



Understanding epidemic data and statistics: A case study of COVID-19

Amirhoshang Hoseinpour Dehkordi¹ | Majid Alizadeh² | Pegah Derakhshan³ | Peyman Babazadeh⁴ | Arash Jahandideh⁵

¹School of Computer Science, Institute for Research in Fundamental Sciences, Tehran, Iran

²School of Mathematics, Statistics and Computer Science, College of Science, University of Tehran, Tehran, Iran

³School of Medicine, Iran University of Medical Sciences, Tehran, Iran

⁴School of Engineering, Islamic Azad University Central Tehran Branch, Tehran, Iran

⁵Adhesive and Resin Department, Iran Polymer and Petrochemical Institute (IPPI), Tehran, Iran

Correspondence

Majid Alizadeh, School of Mathematics, Statistics and Computer Science, College of Science, University of Tehran, Tehran 14155-6455, Iran.

Email: majidalizadeh@ut.ac.ir

Abstract

The 2019 novel-coronavirus (COVID-19) has affected 181 countries with approximately 1197405 confirmed cases (by 5th April). Understanding the transmission dynamics of the infection in each country which got affected on a daily basis and evaluating the effectiveness of control policies are critical for our further actions. To date, the statistics of COVID-19 reported cases show that more than 80% of infected are mild cases of disease, around 14% of infected have severe complications, and about 5% are categorized as critical disease victims. Today's report (5th April 2020; daily updates in the prepared website) shows that the confirmed cases of COVID-19 in the United States, Spain, Italy, and Germany are 308850, 126168, 124632, and 96092, respectively. Calculating the total case fatality rate (CFR) of Italy (4th April 2020), about 13.3% of confirmed cases have passed away. Compared with South Korea's rate of 1.8% (seven times lower than Italy) and China's 4% (69% lower than Italy), the CFR of Italy is too high. Some effective policies that yielded significant changes in the trend of cases were the lockdown policy in China, Italy, and Spain (the effect observed after some days), the shutdown of all non-essential companies in Hubei (the effect observed after 5 days), combined policy in South Korea, and reducing working hours in Iran.

KEY WORDS

biostatistics and bioinformatics, coronavirus, data processing, epidemiology, infection, test statistics, virus classification

1 | INTRODUCTION

Human coronaviruses (HCoV) which cause gastrointestinal and respiratory tract infections were first introduced by the discovery of HCoV-229E and HCoV-OC43, from the nasal cavities of human patients with the common cold, in the 1960s.^{1,2} Other discovered human coronaviruses, which have caused serious respiratory tract infections, include SARS-CoV (in 2003), HCoV NL63 (in 2004), HKU1 (in 2005), MERS-CoV (in 2012), and the latest one SARS-CoV-2 (in 2019) resulting in coronavirus disease (COVID-19).^{3,4} The name originates from the morphology of the virus when viewed under 2D transmission electron microscopy (large pleomorphic spherical particles with the bulbous surface) and stems from the Latin word

"corona," meaning "crown."⁵ Concerning the risk factor, HCoVs vary significantly from the relatively harmless ones (ie, the common cold) to the most lethal ones (MERS-CoV, with more than 30% mortality rate in the infected).⁶ CoVs spread during cold seasons and cause colds with major symptoms, that is, fever, sore throat, and less commonly pneumonia and bronchitis for the more aggressive strains. To date, there are no vaccines or antiviral drugs capable of preventing or treating HCoV infections.⁶⁻⁸

To date, several outbreaks of coronavirus-related diseases have been reported. Severe acute respiratory syndrome (SARS) was the first coronavirus-related outbreak that started in Guangdong, China, in November 2002, and spread to a total of 29 territories, including Hong Kong, Taiwan, Canada, Singapore,

Vietnam, and the United States, within 9 months. It infected a total of 8098 people and killed 774 worldwide.⁹ The second coronavirus-related outbreak happened in the Middle East in April 2012, officially named Middle East respiratory syndrome (MERS). This virus was first identified in a patient from Saudi Arabia, and later, MERS affected several other countries, including Saudi Arabia, South Korea, the United Arab Emirates, Jordan, Qatar, and Oman. Overall, the virus affected 24 countries, with over 1000 cases and over 400 deaths.¹⁰ The outbreak of MERS happened again in South Korea, supposedly from a traveler from the Middle East. It happened during May and July 2015 and infected a total of 186 individuals, with a death toll of 36.¹¹ After 3 years in August 2018, the next MERS outbreak happened in countries of the Arabian Peninsula and resulted in almost 147 infected people and the death of 47. The MERS outbreak had been reported in Saudi Arabia, the United Kingdom, and South Korea.

In December 2019, a pneumonia outbreak was reported in Wuhan, China, and on 31st December, it was attributed to a new strain of HCoV, first named as 2019-nCoV by the World Health Organization (WHO) and later renamed to SARS-CoV-2 by the International Committee on Taxonomy of Viruses. Almost 2 weeks later, on 11th January 2020, Chinese state media reported the first fatality from the newly discovered virus, which led to the infection of dozens more. Until 20th January, multiple countries reported their first cases, including Japan, South Korea, and Thailand. The first confirmed case in the United States came the very next day in the Washington State. As the spread continued, coronavirus presence was confirmed throughout the month of February in the Philippines (2nd February), France (14th February), Iran (21st February), and as reports started in Italy on 23rd February; many more European countries followed the suit, reporting their first confirmed cases. To date, the coronavirus has affected 181 (by 5th April) countries with more than 1100000 confirmed cases and around 65000 people have lost their lives. With the United States, Spain, Italy, and Germany experiencing the worst cases of outbreaks and showing no sign of alleviation, the 2019-2020 outbreak of COVID-19 is now officially recognized as a pandemic by WHO. An outbreak or epidemic often refers to a sudden increase in the occurrence of infectious disease, in a particular time and place. Pandemics are near-global epidemic outbreaks, where multiple countries across the world are involved.¹²

The mentioned rapid trend of spread prompts a lot of concerns and questions such as "How fast is the virus spreading?," "Which policies or efforts could control the disease better?," and "What is the main difference of COVID-19 outbreak with previous epidemics?" Fortunately, the daily case detection changes are available and can be tracked almost in real time on the website provided by authors (<http://iuwa.ir/corona/>). The aim of this study is to provide the transmission trend from China to other countries and to report the daily confirmed cases, fatality causes, and surveillance in every country from the first day of the outbreak until 5th April and, also, to evaluate the effect of each government policy in controlling the outbreak of COVID-19.

2 | METHODS

2.1 | Basic statistics

COVID-19 has currently spread to 181 countries and most national authorities have failed to keep its rapid spread contained.¹³ WHO reports that it began in Wuhan city, located in Hubei province of China, first reported on 21st January.¹⁴ COVID-19 categorizes in three distinctions concerning it is infected host's severity of disease.^{15,16} To date, the statistics of its reported cases show more than 80% of infected had a mild case of disease, whereas around 14% of infected experienced a severe one, suffering from breathlessness and pneumonia. And about 5% are categorized as critical disease patients, their symptoms include septic shock, respiratory failure, and the failure of more than one organ.

Reports on 5th April 2020 show that the United States, Italy, and China have the most confirmed fatal and also recovered cases. The order of confirmed cases after the United States is followed by Spain, Italy, and Germany, which can be seen in Table 1. Confirmed death cases caused by COVID-19 are also observed in 140 different countries (by 5th April), lead in numbers by Italy, Spain, the United States, and France. About 24% of death cases, 26% of confirmed cases, and 32% of recovered cases located in Italy, the United States, and China, respectively, are also shown in Table 1. The overall statistics since 5th April state that there are 1197405 confirmed, 64606 deaths, and 243572 recovered cases, overall. Figure 1 also shows that the COVID-19 spread exists in all continents.

2.2 | Finding linear relations

There is not much known at the moment about COVID-19, so there is a small amount of data about its comprehensive effects and behaviors. In this study, relations are assumed to be linear, when, initially the drawn plot shows obvious linear relations, and later, the

TABLE 1 Top 10 total confirmed, deaths, and recovered cases for 5th April

Country/region	Confirmed	Deaths	Recovered
United States	308850	8407	14652
Spain	126168	11947	34219
Italy	124632	15362	20996
Germany	96092	1444	26400
France	90848	7574	15572
China	82543	3330	76946
Iran	55743	3452	19736
United Kingdom	42477	4320	215
Turkey	23934	501	786
Switzerland	20505	666	6415

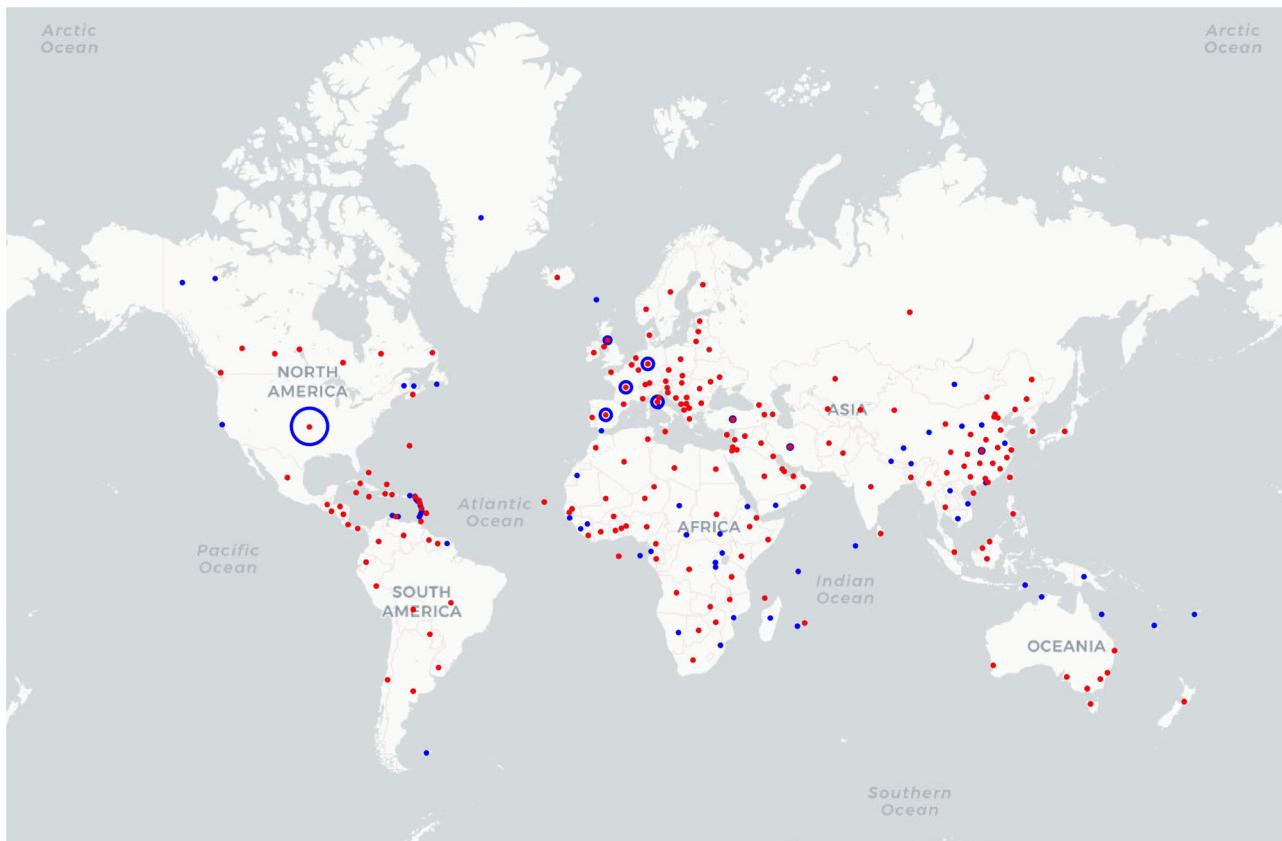


FIGURE 1 Transmission of coronavirus disease 2019-2020 (COVID-19); blue nodes represent regions with confirmed COVID-19 cases, and red nodes represent the regions with COVID-19 causes deaths

fitted linear regression line shows a small enough error to preserve the values given and the linear regression results can be interpreted with relative ease. Besides, fitting regression lines with higher order causes overfitting, resulting from the amount of data. There is no

evidence yet about the relationship of other conditions with the outbreak and its case fatality rate (CFR), so by using linear regression line, policies and behaviors can be compared. In the prediction cases, by using linear regression, we can compare future

Algorithm 1 BRG (Best Linear Regression)

To find linear relation with minimum error of two arrays \mathcal{X}, \mathcal{Y} of numbers, by shifting \mathcal{Y} (maximum number of steps are stp)

```

1: function BRG( $\mathcal{X}, \mathcal{Y}, stp, \eta$ )
2:    $len \leftarrow \text{length}(\mathcal{X}) - stp$                                  $\triangleright \mathcal{N}, x_0, \eta$  are NN, input point and neighborhood function respectively
3:    $min\_val \leftarrow \inf$ 
4:    $best\_date \leftarrow 0$ 
5:    $rma \leftarrow []$                                                $\triangleright \text{length}(\mathcal{X})=\text{length}(\mathcal{Y})$ 
6:    $\mathcal{X}' \leftarrow \text{first } len \text{ elements of } \mathcal{X}$ 
7:   for  $i \in \{1 \dots stp\}$  do
8:      $\mathcal{Y}' \leftarrow i-th \ len \text{ elements of } \mathcal{Y}$ 
9:      $lm \leftarrow \text{LinearRegression}()$ 
10:     $lm.fit(\mathcal{X}', \mathcal{Y}')$                                           $\triangleright$  Fitting best regression line to  $\mathcal{X}', \mathcal{Y}'$ 
11:     $mae \leftarrow lm.meanAbsoluteError$ 
12:     $rma.append(mae)$ 
13:    if  $mae < min\_val$  then
14:       $min\_val \leftarrow mae$ 
15:       $best\_date \leftarrow i$ 
16:       $best\_lm \leftarrow lm$ 
17:
18:   return  $best\_date, best\_lm$ 
```

trends of countries in earlier stages, with the ones in later stages. By considering the above-mentioned statements, we will find the best linear relation between arrays of data. In some cases, the linear relation can be observed but it may exhibit linear relation with some date shift of others (ie, death cases should have a linear relation with earlier values of confirmed cases, given the fact that it should take time from confirmation to death).

CFR could be calculated by the following formula:

$$\text{CFR} = \frac{D_{\text{death}}(T)}{C_{\text{confirmed}}(T - dt)},$$

where D_{death} and $C_{\text{confirmed}}$ functions calculate the value of death cases and confirmed cases at that date, T is the date we want to inspect the CFR, and dt is the mean duration of confirmed to death.

2.3 | Global daily statistics

Figure 2A shows the global confirmed deaths and recovered cases' trend for COVID-19 from 22nd January to 5th April 2020.

Death cases are excessively lower than the confirmed ones, so we normalized (by dividing the value of confirmed deaths and recovered cases to their maximum respectively) it in Figure 2C to investigate all three trends of cases. For the confirmed cases, there is a huge increase since 11th February, the increased tones down from 11th February to the next day. Furthermore, on 13th February, another sharp increase is reported. It can be observed in Figure 2B which shows new cases for each day (and normalized in Figure 2D). The most reliable speculation for this jump is that on that day, China (the country with the most confirmed cases), for the first time, reported the clinically diagnosed cases in addition to laboratory-confirmed cases,¹⁷ in which 13332 clinically diagnosed cases were added to 1148 laboratory-confirmed ones. Since then, China has kept the same reporting method for the confirmed cases. On 23rd and 24th February, confirmed new cases started to increase again. As shown in Figure 2A, the reduction trend is continued (approximately) and the cause of the increase was other countries' growing numbers. So, for more accurate analysis, each country will be investigated separately.

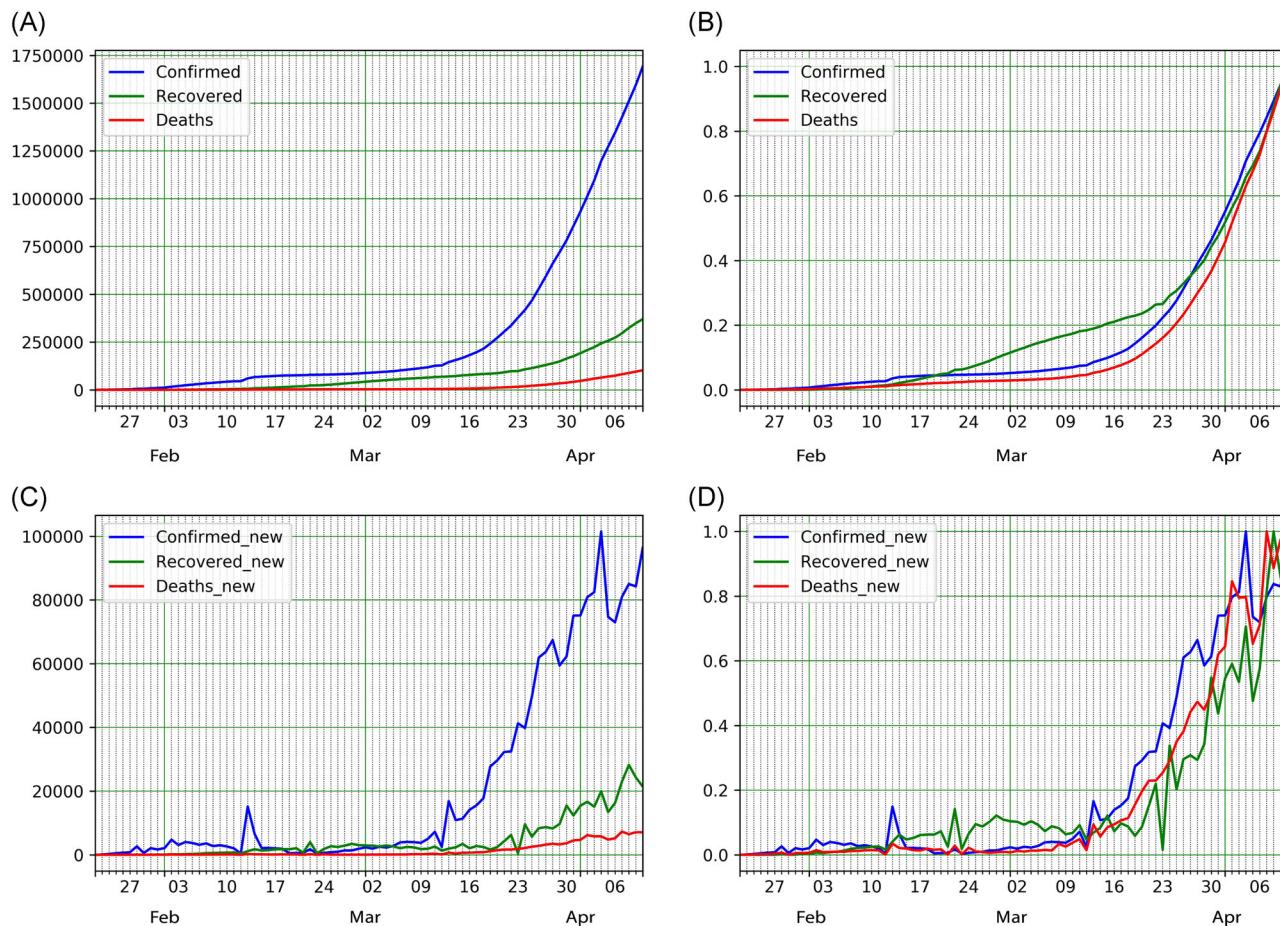


FIGURE 2 COVID-19 global epidemic data and statistics of (A) confirmed, recovered, and death COVID-19 cases, (B) normalized confirmed, recovered, and death COVID-19 cases, (C) new confirmed, recovered, and death COVID-19 cases, and (D) normalized new confirmed, recovered, and death COVID-19 cases

3 | RESULTS

3.1 | China: The mainland

Wuhan city located in Hubei province is reported to be the origin of COVID-19. On 23rd January 2020, a lockdown in Wuhan and other cities in Hubei was implemented to control the outbreak of the COVID-19. A total of 12 other cities in Hubei, consisting of Huangshi, Jingzhou, Yichang, Xiaogan, Jingmen, Suizhou, Xianning, Qianjiang, Xiantao, Shiyan, Tianmen, and Enshi, restricted any form of transportation by the end of 24th January. These decisions were made to prevent the further expansion of COVID-19.

3.1.1 | Confirmed cases

As measured by Backer et al.¹⁸ the incubation period for Wuhan travelers estimated from 2.1 to 11.1 days (the mean incubation period was estimated to be 6.4 days), and also generally the mean incubation period was estimated at 5.2 days which distributed in intervals of 4.1 to 7.0 with 95% confidence.¹⁹ By adding these two values, 11.6 days after 24th January (4th and 5th February), the effects should be manifesting.

Figure 3B depicted the new daily confirmed cases of China outside of Hubei. The peak of the plot is located on 13th February and the daily new cases reduce afterward. This reduction shows that lockdown plays a serious role in the further reduction of cases in China (excluding Hubei province). Even though there is no reason to argue the lockdown's positive impact on Hubei itself, the decrease in new confirmed cases (13th February increase's rationale was described in the previous section) shows that emergency circumstances and movement limitations yield positive results in the reduction of confirmed cases from 10th February. In 13th February 2020, the Chinese government issued a shutdown of all nonessential companies, including manufacturing plants, in Hubei province. Five days later, on 18th February, a drop of new cases could be observed (Figure 3B). Finally, the confirmed new cases in China were negligible from 1st March.

3.1.2 | Deaths

The number of deaths is far lower than the confirmed cases. So, to investigate the relation of confirmed cases trend with the CFR, the normalized plot will be investigated. By observing normalized Hubei province plot of confirmed deaths and recovered cases in Figure 3K, it can be seen that the CFR trend behaves the same as the confirmation, with a shift (in date). Visually, it could be seen that the value of shift in date varies and increases during this time. In earlier cases, the period of confirmation cases leading to death was shorter. It seems, one reason for this variation is that confirmed cases consist only of just laboratory cases and, by adding clinically diagnosed cases (which existed before but did not count beforehand), the time of confirmation to death increases. In other words, the number of confirmed cases gets closer to the real

value, and the cases are announced sooner than they did before. Other possible reasons include the advancements in developing treatments, further delaying fatal cases, and the increase in public awareness, as more people with possible signs of infection come forward to be diagnosed. To estimate the expected value of confirming a case up to the death stage, assuming a linear relationship between the death and confirmed rates, we draw a linear regression line for confirmed and death cases' value, each time increasing the duration and finding the mean absolute error (MAE) of the regression line. Normally, by increasing the duration, following the reduction in investigated points, the MAE is reduced. However, if there is an obvious relation between these two parameters, at the point which they had a correlating relation, MAE will begin to increase (Figure 4A). Wang et al²⁰ estimated the time from the appearance of first symptoms to dyspnea was 5.0 days, to hospital admission 7.0 days, and acute respiratory distress syndrome 8.0 days. Another study found the median days from the first symptom to death as 14 (range 6-41) days.²¹ As seen in Figure 4C, this value is about 11 days in China (excluding Hubei province), Figure 4E depicts value 9 for Hubei and Figure 4A depicts value as 9 days for China. Assuming that the hospital admission is on the same day of confirmation (or a day before confirmation), the mean total of 14 days from the first symptom could be approved in.²¹ By finding the mean day from confirmation to death, it is possible to find out the CFR in China, Hubei, and China (excluding Hubei province) (Algorithm 1 reports best shifting date value and best linear regression line). To find CFR of 5th April for Hubei, confirmed cases on 5th April should be divided by the death cases 9 days prior (which is 27th March) returning 4.7%. Calculating CFR for China (excluding Hubei) till 5th April follows as confirmed cases on the same day divided by death cases of the previous 11 days that date equals 0.9%. Finally, for China's CFR on 5th April, the confirmed cases dated 5th April should be divided by deceased ones of the previous 9 days (27th March) yield 4%.

3.1.3 | Recovered

Recovered cases are defined as active cases-patients recovered after a certain amount of time, with its trend seen in Figure 3J-L. By comparing recovered cases with confirmed ones, a relation is observed after date shifts. Unlike death cases, recovered cases' shifts are initially longer and reduce over time. The assumed reasoning is that as time passes, more medical treatments develop, healthcare providers gain more experience in handling patients' care and as more people are informed, increasing numbers of them get checked in hospitals at the early stages of their disease, resulting in an even more efficient treatment. However, this reduction does not break the linear relation between confirmed cases and recovered ones enough to be significant. To find the mean date shift between confirmed cases and recovered ones, we apply a linear regression line to different dates by shifting them back until a first local minimum MAE is found (Algorithm 1). Hubei province's recovered mean duration value found is 20 days as shown in Figure 4F, the same value for China (excluding Hubei province) is 20 days as shown in Figure 4D, and

finally it is 20 days for China in Figure 4B. To find out the ratio of the recovered cases of Hubei province on 5th April, recovered cases of 5th April are divided by confirmed cases of 20 days prior resulting in 94%. The same calculation applies for China (excluding Hubei) 99%, and finally for China 95%.

3.2 | South Korea: Fast reaction confirmed cases

First confirmed cases of COVID-19 were observed on 20th January in South Korea, but the outbreak started around 18th February

(29 days later), its death and recovered cases' trend is shown in Figure 5A after normalizing, also the new cases' real-valued and normal form are also found in Figure 5B. Newly confirmed, the deceased and recovered cases are also depicted in Figure 5C and normalized in Figure 5D. In Figure 5E, first days of four country's outbreak are compared. New confirmed and death cases are also depicted in Figure 5G,H.

Being one of the first countries reporting the outbreak of COVID-19, the first reported case was on 20th January. However, no outbreak was observed by 18th February, after which there was a significant increase in the number of confirmed cases.

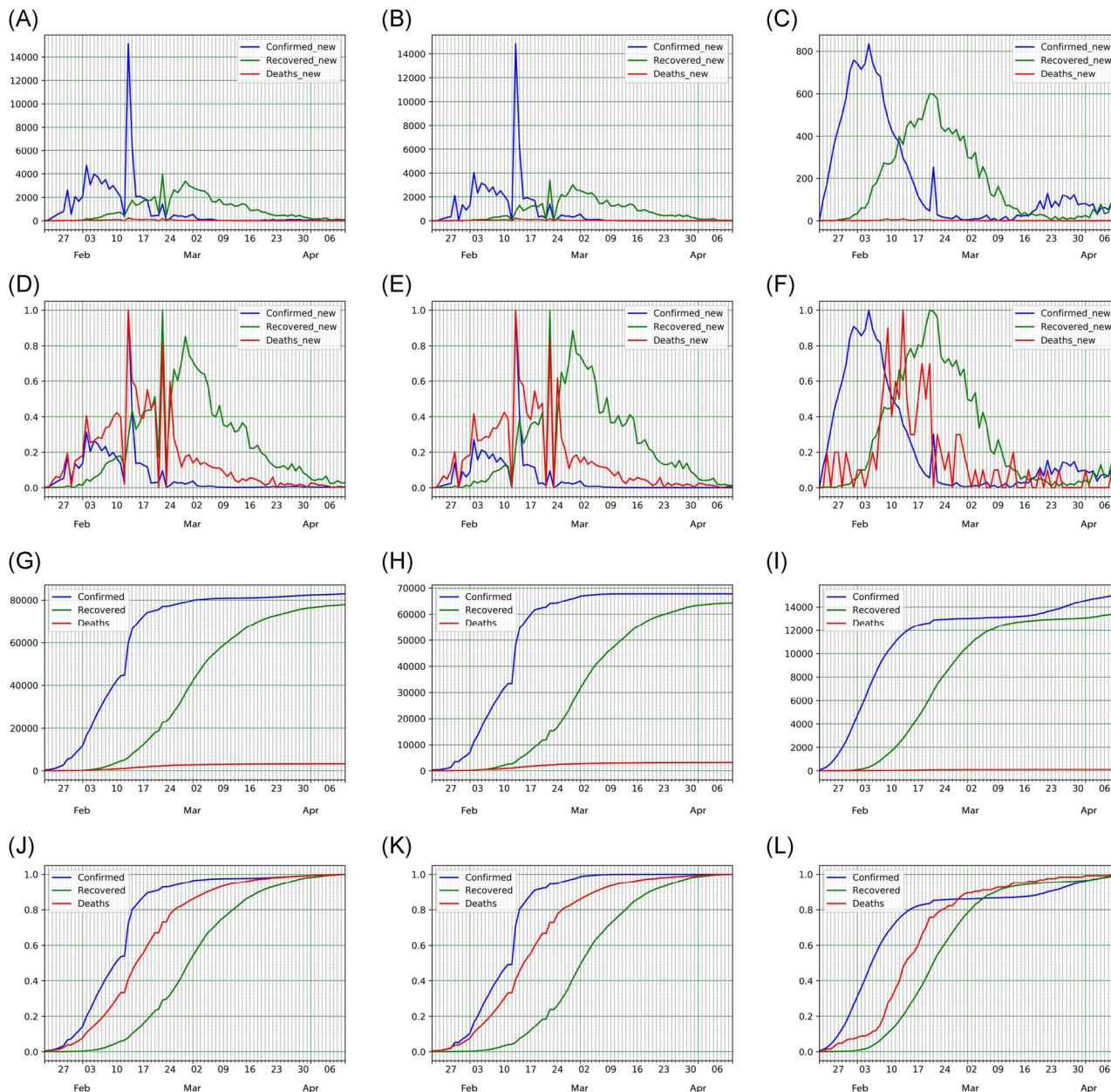


FIGURE 3 COVID-19 epidemic data and statistics of China (mainland), Hubei province, and China excluding Hubei. A-C, New (confirmed, recovered, and death COVID-19 cases). D-F, Normal data (for confirmed, recovered, and death COVID-19 cases). G-I, Confirmed and death cases. J-L, Normal data of confirmed and death cases

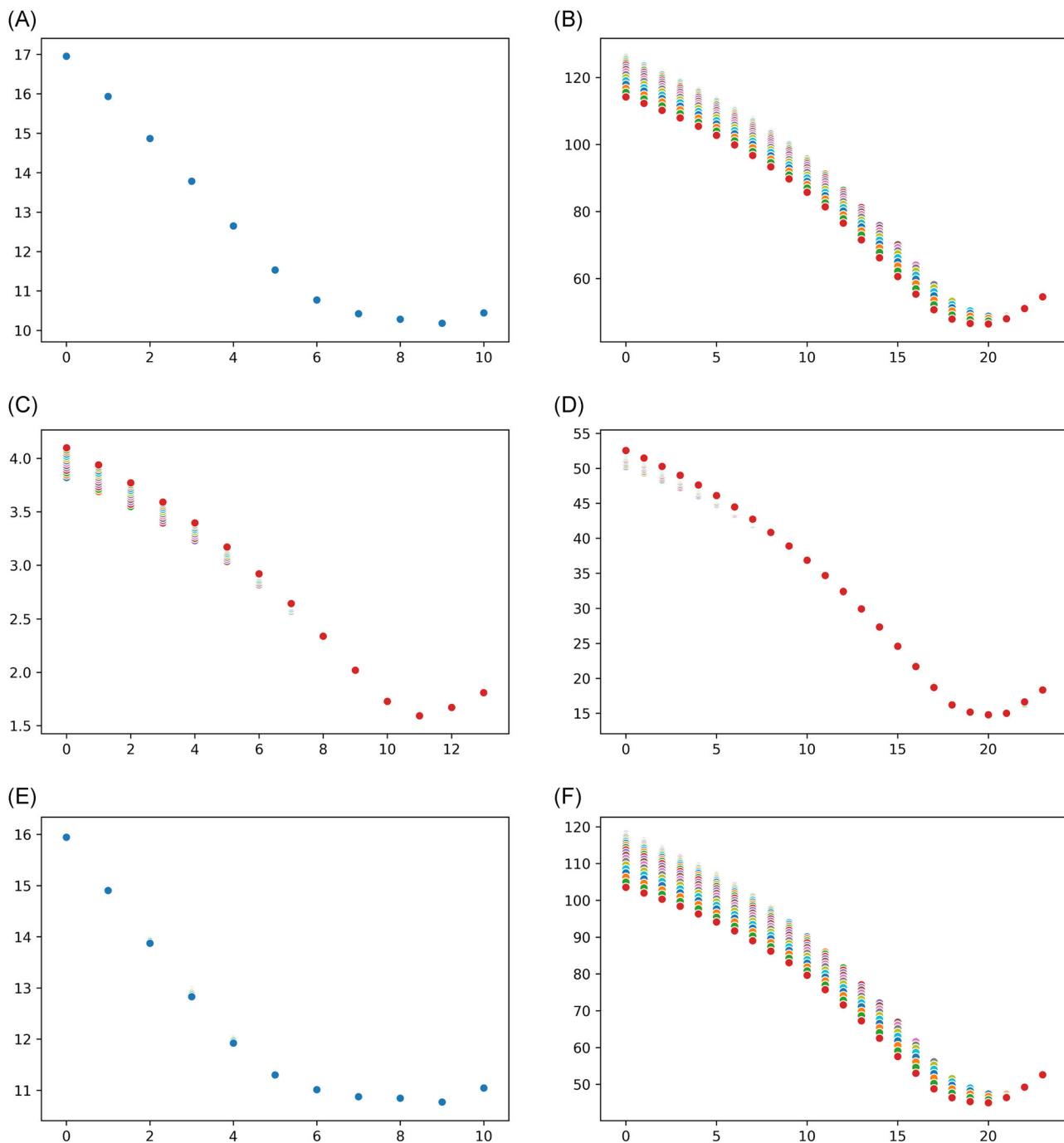


FIGURE 4 Confirmed death and confirmed recovered regression mean absolute error data for COVID-19 transmission through (A,B) China (mainland), (C,D) Hubei province, and (E,F) China excluding Hubei

A comparison between the growth trend in numbers from South Korea from this date with China's is depicted in Figure 5E with the growth rate of South Korea being more than China's. Followed by a faster reduction in South Korea, patterns are approximately the same between the two. COVID-19 spreads from humans to humans;²² therefore, in addition to isolation of people, social avoidance and quarantine policies, and faster detection of infected cases should reduce further growth. Each infected individual, by having contact with others directly or

by proxy (in other words by activity), could infect several people, so to reduce the odds of transmission of the virus, faster detection of infected individuals could play a key role alongside lockdown strategies.

On the basis of the Korean Centers for Disease Control and Prevention, as of 8th March 2020, a total number of 188518 cases have been screened and tested for HCoV infection.²³ More than 10000 tests were conducted on 8th March, and in just 2 days, this number reached to 210144 (indicating more

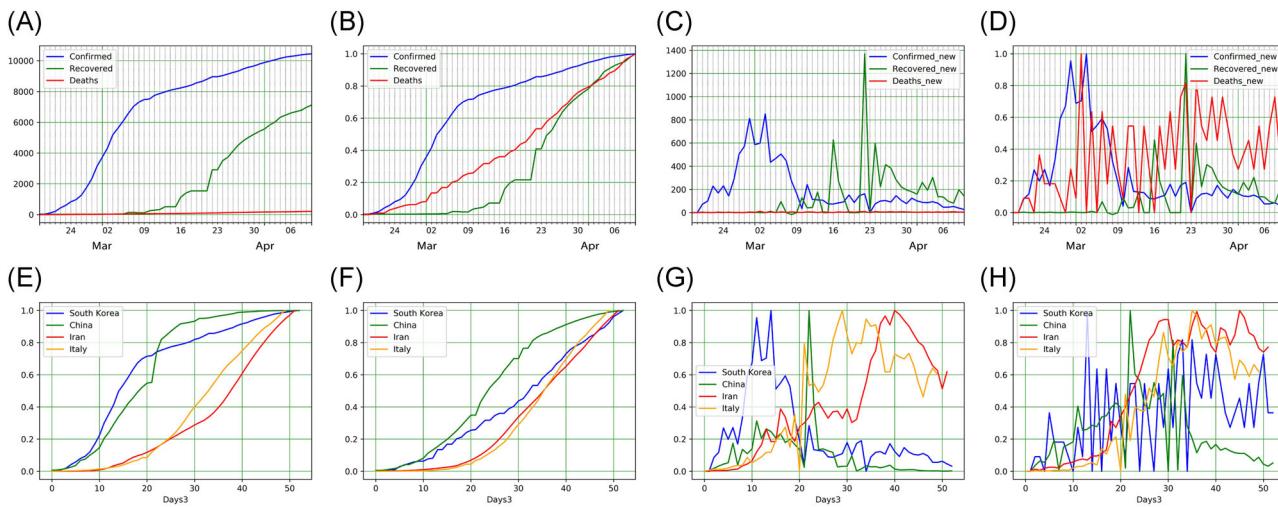


FIGURE 5 COVID-19 epidemic data and statistics of South Korea (A-D), and between counties (South Korea, Iran, Italy, and China). For South Korea: (A) confirmed, recovered, and death COVID-19 cases, (B) normalized data, (C) new data, and (D) normalized new data. To compare between countries: (E) confirmed cases, (F) death cases, (G) confirmed rate, and (H) new confirmed cases

than 20000 tests had been taken in 2 days). Until this date, South Korea reached to 4099 taken tests per million people. The magnitude of this number clearly shows the policy of South Korea and attempts to reduce the duration of detection, through a faster detection strategy.²⁴ Furthermore, the fact that the CFR in the old South Korean affected adults²⁵ (patients above 70 years) is still lower than that of Italian HCoV-affected patients (average 47 years old), which better signifies the role of early detection in controlling the epidemic.

3.2.1 | Deaths

To determine the CFR of confirmed cases, especially in countries in which the COVID-19 is spreading, the knowledge of confirmation to death duration is essential. To demonstrate a wrong approach, dividing the death cases by confirmed ones on a specific date yields the wrong answers for the death cases that might have been confirmed someday prior. By assuming that CFR has a linear (or near-linear) relation with the confirmed rate, the duration of confirmation to death could be evaluated by fitting multiple linear regression lines. The minimum value for MAE appears for 5 days shift (for South Korea), meaning death cases were confirmed 5 days before. To discern the CFR of South Korea, total death cases should be divided by the total confirmed date of 5 days prior (slope of the regression line also shows the rate).

As alluded to before, there are many types of COVID-19 concerning acuteness of the disease. The CFR varies depending on the level of infected people already confirmed. If a country manages to diagnose and confirm the infection of a patient in an earlier stage and begins curating the infected individuals or tallying those with mild COVID-19, the CFR would be comparatively lower. Such perception alludes to the

apparent lower CFRs in South Korea (1.8%) vs other countries with comparably large outbreaks.

3.3 | The United States: The great outbreak

The United States confirmed, the deceased, and recovered COVID-19 cases trend is depicted in Figure 6A, as well as its normal form in Figure 6B. On 10th March, both the confirmed and death cases of the United States increased (about 2.9 and 2.5 times, respectively). The huge increase in death and confirmed cases lead the regression line algorithm to define confirmed to death duration in 0 days. However, it would not be possible unless most dead cases are confirmed after death. Accepting the 3 days for the United States (see Figure 6G), the CFR should be about 4%. The high CFR could be alluded to the detection of the infected individuals during the last stages of the illness or reporting on serious cases with higher CFRs exclusively. Compared with China, Iran, Italy, and South Korea, both confirmed and death cases of the United States show higher orders of incrementation. Comparing the increment rate of countries, the United States seems to be in an earlier stage compared with China (outbreak is nearly ended), South Korea (at the ending stage), and Italy (passing the peak of the outbreak). Experimentally, for COVID-19, most of the countries in which the pandemic happened, (if the strategy does not change many times) two phases of the outbreak could be introduced. The first one is when the virus spreads increasingly (in each day new cases are more than previous), and the second one is the controlling outbreak in which new cases reduced to near zero. The United States is now in the spreading phase (new confirmed and death cases depicted in Figure 6E,F), the more reduction in new cases would cost more, and a trend with more slope costs more to be flattening. China and South Korea passed the first phase and

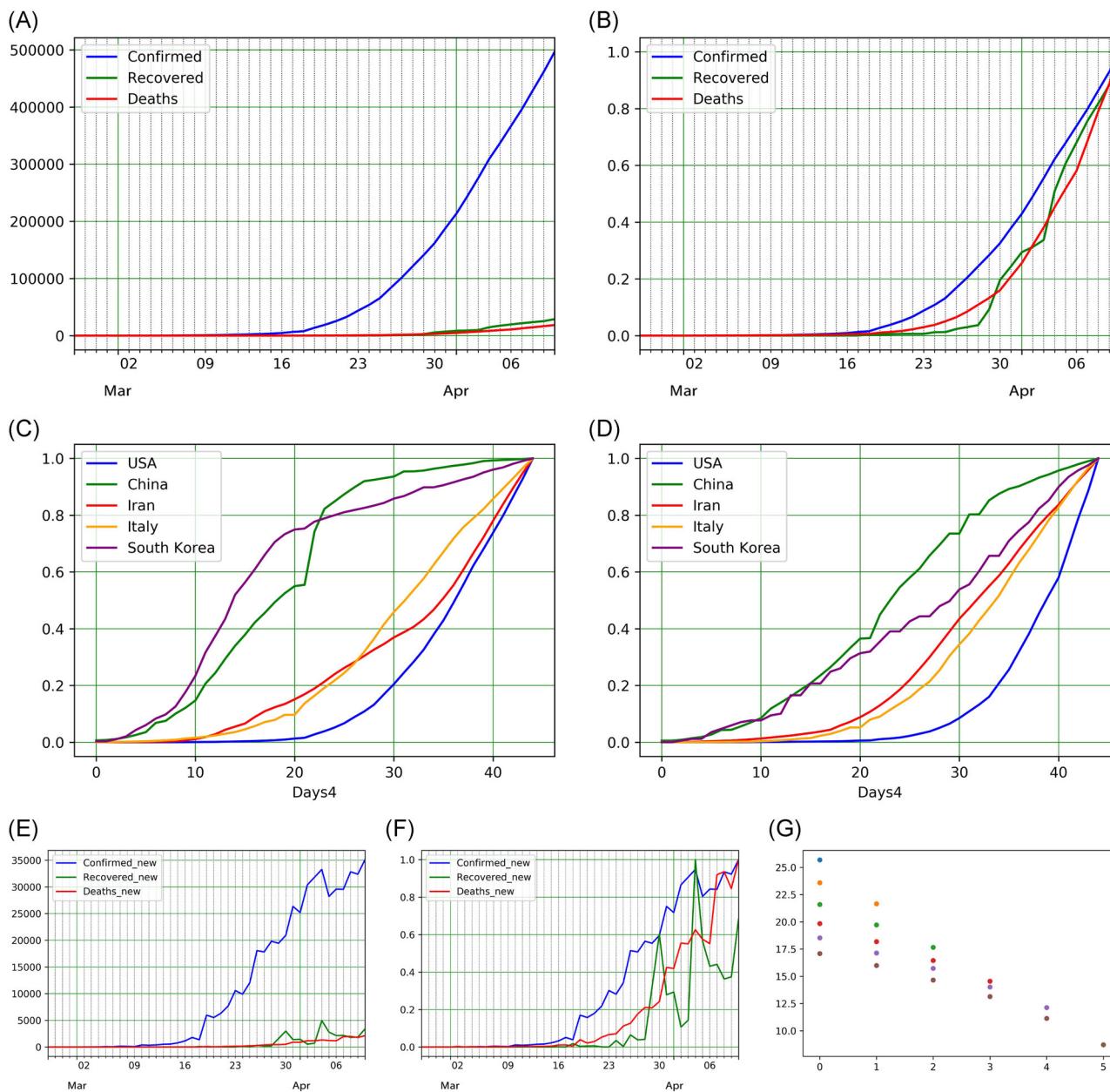


FIGURE 6 COVID-19 epidemic data and statistics of the United States: (A) Confirmed, recovered, and death COVID-19 cases, (B) normalized data for confirmed, recovered, and death COVID-19 cases, (C) confirmed cases of different countries (the United States, China, South Korea, Italy, and Iran), and (D) death cases of the same countries, (E) new confirmed, recovered, and death COVID-19 cases, (F) normalized data for new confirmed, recovered, and death COVID-19 cases, (G) confirmed death regression mean absolute error data for COVID-19 transmission through the United States

currently (5th April) it seems they are at the near end of the second stage. The experience of these two countries shows that the second phase will add about 150% of the first stage's cases but it also could vary depending on the strategies undertaken. Considering this calculation, each day in the first phase adds about 2.5 times the new cases on that date in total. Note that, these two phases could be repeated in a country more than once.

It is advisable to follow the United States in the coming days for more accurate information gathering, considering the latest increment could be caused by wrong data in the previous dates.

3.4 | Italy: High CFR confirmed cases

Italy reported its first confirmed cases of COVID-19 infection on 31st January, later announcing an outbreak around 21 February (21 days later). We have displayed (after normalizing) overall analysis in addition to death and recovered cases' trend in Figure 7B depicting COVID-19s trend in the country. By normalizing the plot, death and recovered cases' trend could be seen in Figure 7B.

The plot for new cases is also given in Figure 7C in real-valued and normal form. A comparison between the growth trend in

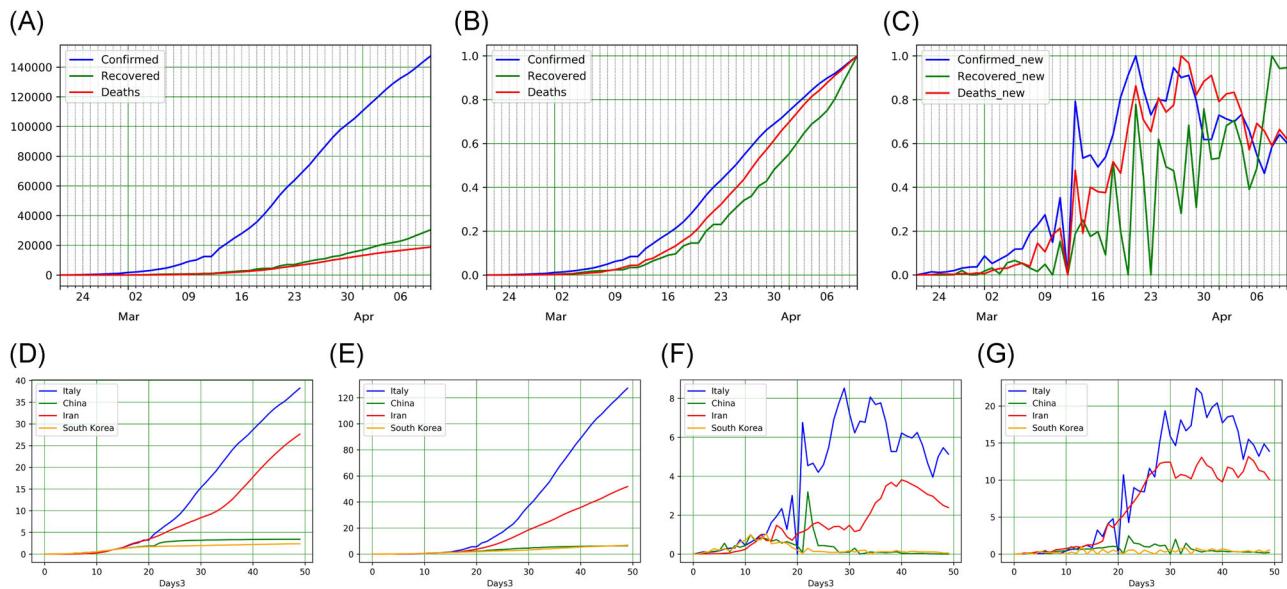


FIGURE 7 COVID-19 epidemic data and statistics of Italy: (A) confirmed, recovered, and death COVID-19 cases, (B) normalized data, (C) normalized new data. Comparison of COVID-19 statistics between China and Italy: (D) Confirmed cases, (E) death cases, (F) new confirmed cases, and (G) new death cases

numbers from Italy from this date with China's in the early stages is depicted in Figure 7D as shown, and the growth rate patterns are approximately the same between the two for the first 14 days, showing identical behavior in the increasing numbers of confirmed cases. After day 14, Italy exhibits an incremental increase in confirmed cases juxtaposed to China's. Lombardy, the center of the outbreak in Italy was locked down on 22nd February.

To analyze the stage of the disease in the Italian HCoV-infected patients (ie, to find out how rapidly the infected people have been screened and diagnosed), the results of the COVID-19 tests of Italy and South Korea can be compared. South Korea has been considered as a reference as its HCoV-related CFR is low enough, indicating that both countries are at the same level of HCoV infection. Till 10th March 2020, a total of 60761 HCoV tests of Italians had been undertaken through their screening program, whereas in South Korea, this number exceeds 210144 tests, more than threefolds. However, the number of the daily taken tests in Italy has been increasing, that is, it was 13000 on 11th March.²⁶

On 9th March, a national quarantine was imposed by the Italian government in which any movement of population has been restricted except necessary ones. Consequently, 11 days later, the increase in new cases in Italy stopped, and it seems Italy is in changing-phase days.

3.4.1 | Deaths

By comparing new cases of Italy and China which is manifested in Figure 7E (and the overall CFR comparison of the two in Figure 7G), it is clear that new death cases of Italy increased next to China's. Calculating the total CFR of Italy (confirmation to death duration

calculated as 4 days), about 14.5% of confirmed cases passed away. Compared with South Korea's rate of 1.8% (eight times than Italy) and China's 4% (72% lower than Italy), the CFR of Italy is too high. Ignoring race and climate as conditions (in which there is no clue of their impact), a strong rationale for this difference should exist. One hypothesis is that some infected individuals are not diagnosed until more serious stages of the disease. This could also explain the increase in confirmed cases, suggesting those infected, remain in contact with others. By comparing hospital beds per 10000 people, the indicator was 115.32 (at 2014), 42 (at 2012), and 34.22 (at 2012) (Iran 15 2014) for South Korea, China, and Italy, respectively.²⁷ Statistically, it is observed that the CFR of COVID-19 has had a direct relationship with age (with the average age of death for these three countries, respectively, 28, 38, and 47 years old²⁸) and it also potentially contributes to the increase in CFRs.

3.5 | Spain: Italy's follower

Massive levels of COVID-19 outbreak were reported around 25th February for Spain, as seen by its trend of confirmed, death, and recovered cases demonstrated in Figure 8A. By comparing the early stages of Spain with Iran, Italy, and South Korea in Figure 8C, the trend in 5th April shows Spain has a larger growth rate than other ones mentioned here. The mean age of Spain is 45 year old and hospital beds per person in the country is 29.65 for every 10000 (at 2013); compared with Italy, the mean age of Spain is slightly smaller and hospital beds per person rate is relatively higher, and consequently, it is expected that the CFR should, therefore, be slightly lower. After calculating the confirmed to death cases duration which is 5 days, the CFR of Spain results in 13.5% for Spain, as expected.

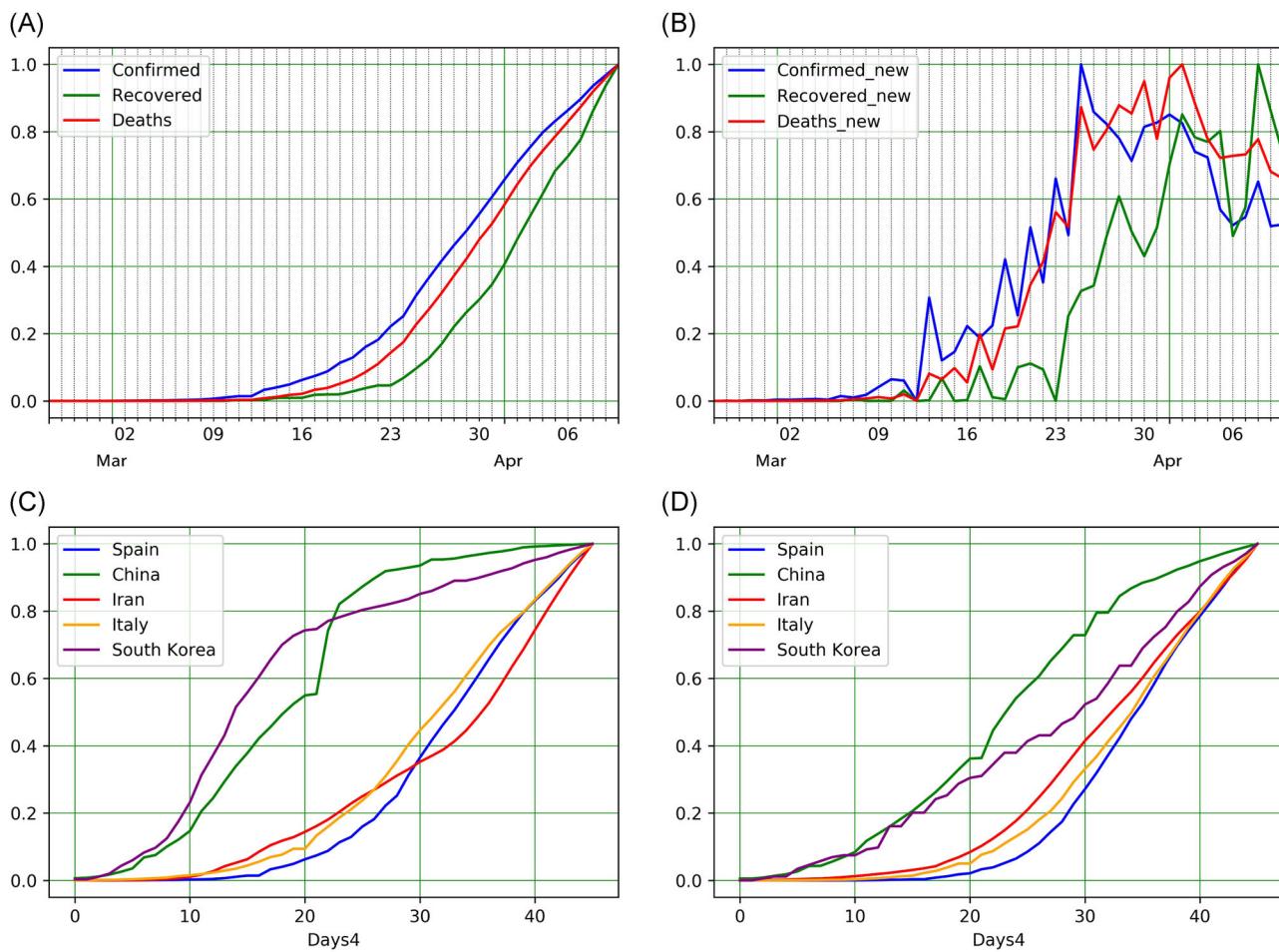


FIGURE 8 COVID-19 epidemic data and statistics of Spain: (A) normalized data for confirmed, recovered, and death COVID-19 cases, (B) new confirmed, recovered, and death COVID-19 cases, (C) confirmed cases of different countries (Spain, China, South Korea, Italy, and Iran), and (D) death cases of the same countries

Following by 9th March national quarantine in Italy, Spain proposed movement restriction on 16th March. As seen in Figure 8B, the increasing new cases of Spain stopped 10 days later. After Italy, Spain is also at the beginning of the phase-shifting period.

3.6 | Other countries

3.6.1 | Iran: Good recovery rate

COVID-19 was first reported in Iran starting with two dead cases on 19th February in Qom (a city near Tehran, Iran's capital), followed immediately by a huge outbreak, displayed in Figure 9A detailing its number of confirmed, deaths, and recovered cases, and normalized in Figure 9B. Officially, Iran has not mandated any city-wide lockdowns, but recently some provinces are refusing nonlocal travelers. Regarding the experiences in China, lockdown strategy will isolate cities to avoid transmitting the growth rate to other cities. However, such policies entail some negative consequences, and therefore, they were not implemented in the early stages of the outbreak in Iran in

contrast to Italy and South Korea, where there was a gap between outbreak and huge outbreak (21 and 27 days). On the 10th day of the outbreak in Iran, the total reported, confirmed, and infected individual cases reached 388.

To find confirmed to death cases mean duration, as before, we calculate the minimum MAE, assuming there is a linear relationship between these two. The linear regression algorithm shows the date shift should be zero, which means on average, death cases confirmed on the same day. By comparing the death trend with confirmed cases in Figure 9B, it is obvious that the trend of CFR was not linear. The reported death vs confirmed is too excessive in the early days. It is possible that the shortage of COVID-19 testing kits or/and the lack of clinical diagnosis approach caused such unreliable data. Countries at this stage mostly had about 7 days of confirmation to the dead period, so following this assumption the COVID-19s CFR in Iran is around 9.4%.

By observing the 24th February report, there were 49 active cases (the cases which are not recovered or dead but confirmed as a COVID-19 infected) reported (at least 14 cases are new) and the value for active cases in the next day was 79 (at least 30 new cases).

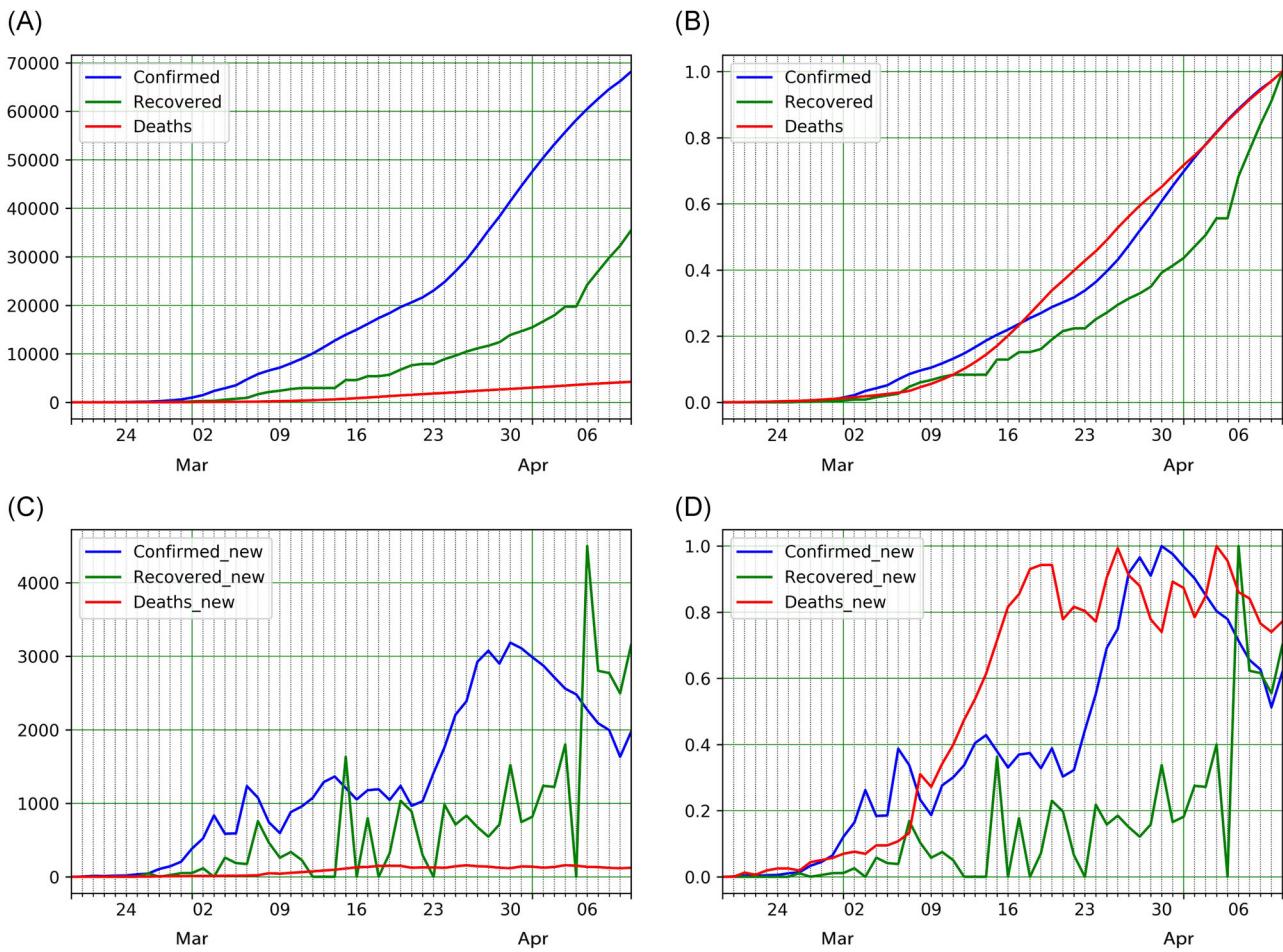


FIGURE 9 COVID-19 epidemic data and statistics of Iran: (A) confirmed, recovered, and death COVID-19 cases, (B) normalized data, (C) new data, and (D) normalized new data

Two days later, the reported recovered cases was 49. Suppose, no one was treated in less than 2 days, this results in a 100% recovery of 14 new cases in 2 days and also no active cases after 2 days, marking it a great experience which was reported by Iran's officials. Otherwise, some cases were treated in 1 day. This makes Iran's phenomena intriguing for investigating the details and contributing factors for such cases to find a relation between an individual's physical conditions and their treatment period. However, errors in the report could also explain this information. Using the regression line to find the mean recovered duration shows that the recovery duration for Iran should be around 1 day. Also seen in Figure 9C, the rate of recovery in Iran is quite significant at 37%.

On the 20th day of the outbreak, a temporal reduction in new cases could be observed starting on 6th March, persisting until 9th March. Evidently, it is the impact of a nationwide implemented policy reducing working hours 6 days prior (5.4 days is the mean incubation period¹⁸). After 4 days, with the arrival of weekends and the following two additional holidays, reports indicate, many people took trips during these dates which are widely assumed as a reason for confirmed cases' rise.

Although many companies implemented forms of remote work for their employees (which is estimated to have a positive impact on preventing the growth outbreak), yet many governmental offices working hours returned to their preoutbreak times (naturally assumed to negatively impact the prevention of disease's spread).

3.6.2 | Japan

The first confirmed cases of COVID-19 in Japan returned from Wuhan on 6th January. However, the outbreak in Japan sped up on 15th February. Figure 10A shows the trend of the outbreak after 15th February in Japan, whereas Figure 10B is normalized to compare confirmed, deceased, and recovered cases. The trend shows that the spread of COVID-19 in Japan behaves at a lower rate than China's first dates. CFR is estimated that it is about 4.1%. The CFR is good considering the mean age in Japan, which is 48. By finding out the hospital per bed value of Japan which is 134 (reported in 2012 by WHO), the relation of death cases compared with hospital per bed is more clear. Although

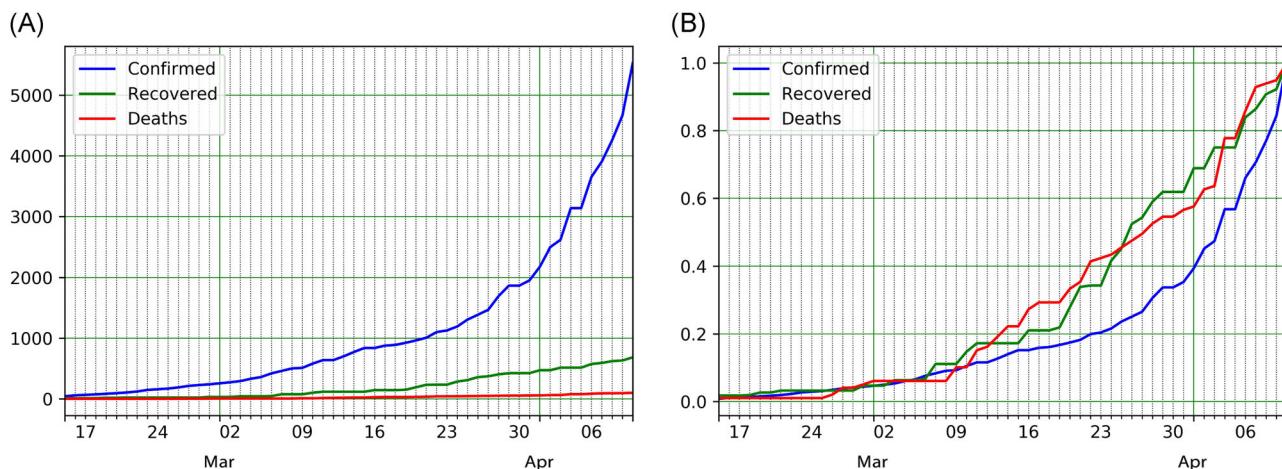


FIGURE 10 COVID-19 epidemic data and statistics of Japan: (A) confirmed, recovered, and death COVID-19 cases, (B) normalized data for confirmed, recovered, and death COVID-19 cases

Japan managed the early stage exponential rate of COVID-19, a reduction in new cases is not observed.

3.6.3 | France: Unexpected recent changes

France underwent two increments in its rate of confirmed COVID-19 cases. The first one was on 25th February and, therefore, chosen as the starting point of the outbreak. The second was from 4th April as shown in Figure 11A. By normalizing data displayed in Figure 11B and calculating the confirmed to death duration which is 7, CFR for France on 5th April results in 19.9%. Considering the mean age as a parameter, which in France is lower than Spain (42 and 45, respectively), in addition to beds per person which is 64.77 (at 2013)²⁷ (higher than Spain), CFRs are expected to be lower than Spain, but the unexpected growth in death cases from 2nd April increases the CFR of France even more than Spain's. The unexpected change also had an impact on the regression line. By comparing the growth of confirmed cases with China, Iran, Italy, and South Korea as depicted in Figure 11C, France has a very high slope in both confirmed and death. More than 1000 new death cases per day from 2nd April and more than 25000 new confirmed cases on 4th April contribute to such a faster increment.

3.6.4 | Germany: Handled CFR

By observing Figure 12A and comparing the trend with China, Italy, Iran, and South Korea, which is depicted in Figure 12B, Germany had the same COVID-19 trend as Italy. The total number of deaths in Germany in 17th March is 24. By dividing the deaths of 5th April to the confirmed on 30th March, 2.36% is the CFR, which is too low at considering the mean age of Germany which is 47. The possible reason for this low CFR is earlier confirmation, which leads to earlier hospitalization for the infected.

4 | DISCUSSION

The current outbreak of the novel-coronavirus (COVID-19), epicentered in Hubei province of the People's Republic of China, has spread to many other countries. The case detection rate is changing daily and can be tracked in almost real time on the mentioned website.

Reports on 5th April 2020 show that the United States, Italy, and China have the most confirmed, fatal, and recovered cases, respectively, and in terms of confirmed cases, Spain, Italy, Germany, and France are following the United States. Confirmed death cases lead in numbers by Italy and followed by Spain, the United States, and France. The daily statistics showed that lockdown is effective in the reduction of incidence of confirmed cases with COVID-19 after about 10 days in China, Italy, and Spain. South Korea is one of the first countries reporting the cases after China and the growth pattern of confirmed cases is the same as China's. However, they implement some policies such as in addition to isolation of people, social avoidance, and quarantine policies for infected, and faster detection of infected cases which were effective in a decrease in the new confirmed case and also case fatality. Italy and China have approximately the same growth rate patterns for the first 14 days. The lockdown strategy of Lombardy (the center of outbreak in Italy) seems to have had a positive effect on other municipalities. Unlike China's growth pattern, Iran's incremental trend continued to rise until the 20th day of the outbreak, even though a temporal reduction in new cases could be observed due to a nationwide implemented policy reducing working hours. The CFR in China was 4% (for Hubei was 4.7%). The highest and the lowest CFRs belonged to Italy (14.3%), and South Korea (1.8%), respectively, which could represent the effectiveness of their policies in control of the COVID-19. The United States has more than a quarter of confirmed cases. It seems the leading country in confirmed cases of COVID-19 should propose a new policy to reduce new cases and go to the next phase of the pandemic.

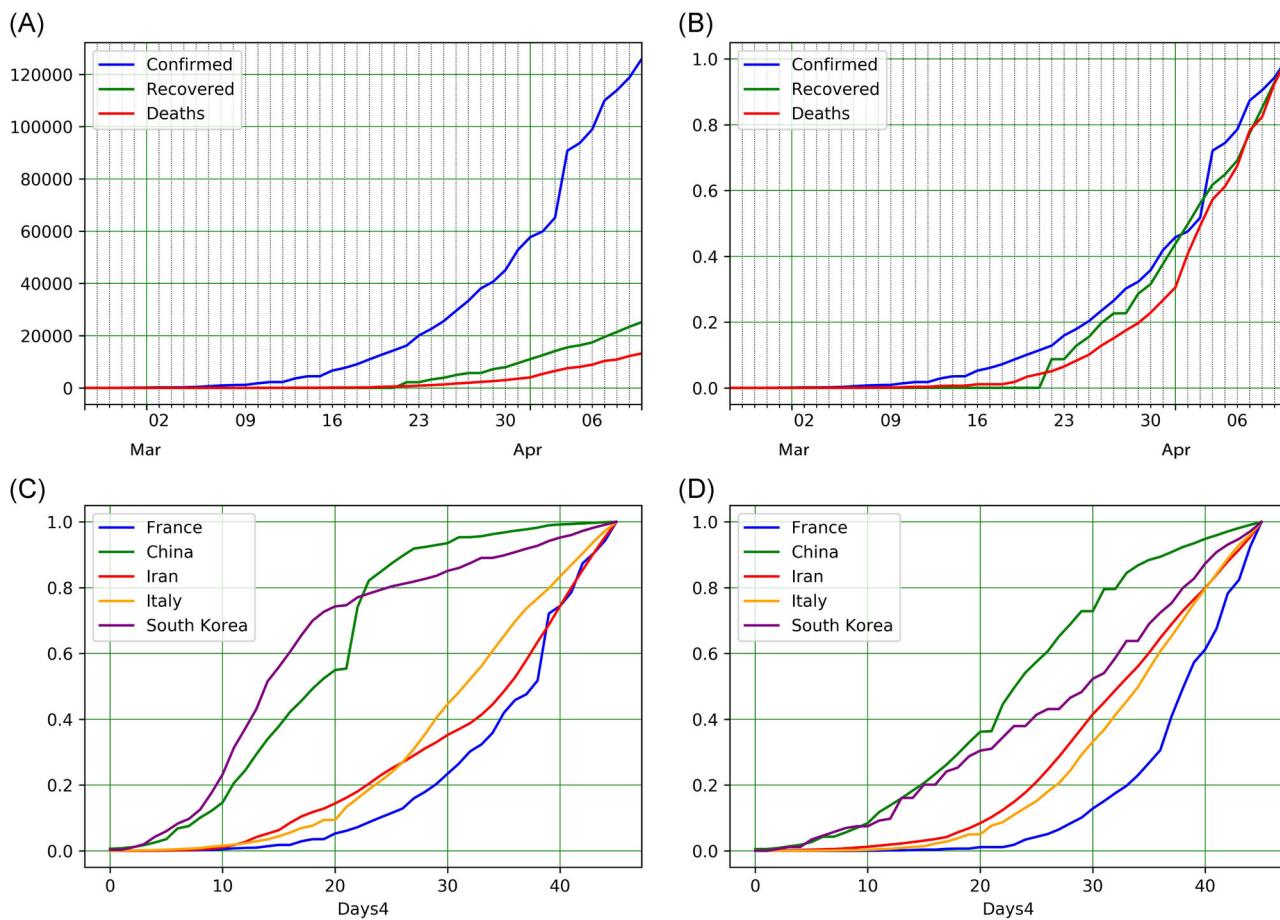


FIGURE 11 COVID-19 epidemic data and statistics of France: (A) confirmed, recovered, and death COVID-19 cases, (B) normalized data for confirmed, recovered, and death COVID-19 cases, (C) confirmed cases of different countries (France, China, South Korea, Italy, and Iran), and (D) death cases of the same countries

Social distancing is one of the most effective policies to control the past epidemic disease by limiting human to human transmission and reducing mortality and morbidity.^{29,30} However, studies suggest that a combination of multiple policies can boost effectiveness.

For instance, New York City Department of Health implemented different policies during the influenza pandemic in 1918-1919 at the same time and they have the lowest rate of mortality on the eastern seaboard of the United States.³¹

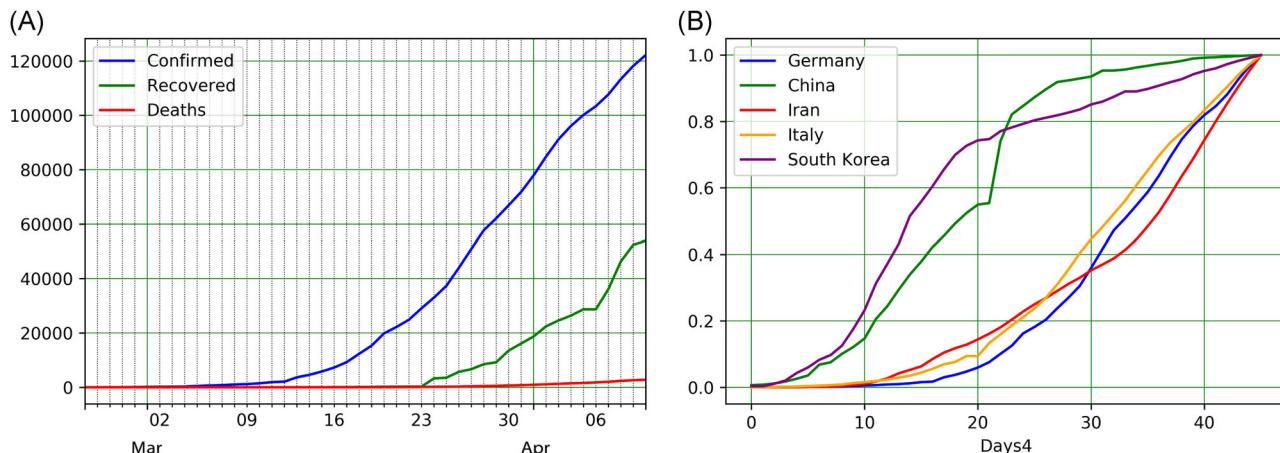


FIGURE 12 COVID-19 epidemic data and statistics of Germany: (A) confirmed, recovered, and death COVID-19 cases, and (B) confirmed cases of different countries (Germany, China, South Korea, Italy, and Iran)

During the COVID-19 outbreak, researchers predicted that the mass movement of people at the end of the Lunar New Year holiday would increase the spreading of disease. Facing this concern, government of China implemented policies which was helpful in controlling the disease such as, extending the holiday so that the holiday would be long enough to shelter the incubation period of COVID-19, isolation of confirmed cases in hospitals, quarantining mild or asymptomatic persons in different hospitals, home-based quarantine of people from Hubei province (epicenter of the epidemic), and the most important one was to prevent individuals with asymptomatic infections from spreading the virus.^{19,22} Iran is facing this concern as an important upcoming event in Iran is Nowruz which is the Iranian New Year, which recommended prompted policies from government.

ORCID

Amirhoshang Hoseinpour Dehkordi  <http://orcid.org/0000-0002-6386-7037>

Majid Alizadeh  <http://orcid.org/0000-0003-2644-5959>

REFERENCES

- Myint SH. Human coronaviruses: a brief review. *Rev Med Virol*. 1994; 4(1):35-46.
- Tyrrell DAJ, Bynoe ML. Cultivation of viruses from a high proportion of patients with colds. *Lancet*. 1966;287:76-77.
- Lim YX, Ling Ng Y, Tam JP, Liu DX. Human coronaviruses: a review of virus-host interactions. *Diseases*. 2016;4(3):26.
- Ashikujaman S. Coronavirus: a mini-review. *Int J Curr Res Med Sci*. 2020;6(1):8-10.
- Goldsmith CS, Tatti KM, Ksiazek TG, et al. Ultrastructural characterization of SARS coronavirus. *Emerg Infect Dis*. 2004;10(2):320-326.
- Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. In: Maier H, Bickerton E, Britton P, eds. *Coronaviruses*. New York, NY: Humana Press; 2015:1-23.
- Forgie S, Marrie TJ. Healthcare-associated atypical pneumonia. *Semin Respir Crit Care Med*. 2009;30:67-85.
- Liu P, Shi L, Zhang W, et al. Prevalence and genetic diversity analysis of human coronaviruses among cross-border children. *Virol J*. 2017; 14(1):230.
- Smith RD. Responding to global infectious disease outbreaks: lessons from SARS on the role of risk perception, communication and management. *Soc Sci Med*. 2006;63(12):3113-3123.
- World Health Organization. Case-control study to assess potential risk factors related to human illness caused by Middle East Respiratory Syndrome Coronavirus (MERS-CoV). 2014.
- World Health Organization. Middle East Respiratory Syndrome Coronavirus (MERS-CoV): summary of current situation, literature update and risk assessment. Technical report, World Health Organization. 2015.
- Green MS, Swartz T, Mayshar E, et al. When is an epidemic an epidemic? *Isr Med Assoc J*. 2002;4(1):3-6.
- Johns Hopkins CSSE. 2019 Novel coronavirus covid-19 (2019-ncov) data repository. 2020.
- World Health Organization. Coronavirus disease 2019 (COVID-19) situation report. 2020.
- Gong J, Dong H, Xia SQ, et al. Correlation analysis between disease severity and inflammation-related parameters in patients with covid-19 pneumonia. *medRxiv*. 2020. <https://doi.org/10.1101/2020.02.25.20025643>
- Liu W, Wang F, Li G, et al. Analysis of 2019 Novel Coronavirus Infection and Clinical Characteristics of Outpatients: an Epidemiological Study from the Fever Clinic in Wuhan, China. *SSRN Electron J*. <https://doi.org/10.2139/ssrn.3539646>
- World Health Organization. Coronavirus disease 2019 (COVID-19) situation report. 2020.
- Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020. *Euro Surveill*. 2020;25(5): 10-15.
- Li Q, Guan X, Wu P, et al. Early transmission dynamics in wuhan, china, of novel coronavirus-infected pneumonia. *N Engl J Med*. 2020; 382:1199-1207.
- Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA*. 2020;323:1061.
- Wang W, Tang J, Wei F. Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China. *J Med Virol*. 2020;92:441-447.
- Chen S, Yang J, Yang W, Wang C, Bärnighausen T. COVID-19 control in China during mass population movements at New Year. *Lancet*. 2020;395:764-766.
- Korean Centers for Disease Control and Prevention. The Update of COVID-19 in Republic of Korea. 2020.
- worldometers. Coronavirus testing: criteria and numbers by country. 2020.
- Korean Centers for Disease Control and Prevention. 114 additional cases are confirmed. 2020.
- Umberto Rosini Bruce Becker. Dati COVID-19 italia. 2020.
- World Health Organization. Hospital beds (per 10 000 population). 2020.
- worldometers. Countries in the world by population (2020). 2020.
- Ahmed F, Zviedrite N, Uzicanin A. Effectiveness of workplace social distancing measures in reducing influenza transmission: a systematic review. *BMC Public Health*. 2018;18(1):518.
- Caley P, Philp DJ, McCracken K. Quantifying social distancing arising from pandemic influenza. *J R Soc Interface*. 2008;5(23):631-639.
- Markel H, Lipman HB, Navarro JA, et al. Nonpharmaceutical interventions implemented by us cities during the 1918-1919 influenza pandemic. *JAMA*. 2007;298(6):644-654.

How to cite this article: Hoseinpour Dehkordi A, Alizadeh M, Derakhshan P, Babazadeh P, Jahandideh A. Understanding epidemic data and statistics: A case study of COVID-19. *J Med Virol*. 2020;92:868-882. <https://doi.org/10.1002/jmv.25885>