomatic, or are not isolated and monitored or under controlled, development of disease in these countries will continue. While the disease currently reported in countries that have a relatively low rate of death, we believe that the development of the disease in these countries has been basically under controlled, because most of infected people or infection source, if not all, have been identified and have been stopped from continue infecting other people.

Due to the back and forth revision and correction of the data as announced by the official sources, we are not certain that all the data are error-free; however, we feel that these data as a whole are reliable. The estimation of the days of disease duration is based on the fact that countries have taken measures of social distancing and wearing personal protection equipment. If these measures are stopped early when the disease is still spreading, the COVID may continue its pandemic for a very long period of time.

Conclusions

In summary, places beginning with a high mortality of COVID-19 pandemic means a long painful period of the disease pandemic will come. In places where the mortality rate is relatively high, health professionals should be prepared for a longer period of fighting with this pandemic. Any carelessness or loosing of control measures may cause inestimable losses of lives in these places of high mortality.

Abbreviations

2019nCoV

2019 novel coronavirus

ANOVA

Analysis of Variance.

COVID-19

Coronavirus disease 2019

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All the data are available at the public databases as provided in the manuscript.

Funding

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.

Competing interests

The authors declare no competing financial interests.

Authors' Contributions

YJ, WG, YW, LA, HC, DS conceived and designed the experiments. TG, LY, TS, WG performed data searching and collection. TG, LY, TS, YJ, WG analyzed the data. WG, YW, HC, DS contributed analysis tools. TG, LY, SD, LA, CG, SH, WG wrote the manuscript. All authors revise and approve and manuscript.

Acknowledgments

Not applicable

References

1. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet*. 2020; pii: S0140-6736(20)30185-9.
2. Helmy YA, Fawzy M, Elaswad A, Sobieh A, Kenney SP, Shehata AA. The COVID-19 Pandemic: A Comprehensive Review of Taxonomy, Genetics, Epidemiology, Diagnosis, Treatment, and Control. *J Clin Med*. 2020 Apr 24;9(4). pii: E1225.
3. Xu J, Pan T, Qi X, et al. Increased mortality of acute respiratory distress syndrome was associated with high levels of plasma phenylalanine. *Respir Res*. 2020;21:99. <https://doi.org/10.1186/s12931-020-01364-6>.
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. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020; pii: S0140-6736(20)30183-5.
5. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*. 2020; pii: S0140-6736(20)30260-9.
6. World Health Organization. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.
7. Wang L, Li J, Guo S, et al. Real-time estimation and prediction of mortality caused by COVID-19 with patient information based algorithm. *Sci Total Environ*. 2020;727:138394. <https://doi.org/10.1016/j.scitotenv.2020.138394>.
8. Wang L, Cao Y, Ren M, Chen A, Cui J, Sun D, Gu W. Sex Differences in Hazard Ratio During Drug Treatment of Non-small-cell Lung Cancer in Major Clinical Trials: A Focused Data Review and Meta-analysis. *Clin Ther*. 2017 Jan;39(1):34–54.
9. Carvajal G, Roser DJ, Sisson SA, Keegan A, Khan SJ. Modelling pathogen log10 reduction values achieved by activated sludge treatment using naïve and semi naïve Bayes network models. *Water Res*. 2015;85:304–15.

Figures

Figure 1

Disease patterns of four categories of regions of COVID-19 pandemics. 1A-B. estimated days of disease durations and log10 of patient numbers for Hubei, three provincial regions where the disease has not ended (in purple color), four regions of potential ending of pandemics (in blue color), and 6 regions of ended pandemics. 1C. Disease pattern in Hubei province. 1D. Disease patterns of regions that the disease has not yet come to the end. 1E. The disease patterns of regions that are potentially ending the disease pandemic. 1F. Patterns of regions that the disease pandemic has come to the end.

Figure 2

Patterns of infection rates of four categories of regions of COVID-19 pandemics. 2A. Overall comparison among four groups. 2B. Daily disease infection rates in Hubei province. 2C. Daily disease infection rates of regions that the disease has not yet come to the end. 2D. Daily disease infection rates of regions that are potentially ending the disease pandemic. 2E. Daily disease infection rates of regions that the disease pandemic has come to the end. Consider adding a statement about what the vertical axis represents.

Figure 3

Patterns of death rates of four categories of regions of COVID-19 pandemics. 3A. Overall comparison among four groups. 3B. Daily death rates in Hubei province. 3C. Daily death rates of regions that the disease has not yet come to the end. 3D. Daily death rates of regions that are potentially ending the disease pandemic. 3E. Daily death rates of regions that the disease pandemic has ended.

Figure 4

The average infection and death rates by different disease duration groups. 4A. The infection rate (left) and death rate (right) of four groups. 4B. The infection rate and death rate of Hubei province. 4C. The infection rate and death rate of regions in group of not ended. 4D. The infection rate and death rate of regions in group of potentially ended. 4E. The infection rate and death rate of regions in group of pandemic ended.

Figure 5

Pandemic days and disease infection and death rate of COVID-19. 5A. The relationship between death rate and the pandemic duration in different regions in China. The numbers on the vertical axis represent the days of disease duration; the numbers on the horizontal axis show death rate. 5B. The relationship between patient numbers and the pandemic duration in China. The numbers on the vertical axis represent the Log10 of patient numbers; the numbers on the horizontal axis show days of disease duration. 5C. The relationship between death rate of the first 25 days and the pandemic duration. The number on the vertical bar represent the days of disease duration; the number on the horizontal bar shows death rate. 5D. Estimated COVID-19 duration in different countries. Each country has two estimated numbers. The blue bar represents the minimum days while the green bar represents the potential maximum days.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalTable1.Sourcesofdata.docx](#)
- [SupplementalTable2.Dataofprovinceandcities.xlsx](#)
- [SupplementalTable3.Dataafteroutliers.xlsx](#)
- [SupplementalTable4.Datafromcountries.xlsx](#)