Global Comparison of Changes in the Number of Test-Positive Cases and Deaths by Coronavirus Infection (COVID-19) in the World

Akihiro Hisaka [1], *, Hideki Yoshioka [1], Hiroto Hatakeyama [1], Hiromi Sato [1], Yoshihiro Onouchi [2] and Naohiko Anzai [3]

[1]   Clinical Pharmacology and Pharmacometrics, Graduate School of Pharmaceutical Sciences, Chiba University, 1-8-1, Inohana, Chuo-ku, Chiba, 260-8675, Japan; hisaka@chiba-u.jp (A.H.); ysokhdk@gmail.com (H.Y.); h-hatakeyama@chiba-u.jp (H.H,); hiromi-s@chiba-u.jp (H.S.)

[2]   Department of Public Health, Graduate School of Medicine, Chiba University, Chiba, 1-8-1, Inohana, Chuo-ku, Chiba, 260-8670, Japan.; onouchy@chiba-u.jp

[3]   Department of Pharmacology, Graduate School of Medicine, Chiba University, Chiba, 1-8-1, Inohana, Chuo-ku, Chiba, 260-8670, Japan.; anzai@chiba-u.jp

*   Correspondence: hisaka@chiba-u.jp; Tel.: +81-43-226-2880

Abstract: Global differences in changes in the numbers of population-adjusted daily test-positive cases (NPDP) and deaths (NPDD) by COVID-19 were analyzed for 49 countries. The changes per population of a hundred million were compared, adjusting by the beginning of test-positive cases increase (BPI) or deaths increase (BDI). Notable regional differences of more than 100 times in NPDP and NPDD were observed. The trajectories of NPDD after BDI increased exponentially within 20 days in most countries. A machine learning analysis suggested that NPDD on 30 days after BDI was the highest in Western countries (1180), followed by the Middle East (128), Latin America (97), and then Asia (7), and furthermore that, NPDD in Western countries with a positive rate of the PCR test of less than 7.0% attenuated to only 15%. The cause behind differences between regions might be complex, however, investigation of the host genetic factors would be warranted. The lower positive rate would be caused by aggressive testing policy and associated with longer lag times between BPI and BDI. Our analysis suggested that the positive rate need to be 7% or less by extensive tests to reduce deaths effectively. As the number of infected people is growing rapidly, earlier expansion of the test capacity is indispensable.

Keywords: COVID-19; coronavirus; infectious disease; infection management; PCR test; mortality; kinetic analysis

1. Introduction

Coronavirus diseases 2019 (COVID-19) has been spreading globally and declared as a pandemic by WHO on March 11, 2020 [1]. Following the outbreak in Wuhan, China, in December 2019 [2], the number of test-positive cases in Italy, the United States, Spain, France, and Germany have surpassed the number of Chinese cases and more countries will soon follow. Nations and cities in which outbreaks have occurred have adopted strong measures such as overall shutdowns. However, the number of victims is growing very rapidly. The cumulative number of test-positive cases worldwide reached 1,619,964 on April 10, 2020 [3] and is currently increasing

at a daily rate of about 80,000 new cases. The cumulative number of cases in Asia is currently around 270,000, which accounts for 16.5% of the world's cases. The number of cases in Europe and North America is higher than in Asia, accounting for 48.8% and 30.9% of the world's cases, respectively. Currently, the densities of cumulative test-positive cases are 18.4 and 14.5 times higher in Europe and North America than in Asia, respectively, and the gaps are constantly widening. In contrast, the density of test-positive cases in Africa is very low, at only 0.17 times that of Asia now. The number of test-positive cases is ever-increasing and depends on the timing of the outbreak, and differences in medical environments and policies between countries. Thus, it is difficult to interpret simple differences in the number of cases between continents. However, large biases between different continents are a fact and serious considerations must be given to how to reasonably terminate the infections of different degrees depending on the continental, and at the same time, the possibility of spreading to areas where the infection is not yet widespread. [4]

In contrast, the number of test-positive cases may be affected by the number of conducted PCR tests, which is now widely accepted as the golden standard method for confirming COVID-19 [5]. Countries that actively conduct PCR tests believe that testing is indispensable to identify infected individuals and quarantine them. However, even in countries where the infection has widely spread, it is difficult to test everyone because the PCR test is rather laborious and the infection rate in such countries is only approximately 0.1%. In addition, even if a virus is present in the body, false negatives may occur [6], and so it is hard to ensure that every infected individual is quarantined. Conversely, some countries seem reluctant to the PCR test because patients without positive symptoms have a low risk of spreading the infection, and aggressive test to those people requires waste resources and may cause disruption of the medical system. However, in such countries, not quarantining infected people may increase the risk of an explosive outbreak. Since different countries have implemented different testing policies, aiming to investigate the state of the current COVID-19 by analyzing the number of test-positive cases alone may lead to bias.

This study aimed to investigate regional differences in the global COVID-19 outbreak by analyzing the number of deaths, which is a more direct consequence of the infection, in addition to the number of test-positive cases in various countries. Generally, the basic reproduction number (R_0) is preferred to assess outbreak risks of infectious diseases as well as COVID-19 [7]. However, estimating R_0 requires an estimation of the number of infected individuals, infection rate, quarantine rate, and recovery rate, and these parameters would be potentially time-dependent, non-linear, and diverse among populations. Moreover, the reliability of estimated R_0 is difficult to verify due to the uncertainty of the information used for the calculation. Thus, it is not easy to compare R_0 among different countries in a timely manner. For this reason, we have focused our analysis on changes in the number of test-positive cases and deaths adjusted by the population of each country and considering the timing of the outbreak.

2. Materials and Methods

According to the Kermac-McKendrick model [8], the number of infected subjects at time t, normalized to the total susceptibility population, $I(t)$, obeys the following Malthus' law:

$$I(t) = I(0)\, e^{(\beta-\gamma)t}, \tag{1}$$

where β is the infection rate and γ is the recovery and isolation rate. Here, infected subjects include those who are asymptomatic. Immediately after the initiation of the outbreak, which is

typically caused by the infection being brought in from abroad, $\gamma = 0$ because neither recovery nor isolation is possible. Currently, in most countries, the PCR testing starts when symptoms of infection appear, so the period during which γ is 0 continues until the symptoms become apparent after infection occurs. We call this "the hidden phase." After detection, the isolation rate increases with intensive testing, and the number of test-positive cases explosively increases. We call this "the explosive increase phase." Thereafter, the quarantine rate increases and is tentatively kept at a constant value depending on the testing system of each country. Some asymptomatic people may recover in this period and it also raises γ. The ratio of the number of test-positive cases to infected cases becomes approximately constant due to equilibrium of these rates. Thus, an increase due to Malthus' law is observed, and we call it "the exponential increase phase." From several days to about two weeks later, due to social responses to the infection [9], β gradually decreases. Simultaneously, deaths and recovered individuals appear mostly from quarantine or hospitalized people. We call this "the deceleration phase." As the deceleration phase continues, β becomes smaller than γ, and the number of infected people begins to decrease, and later the infection converges.

It should be noted that the number of test-positive cases analyzed in this study is smaller than the number of infected people and is observed only after the hidden phase. The ratio of test-positive cases and infected cases should be time-dependent especially in the initial period. The number of deaths is only observed after the deceleration phase. The basic reproduction number is represented by $R_0 = \beta/\gamma$, but its meaning is completely different depending on the phase. To estimate R_0, it is necessary to select an appropriate timing and to estimate the detection ratio and recovery rate in each country, and thus this was not analyzed in this study.

Daily numbers of test-positive cases and deaths by COVID-19 reported from all countries worldwide were downloaded from the webpage of the European Center for Disease Prevention and Control (ECDC) [10] as of April 8, 2020. The analysis was narrowed down to the top 50 infected countries as of March 29. Because the Princess Diamond cruise ship was excluded, the actual analysis was performed for 49 countries. The number of population-adjusted daily test-positive cases (NPDP) and the number of population-adjusted daily deaths (NPDD) were calculated as 5-day moving average for a population of a hundred million. The first and last points were calculated as a three-day moving average. The beginning of test-positive cases increase (BPI) was defined as a point at which the average NPDP for two continuous days exceeded 10. The beginning of deaths increase (BDI) was defined as a point at which the average NPDD for two continuous days exceeded 1 for Asian countries and 5 for non-Asian countries, respectively. Vietnam, Russia, and South America did not meet the criteria for BDI. Due to a small population, the BDI of Singapore, Luxembourg, and Iceland was not reliable (too long) and thus these countries were excluded from BDI related analysis. The virus testing statistics by country was obtained from English version of Wikipedia [11].

Regression analysis was performed on the trajectories of NPDD after BDI using the Gradient Boosting Decision Tree algorithm [12]. A dummy covariate representing a global region or a range of the positive rate of the PCR test was used. The analysis was performed on Python (version 3.7.3) [13, 14] using GradientBoostingRegressor function of the scikit-learn library (version 0.22.2) [15]. The 95% confidence interval was calculated by repeating the country-level bootstrap analysis of 1,000 times.

3. Results

Figure 1 shows changes in NPDP and NPDD per 100 million in each country from the BPI. Most countries showed a rapid increase in NPDP for about 5 days after the BPI, which was probably the explosive increase phase. From 5 to 20 days from BPI, an exponential increase was

observed in many countries. Asian (excluding the Middle East, the same applies hereinafter) and non-Asian countries (Europe, Middle East, Americas, Africa, and Oceania, the same applies hereinafter) showed very different trajectories later than 20 days after BPI, i.e. the deceleration phase. In Asian countries, NPDP fluctuated mostly in the range from 10 to 1,000. In some countries, the fluctuation has continued for more than 70 days. In most non-Asian countries, however, NPDP increased steeply and then reached the plateau of 1000 or more by 30 days after BPI. In some European countries, NPDP reached almost 10,000. A plateau of NPDP does not indicate a plateau in the number of cumulative test-positive cases, rather its linear increase. The NPDP analysis found no noticeable regional differences between Europe, Africa, the Americas, and the Middle East. The outbreak is yet in the very early phase in the South Asian countries, i.e. Pakistan and India, but it seems that their NPDP time courses are more like the East Asian countries than the European countries currently. Fortunately, the NPDPs in Africa and in Oceania (Australia and South Africa, respectively) entered a decreasing phase recently although only one country in these areas was analyzed.
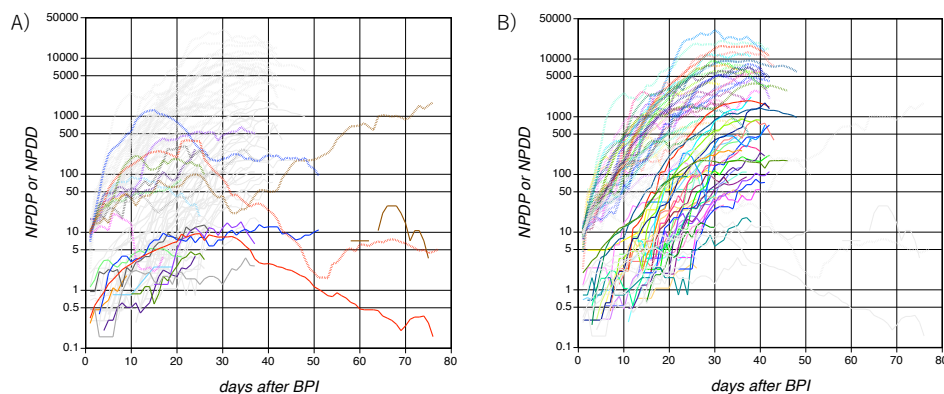
Figure 1. Time courses of NPDP and NPDD after BPI by COVID-19 in countries in Asia (except for the Middle East. Panel A) and in non-Asian countries (Panel B). The upper dotted lines and lower solid lines represent NPDP and NPDD, respectively. The colors of NPDP and NPDD are corresponding to each country. Lines in very light gray indicate lines in non-Asian and Asian countries in Panel A and B, respectively, for comparison. Refer to Figures 2 and 3 to identify each country.

In the change in NPDD, unlike NPDP, almost no initial explosive increase was observed. The increase in NPDD in Asian countries was modest, with values of about 10 or less on the 20th day after BPI. In some countries, NPDD was less than 1. In contrast, in most non-Asian countries, NPDDs rose significantly, as high as 10-2000 at 30 days after BPI. However, there are still many countries where the period after BPI is still shallow, and the value is not yet definitive. In many countries, NPDD also reached a plateau on 30 days after BPI, and the cumulative number of deaths had shifted from an exponential increase to a linear increase.

The beginning of NPDD increase was somewhat later than the beginning of NPDP increase, and the lag time from BPI to BDI varied from 1 to 25 days among the countries. As will be shown later, this lag time was related to the positive rate of the PCR test in each country. Figures 2 and 3 show individual NPDD and NPDP profiles for countries ordered by the positive rate of the PCR test, with information of the lag time. In Asian countries, the relationships between the seriousness of the infection and the positive rate were not clear. Nevertheless, current NPDDs were very low in countries with the lowest positive rate such as Taiwan, China, and Vietnam. A similar tendency was also observed in the countries in the Middle East and Latin America. In

Western countries, it was apparent that higher NPDD occurred in the countries with a higher positive rate and shorter lag time.
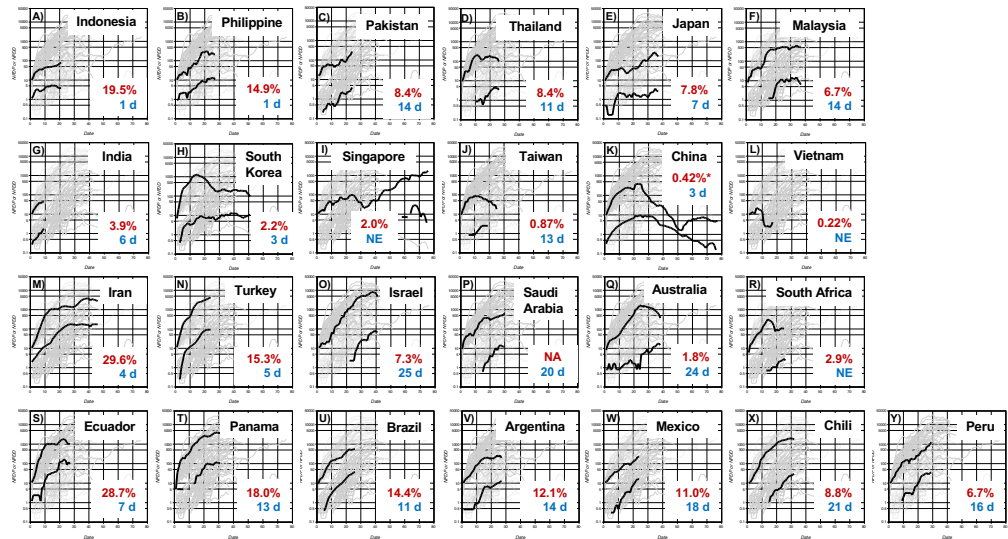
Figure 2. Time courses of NPDP and NPDD by COVID19 after BPI in countries in Asia (Panels A to L), the Middle East (Panels M to P), Oceania (Panel Q), South Africa (Panel R) and Latin America (Panels S to Y). Numbers in red and blue represent the positive rate of the PCR test and the lag-time between BPI and BDI, respectively. The countries are placed in decreasing order of the positive rate in each region. The upper and lower bold lines in each panel are NPDP and NPDD, respectively, except for Vietnam where no death has been recorded. The scales in all panels are the same as those in Figure 1. *: The positive rate in Guangdong. NE: Not evaluable. NA: Not available.
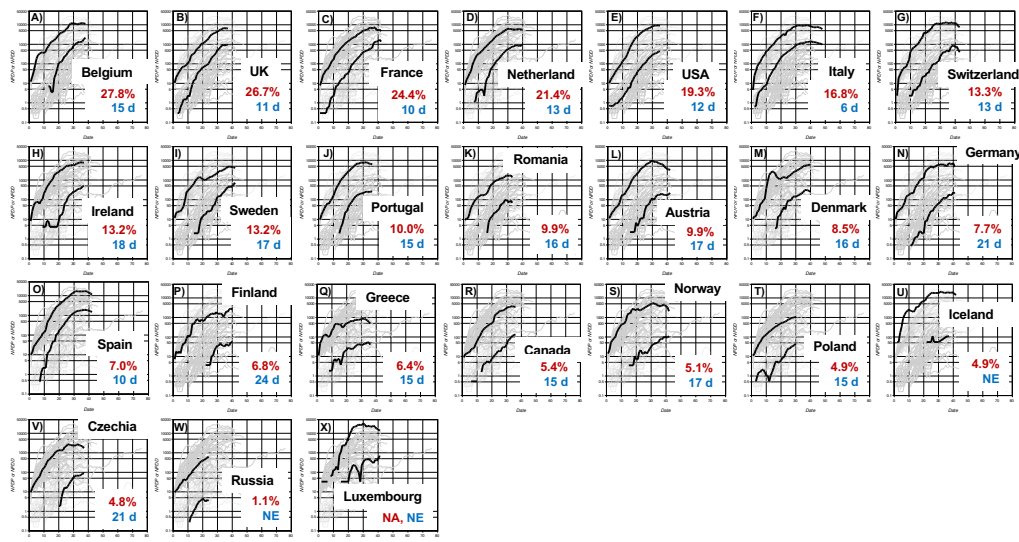
Figure 3. Time courses of NPDP and NPDD after BPI by COVID19 in countries in Europe and North America. Numbers in red and blue represent the positive rate of the PCR test and the lag-time between BPI and BDI, respectively. The countries are placed in decreasing order of the

positive rate. The upper and lower bold lines in each panel are NPDP and NPDD, respectively. The scales in all panels are the same as those in Figure 1. NE: Not evaluable. NA: Not available.

There would be no extreme difference in the period between timings of infection and death in countries where medical care has reached a certain level. Therefore, compared with the countries with a long lag time, in countries where the lag time between BPI and BDI was short, the true infection may have occurred somewhat earlier than was detected by the increase in NPDP [16]; i.e. "the hidden phase" should have become longer. For this reason, we also analyzed changes in NPDD after BDI (Figure 4). The trajectories of NPDD after BDI were consistent for all the countries in the world. NPDD showed an exponential increase until approximately 20 days after BDI and then reached a plateau by 25 days. It is encouraging that the peaks of NPDD for Italy and Spain seem to have passed. At present, only China has shown a clear decline in NPDD. By this analysis, it became somewhat apparent that NPDD trajectories of countries in the Middle East and Latin America follow an intermediate transition between Asian and Western countries.
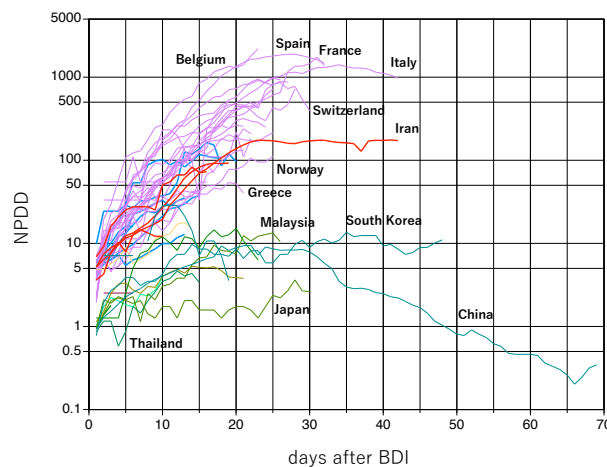
Figure 4. Time courses of NPDD after BDI by COVID-19. Green, red, purple and blue lines represent countries in Asia, the Middle East, Western and Latin America, respectively. Some noticeable countries are labeled but refer to Figures 2 and 3 to identify each country.

Figure 5 shows the results of a machine learning analysis on the transition of NPDD after BDI. Due to differences in BDI criteria between Asian countries and non-Asian countries, BDI in non-Asian countries is approximated to be a few days earlier than the currently observed value. Even though such a difference was taken into the consideration, distinct global regional differences were demonstrated by the analysis because the 95% confidence intervals (CI) did not overlap at all between Asian and non-Asian countries (Figure 5A). The analysis estimated that the medians of NPDD on 30 days after BDI were 1180, 128, 97 and 7 in Western countries, the Middle East, Latin America, and Asia, respectively. Africa was not analyzed because only one country was included. The effect of a positive rate of the PCR test on NPDD in Western countries was also analyzed by machine learning (Figure 5B). In countries with a positive rate of less than 7%, a number of deaths after 30 days after BDI was expected to be 15% of countries with a positive rate of equal or more than 17.0%. No difference in NPDD was detected between countries with a positive rate of 7.0 to 16.9% and 17.0 to 28.0%. The positive rate was the lowest, at 1.8%, in

Australia which has successfully suppressed deaths to the smallest in Western countries at present in this analysis.
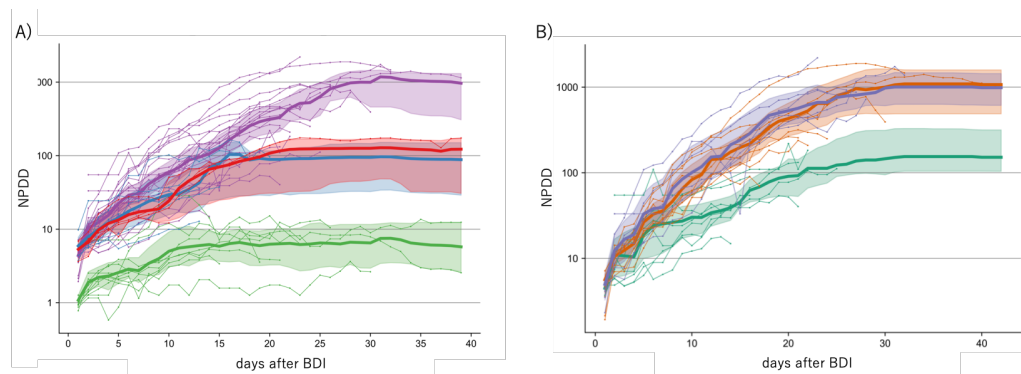
Figure 5. Estimation of the time course of NPDD after BDI by machine learning analysis classified by the global region (Panel A) and by the positive rate of the PCR test (Panel B). In Panel A, yellow-green, red, blue, and purple lines indicate countries in Asia (excluding the Middle-East), the Middle East, Latin America, and Western (Europe, Oceania, and North America), respectively. In Panel B, only Western countries were analyzed. Green, orange and blue lines indicate the positive rate of 0.0 to 6.9%, 7.0 to 16.9% and 17.0 to 28.0%, respectively. Each colored area represents 5 to 95% confidence interval of the median estimated by bootstrap analysis.

To examine whether the lag time from BPI to BDI is related to the aggressiveness of the PCR test in each country, we compared the lag time and the positive rate of the PCR test in Western countries. As shown in Figure 6, a statistically significant negative correlation was observed between the lag time and the positive rate of the test (p < 0.01 by t-test).
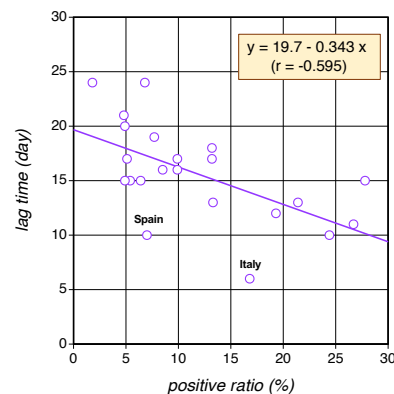
Figure 6. Correlation between the positive rate of the PCR test and the lag time between BPI and BDI in Western countries. The correlation was statistically significant by t-test (p < 0.01).

4. Discussion

In this study, remarkable regional differences in changes in NPDP and NPDD became apparent between Asian countries (excluding the Middle East) and non-Asian countries (Figure 1). In NPDDs of countries in the Middle East and Latin America, small differences were also

found from the countries in the Asian or Western countries (Figure 4 and 5). COVID-19 outbreaks occurred everywhere on the Earth which kinetically means that R_0 is considerably higher than 1.0 in any country without precautionary measures and efficient quarantine. However, once appropriate social measures were undertaken, considerable regional differences appeared from 20 days after the outbreak. Between-country differences and fluctuations in NPDP in Asia may be due to accidental outbreaks or intermittent inputs of infected people from overseas. Attention should be paid to recent rises in NPDPs in several Asian countries, such as China and Singapore, where NPDP had been well controlled for a while. Infection with different types of SARS-Cov-2 [17] would need to be considered because some small outbreaks in Asian countries synchronized with outbreaks in Europe. Nevertheless, it seems that the later outbreaks in Asian countries were smaller than the first one, fortunately.

In contrast, non-Asian countries showed steep increasing trajectories for both NPDP and NPDD. Especially in some European countries where the serious outbreaks occurred, NPDP and NPDD have reached very high numbers, approximately 10,000 and 1,000, respectively, by 30 days after BPI. However, at the same time, it should be noted that NPDP has recently gone into the decline phase in some European countries such as France, Italy, Switzerland, Portugal, Spain, Greece, Norway, and Austria (Figure 3). In China, the peak of NPDP occurred on February 6 and, given the recent partial deblocking of Wuhan, the containment in June or July would not be impossible in Europe.

The reason behind the smaller spread of the virus in Japan, compared to Europe and other Western countries, has been discussed. It was suspected that the smaller number of PCR tests carried out in Japan is responsible for this [18]. The results of this analysis showed that Japan, as well as all Asian countries, were considerably slower in the spread of the infection compared with the European countries when test-positive cases were adjusted by the population. In addition, similar regional differences were observed not only in NPDP but also in NPDD, indicating that the difference in the number of tests did not significantly affect the evaluation of the differences in the rate of infection between Asian and non-Asian countries. In addition, it should be noted that the positive rate of the PCR test in Asian countries is generally not higher than Western countries for the sake of a smaller number of test-positive cases (Figures 2 and 3).

The lag time between BPI and BDI differed significantly, even within non-Asian countries, where the trajectories of NPDP were similar among countries. Such differences might be caused by differences in the medical environments and their capacity, as well as differences in the policy of PCR testing because a long lag time would be caused by a delay in the detection of infected citizens. On the other hand, the country's testing policy may affect the positive rate of the PCR test. A country that only tests people admitted to hospitals will have a higher positive rate than a country that tests all citizens, whether or not they are showing symptoms, other things being equal. A clear reverse correlation was observed between the lag time and the positive rates of the PCR test in this study (Figure 6), suggesting that both indexes are appropriate as a measure of the testing policy of each country. In the correlation analysis, deviations were somewhat larger in Italy and Spain where the serious spread of the infection occurred. It may be interpreted as a result of the rapid expansion of the testing capacity after the infection have spread to a significant degree. It is reasonable that an earlier expansion of the testing capacity would be better to attenuate consequences of the infection effectively. The relationship between test aggressiveness and consequences of an infection is a very important perspective, perhaps not previously considered in other infectious disease outbreaks.

The machine-learning analysis estimated that NPDD in Western countries with a positive rate of the PCR test of less than 7.0% attenuated to only 15% compared to countries with a higher positive rate (Figure 5B). The analysis suggested that the positive rate need to be 7% or less at

least to reduce deaths effectively. In this study, it is difficult to define the target positive rate for effective attenuation of deaths, however, 7% is probably not enough because some countries with a positive rate of less than 2% such as Australia and Russia, NPDDs were effectively suppressed further (Figure 3). Moreover, in Asia, deaths by COVID-19 has been almost successfully contained currently in China and Thailand where the positive rate was less than 1% (Figure 2). Thus, our analysis suggested that extensive tests could have a significant effect on reducing the number of victims. Nevertheless, the relationships between the testing policy and the positive rate, and between the positive rate and spread of infection need to be verified thoroughly in the future.

It is unclear whether the large interregional differences in responses to the COVID-19 observed in this study were due to genetic differences between the inhabitants. Ethnic differences in relation to responses to infectious diseases have been rarely reported. However, this possibility cannot be denied because distributions of the race and of observed infection increasing rates were in good agreement globally. Habitants in East Asia and Southeast Asia are mostly Mongoloids, whereas Western countries are more heterogenic but substantially dominated by Caucasians. The Middle East, Pakistan, and India are classified generally as Caucasians, but they differ considerably from Europeans. In this analysis, Pakistan and India were closer to the results in Asian countries, while Iran showed an intermediate increasing infection rate between Asian and non-Asian countries. Although machine learning analysis classified the Middle East and Latin America as an independent group from Europe and other parts of Asia (Figure 5A), the period after BDI is not yet sufficient to conclusively determine if countries in these regions are different from Europe or Asia. Negroids do not account for most of the population of any nation in this analysis, and it is not yet possible to predict whether the infection will spread in Africa at the same rate as in Europe. In multiethnic countries, such as the United States, the racial backgrounds of the patients need to be analyzed carefully. However, such information is not yet available.

SARS coronaviruses utilize ACE2 as an essential receptor for cell fusion for infection [19]. *ACE2*, coded on Xp22 at the location where genes are known to escape X-inactivation, which may contribute to the phenotypic differences between sexes and the tissue-specific differences in X-inactivation [20]. An intronic variant of *ACE2* (rs2285666, also known as G8790A) has been associated with hypertension [21]. Wu YH et al. reported that the serum ACE2 level in individuals with AA genotype was 1.5-fold higher [22]. The ACE2 expression in the tissues was not measured in this study. According to the 1000 Genomes project, the allele frequency of A was the lowest in African (21.1%) and European (23.5%), and the highest in East Asian (53.7%). The frequencies in South Asian (47.9%) and Latin American (33.6%) are middles. If the A allele or other linked variant alleles have a protective effect against SARS-Cov-2, the different allele frequencies among ethnicities might be relevant to the regional differences of COVID-19 epidemiology observed in this study. With regard to the transmission of HIV, the C-C chemokine receptor 5 (CCR5) serves as the predominant co-receptor for a viral entry during the initial transmission. It was well established that a homozygous Δ32 mutation in the *CCR5* gene prevents cell surface expression of CCR5 and thus confers resistance to infection with CCR5-tropic HIV strains [23].

Genetic variants could also modulate patients' responses to treatments for viral infection and some of them have been considered to explain the ethnical difference in treatment efficiencies. It has been reported that the homozygotes of C allele at rs12979860 (C/T) near *IL28B* gene responded twice better in sustained virological response (SVR) more favorably to peginterferon (pegIFN) and ribavirin (RBV) therapy for chronic hepatitis C and B than the heterozygotes [24]. East Asians responded better than European Africans and the C allele frequency in East Asian (92.0%) is higher than those of European (69.1%) and African (33.1%). On the other hand, it has been reported that rs4042 in *CXCL1* gene is related to the levels of some chemokines in the blood

[25] which may suppress the progression of fibrosis in hepatitis C [26]. A frequency of the favorable A allele is higher in order of African, Asian, and Caucasian [27]. Thus, Caucasians may be more likely to suffer from severe fibrosis compared with the other ethnicities.

It has been hypothesized that regional differences in COVID-19 impact could be partially explained by different national policies such as the Bacillus Calmette-Guérin (BCG) childhood vaccination [28]. In general, vaccines exert immune responses specific to a targeted pathogen by producing antibodies. However, it has been reported that BCG vaccination induced genome-wide epigenetic reprogramming of human monocytes which was accompanied by significantly altered responses of innate immune cells [29]. The "trained immunity" by BCG vaccination can protect against a non-related viral infection with an attenuated yellow fever virus vaccine strain by IL-1β-mediated responses. Higher IL-1β levels are also highly relevant to the direct protective effect against viral infection, such as the herpes simplex virus [30]. Even though further investigations are necessary, these observations could support the hypothesis that BCG may contribute to the regional differences in COVID-19.

5. Conclusions

In this study, we showed that the number of population-adjusted cumulative deaths due to COVID-19 increased exponentially for about 20 days from BDI, and then increased linearly after 25 days, in most countries in the world. About 30 days after BDI, significant regional differences in fatalities were found between Asia, the Middle East, Latin America, and Western countries. This regional difference may be affected by governmental policies, age distribution, welfare systems (including BCG vaccination), the medical environment, and social behaviors in each country. However, the possibility of racial differences should also be considered. It is possible that host genetic factors influence individuals' susceptibility, response to treatment as well as prognosis in SARS-Cov-2 infection and investigation of these factors is an urgent issue for more efficient medical care and public health policymaking. In the future, it will be very important to quantitatively explore factors that affect the rate of COVID-19 by further analyzing country-specific information. Although the change in the number of test-positive individuals is an important indicator of the rate of spread of infection, a multiple-fold discrepancy with the number of deaths exists, and thus careful estimation of the number of infected subjects is warranted. Extensive PCR testing carries the risk of causing social disruption by magnifying the initial explosive increase, but our analysis demonstrated that it would be effective for reducing the number of deaths. Although further study is necessary to verify our analysis, earlier expansion of the test capacity seems indispensable. The machine learning analysis presented here is still preliminary but, as the search for infection spreading factors progresses, prediction accuracy will be improved further, and it is expected to be useful for constructing a rational strategy to terminate COVID-19.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, readme.rtf, AnalysisOfCOVID-19_Hisaka.xlsx, PCRtest.xls, data.csv, sample_script.ipynb, sample_script.ipynb.pdf

Author Contributions: Conceptualization and methodology, A.H.; analysis and investigation, A.H and H.Y.; writing—original draft preparation, A.H., H.H., H.S., ; writing—review and editing, A.H., Y.O., N.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1   Ducharme, J. *World Health Organization Declares COVID-19 a 'Pandemic.' Here's What That Means*. Times: Available on line: https://time.com/5791661/who-coronavirus-pandemic-declaration/ (assessed on 10, April, 2020).

2   Li, Q.; Guan, X.; Wu, P.; Wang, X.; Zhou, L.; Tong, Y.; Ren, R.; Leung, K.S.M.; Lau, E.H.Y.; Wong, J.Y.; et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med*. **2020**, *382(13),* 1199-1207.

3   Daily report by Baidu. Available on line: https://voice.baidu.com/act/newpneumonia/newpneumonia/?from=osari_pc_1 (assessed on 10, April, 2020).

4   *COVID-19 in Africa: 'Prepare for the worst, but hope for the best'* Healio: Available on line: https://www.healio.com/infectious-disease/emerging-diseases/news/online/%7B5ff21bfd-f159-4d75-94ea-a660e985683a%7D/covid-19-in-africa-prepare-for-the-worst-but-hope-for-the-best (assessed on 10, April, 2020).

5   Hong, K.H.; Lee, S.W.; Kim, T.S.; Huh, H.J.; Lee, J.; Kim, S.Y.; Park, J.S.; Kim, G.J.; Sung, H.; Roh, K.H.; Kim, J.S.; Kim, H.S.; Lee, S.T.; Seong, M.W.; Ryoo, N.; Lee, H.; Kwon, K.C.; Yoo, C.K. Guidelines for Laboratory Diagnosis of Coronavirus Disease 2019 (COVID-19) in Korea. *Ann Lab Med.* **2020,** (*in press*).

6   Xiao A.T.; Tong Y.; Zhang, S. False-negative of RT-PCR and prolonged nucleic acid conversion in

COVID-19: Rather than recurrence. *J. Med. Virol.* **2020,** (*in press*).

7   Park, M.; Cook, A.R.; Lim, J.T.; Sun, Y.; Dickens, B.L. A Systematic Review of COVID-19 Epidemiology Based on Current Evidence. *J Clin Med*. **2020,** *9(4),* pii: E967.

8   Kermack, W.O.; McKendrick, A.G. Contributions to the mathematical theory of epidemics – I, *Proceedings of the Royal Society Series A*, **1927**, *115,* 700–721.

9   Zhou, W.K.; Wang, A.L.; Xia, F.; Xiao, Y.N.; Tang, S.Y. Effects of media reporting on mitigating spread of COVID-19 in the early phase of the outbreak. *Math Biosci Eng*. **2020** *17(3)*, 2693-2707.

10   European Centre for Disease Prevention and Control. Available on line: https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide (assessed on 5, April, 2020).

11   Wikipedia (English) "COVID-19 testing". Available on line: https://en.wikipedia.org/wiki/COVID-19_testing (assessed on 11, April, 2020).

12   Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat*. **2001**, *29, 5,* 1189-1232.

13   Python Software Foundation. Available on line: https://www.python.org/

14   Pedregosa, F.; Varoquaux, G.; Gramfort, A.; et al. Scikit-learn: Machine Learning in Python, *JMLR.* **2011**, *12, 85*, 2825-2830.

15   scikit-learn. Available on line: http://scikit-learn.org/

16   Bommer, C;, Vollmer S. Average detection rate of SARS-CoV-2 infections is estimated around six percent. Available on line http://www.uni-goettingen.de/de/document/download/ff656163edb6e674fdbf1642416a3fa1.pdf/Bommer%20&%20Vollmer%20(2020)%20COVID-19%20detection%20April%202nd.pdf (assessed on 13, April, 2020).

17   Tang, X.; Wu, C.; Li, X.; Song, Y.; Xinmin, Yao.; Wu, X.; Duan, Y.; Zhang, H.; Wang, Y.; Qian, Z.; Cui, J.; Lu. J. On the origin and continuingevolution of SARS-CoV-2. National Science Review. Available on line nwaa036, https://doi.org/10.1093/nsr/nwaa036 (assessed on 13, April, 2020).

18   Reidy G. A Coronavirus Explosion Was Expected in Japan. Where Is It? Bloomberg: Available on line https://www.bloomberg.com/news/articles/2020-03-19/a-coronavirus-explosion-was-expected-in-japan-where-is-it (assessed on 10, April, 2020).

19   Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature (2020). in press, available on line https://doi.org/10.1038/s41586-020-2180-5

20   Talebizadeh, Z.; Simon, S.D.; Butler, M.G. X chromosome gene expression in human tissues: male and female comparisons. *Genomics* **2006**, *88*, 675–681.

21   Yang, M.; Zhao, J.; Xing, L.; Shi, L. The association between angiotensin-converting enzyme 2 polymorphisms and essential hypertension risk: A meta-analysis involving 14,122 patients. *J Renin Angiotensin Aldosterone Syst.* **2015**, *16(4)*, 1240-4.

22   Wu, Y.H.; Li, J.Y.; Wang, C.; Zhang, L.M.; Qiao, H. The ACE2 G8790A Polymorphism: Involvement in Type 2 Diabetes Mellitus Combined with Cerebral Stroke. *J Clin Lab Analysis.* **2017**, *31,* e22033.

23   Allers, K.; Schneider, T. CCR5Δ32 mutation and HIV infection: basis for curative HIV therapy. *Curr Opin Virol.* **2015,** *14*, 24-9.

24   Ge, D.; Fellay, J.; Thompson, A.J.; Simon, J.S.; Shianna, K.V.; Urban, T.J.; Heinzen, E.L.; Qiu, P.; Bertelsen, A.H.; Muir, A.J.; Sulkowski, M.; McHutchison, J.G.; Goldstein, D.B. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature.* **2009**, *461(7262)*, 399-401.

25   Nischalke, H.D.; Berger, C.; Luda, C.; et al. The CXCL1 rs4074 A allele is associated with enhanced CXCL1 responses to TLR2 ligands and predisposes to cirrhosis in HCV genotype 1-infected Caucasian patients. *J Hepatol.* **2012**, *56,* 758–64.

26   Johansson, S.; Talloen, W.; Tuefferd, M.; Darling, J.M.; Scholliers, A.; Fanning, G.; Fried MW.; Aerssens, J. Plasma levels of growth-related oncogene (CXCL1-3) associated with fibrosis and platelet counts in HCV-infected patients. *Aliment Pharmacol Ther.* **2015,** *42(9)*, 1111-21.

27   Cariaso, M.; Lennon, G. SNPedia: a wiki supporting personal genome annotation, interpretation and analysis. *Nucleic Acids Res. 2012*. **40**, D1308–12.

28   Zwerling, A.; Behr, M.A.; Verma, A.; Brewer, T.F.; Menzies, D.; Pai, M. The BCG World Atlas: a database of global BCG vaccination policies and practices. *PLoS. Med*. **2011**, *8(3)*, e1001012.

29   Arts, R.J.W.; Moorlag, S.J.C.F.M.; Novakovic, B.; Li, Y.; Wang, S.Y.; Oosting, M.; Kumar. V.; Xavier, R.J.; Wijmenga, C.; Joosten, L.A.B.; Reusken, C.B.E.M.; Benn, C.S.; Aaby, P., Koopmans, M.P.; Stunnenberg, H.G.; van Crevel, R.; Netea, M.G. BCG vaccination protects against experimental viral infection in humans through the induction of cytokines associated with trained immunity. *Cell. Host. Microbe*. **2018**, *23(1)*, 89-100.e5.

30   Sergerie, Y.; Rivest, S.; Boivin, G. Tumor necrosis factor-alpha and interleukin-1 beta play a critical role in the resistance against lethal herpes simplex virus encephalitis. *J. Infect. Dis.* **2007**, *196(6)*, 853-860.