GAVIN PUBLISHERS

OPEN ACCESS

Journal of Biostatistics & Biometrics

Research Article

Kravets S, et al. J Biostat Biom: JBSB-110.
DOI: 10.29011/JBSB-110.100010

Ever Changing Statistics of COVID-19 and its Movement

Sasha Kravets[1], Xiaohan Mei[1], Avisek Datta[1], Debarghya Nandi[1], Rathi Dasgupta[2], Dulal Bhaumik[1*]

[1]Division of Epidemiology and Biostatistics, University of Illinois at Chicago, Chicago, USA

[2]Data-Core Systems Inc., Philadelphia, USA

[*]**Corresponding author:** Dulal Bhaumik, Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago, Chicago IL, USA

Citation: Kravets S, Mei X, Datta A, Nandi D, Dasgupta R, et al. (2020) Ever Changing Statistics of COVID-19 and its Movement. J Biostat Biom: JBSB-110. DOI: 10.29011/JBSB-110.100010

Received Date: 05 August, 2020; **Accepted Date:** 11 August, 2020; **Published Date:** 17 August, 2020

Abstract

This study focuses on the concordance of different data sources (e.g. World Health Organization, Centers for Disease Control and Prevention etc.) used for reporting of country wise confirmed COVID-19 infection cases and death rates. It discusses how to estimate related covariates in order to predict COVID-19 confirmed cases and death rates when reporting systems are questionable. It attempts to explain why the pattern of early outbreak of COVID-19 is different from the stabilized portion of the outbreak curve, and examines a long tail nature of its slowing down effect. It also examines the role of temperature, pH level, and humidity in slowing down the spread of COVID-19, and a particular vaccine that many South Asian countries use to protect children from tuberculosis. To illustrate these ideas, the study focuses on China, Italy, Spain, France, South Korea, and the United States, and reveals the policies for varying trends of confirmed cases and death rates of COVID-19.

Keywords: BCG vaccine; Concordance; Confirmed cases; COVID-19; Death rate

Introduction

What started in the Hubei province of China last December has now become a household name all over the world. The effects of COVID-19 are no longer confined only to China but have spread to almost every country in the world; killing thousands of people, ruining the world's economy, putting millions of people under starvation, and leading to uncertain times. Several steps such as isolation, lock down, social distancing rules, regulation of the use of masks and gloves, and repeated use of hand washing and sanitization, are being used both at the individual and mass levels to control the wild spread of the virus and the death rate. Depending on the adherence to the recommendations given by health officials and administrations, several countries have started seeing benefits in terms of reduction of both confirmed new cases and death rates. In the absence of any treatment or vaccine effective particularly for COVID-19, eradication of this infection is elusive.

Practices of alternatives such as Hydrocloroxychloroquine have shown adverse effects in veterans, the effectiveness of convalescent plasma treatment is yet to be evaluated. The question that everybody is asking is "when is it going away" and when we will get back to normal life? Countries such as China, South Korea, and Italy, that were affected early by COVID-19 are trying to come back to "new" normal life, however no such signs are seen in the United States. In order to answer that question, we need a better understanding of the nature of the virus, the pattern of outbreak, how long it takes to reach peak infection rates, and when it will stabilize to a new unknown level. While exploring this, we will investigate if environmental factors such as temperature, humidity etc. play any role in spreading the infection. We will also investigate the varying rates of COVID-19 infection across countries and the role of the BCG vaccination.

Data Collection/Cleaning

We collected daily data reported by the World Health Organization (WHO), beginning from January 22 - April 20, 2020. Daily reports are publicly available through the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. Prior to February 1st, 2020, the United States only reported daily data by state, but once confirmed cases began to rapidly increase, the WHO reported data by cities, and after March 1st, data was reported by county within each state. Country data was reported for the whole country, with exceptions such as Australia or Canada that reported data by province or region. Additionally, for countries who own territories,

we grouped those territories to be labelled as their own country. We investigated state level data for the United States, and country level data for the remaining countries of interest. Because daily reports of cases and deaths were reported on a cumulative basis each day, we created the daily cases and deaths by differencing the cumulative data with the day prior.

Data Concordance

This was an observational study, so we depended on several data reporting systems, and evaluated concordance between those different reporting systems. We investigated data reported by the WHO and Centers for Disease Control and Prevention (CDC) types of organizations of respective countries. It should be noted that there is no independent organization that collected data or verified data reported by China. No wonder, we see a good concordance between the WHO data and the National Health Commission of the People's Republic of China (NHCC) data on COVID-19, with a lagging of two days. For France, Spain, Italy, and South Korea, we examine concordance between data reported by the European CDC (ECDC) and the WHO. Each data source varied in its timeframe of data availability and reporting, and those minor differences in data collection lead to some discordance of data on a daily basis. However, we can account for a 1 to 2 days lag of data reporting and proceed with an evaluation of data concordance. In Figure 1, without accounting for any daily lag in data reporting, we can see that the data reporting for four countries of interest (Italy, France, Spain, and South Korea) is almost identical.

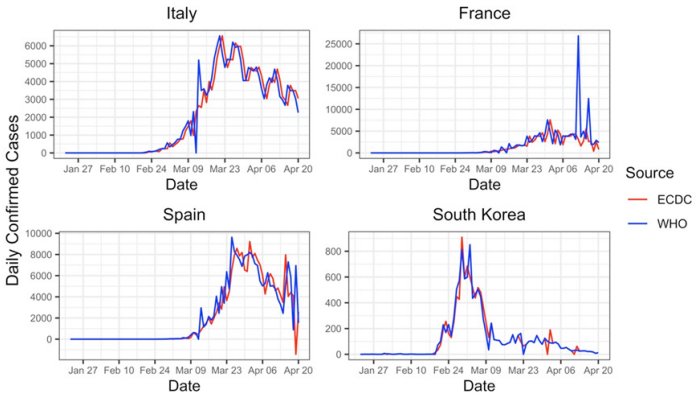

Figure 1: Concordance between data reported by the ECDC and the WHO.

Similarly, Figure 2 shows that the data provided for the United States is extremely similar between the ECDC and the WHO.

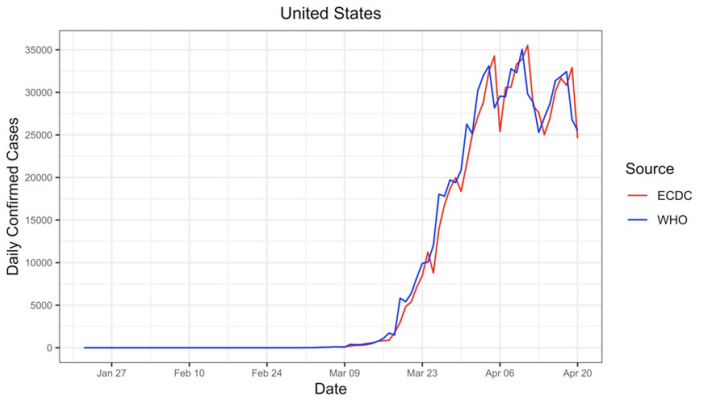

Figure 2: Concordance between data from the ECDC and the WHO for the United States.

The concordances between the WHO and ECDC data are more than 90 percent for each country (Italy: 0.9642, Spain: 0.9001, United States: 0.9913, South Korea: 0.9682), except France. The CCC computed for France is 0.4381. Graphically we can see that there is concordance in the number of daily cases between the WHO and ECDC, however, the spike in cases on April 12, 2020 greatly inflates the average number of cases reported by the WHO, thereby lowering the CCC. Prior to April 12, the CCC is 0.8624. There are some obvious data entry errors, for example, the reported daily confirmed cases in Spain on April 19 are -1430. Both the ECDC data and WHO data report such errors and tend to correct these mistakes within several days.

Statistical Modelling of COVID-19 Data

After observing the high concordance between data from the WHO and CDCs of corresponding countries, we chose to work with data provided by the WHO. We first looked at the daily reported confirmed cases of COVID-19 in Hubei province from January 22 - March 5. In order to model the trend of daily cases, we fit an autoregressive model of order 1. An AR (1) model assumed that each days daily confirmed cases were correlated with the previous days' number of confirmed cases. Estimated model parameters were used for prediction of daily-confirmed cases for the aforementioned duration. A Chi-Square test was conducted to assess the fit of the predicted model and found that the model fit well for the predicted daily cases in Hubei province (p-value = 0.16).

Figure 3 shows the reported confirmed daily cases and the predicted daily cases. We can see that from January 22 through

approximately February 12, the predicted model shows a fairly stabilized amount of daily cases, around 2000 cases per day. The model then captures the large spike seen on February 13. This spike can be attributed to a change in the evaluation of disease as defined by the NHCC. The model then predicts a larger amount of daily cases compared to those reported from February 18 - March 5 in order to adjust for the under estimation of cases prior to the spike on February 13. Using these predicted daily cases, we can compute the predicted cumulative cases.

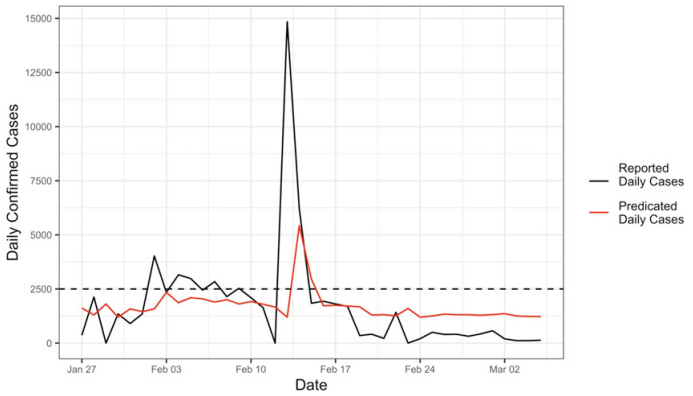

Figure 3: Reported and Predicted Daily Confirmed Cases of COVID-19 in Hubei Province.

Discrepancies in Daily Confirmed Cases and Death Rates

Multiple discrepancies in data reporting systems are observed across countries and over time. For example, in almost all countries, confirmed COVID-19 cases are/were determined by testing only those patients who have/had severe symptoms. This approach of detecting confirmed cases completely ignores asymptomatic and less severe cases. Thus, many true cases remain unnoticed and uncounted, and as a result, the death rate (number of deaths divided by number of confirmed cases), becomes inflated. While following this approach, some countries (e.g. United States, Japan, and South Korea) adapted their testing strategies for a more systematic testing approach. The large inflation of reported confirmed cases in Hubei province on February 13 can be contributed to this reason (Figure 3). It is extremely hard, probably impossible, to develop a standard time invariant measure for computing the daily death rate for all countries.

Instead, we rely on the reported data and compute the daily death rate by taking the ratio of the number of deaths to the number of confirmed cases, with a lagging of k days. Thus, on day t, if X_t is the number of deaths, and Y_(t-k) is the number of confirmed cases on day t-k, the death rate on day t is X_t/Y_(t-k), where k is determined to minimize the variation of X_t/Y_(t-k) over t. For a constant death rate, X_t/Y_(t-k) will not vary over time, thus a Chi-square test can be used to check the significance of fluctuations. Even though it is expected that the value of k may vary between countries, we chose k = 6 for Italy, France, Spain, South Korea, and the United States, and k = 7 for China, to minimize the variation over time. In what follows, we compare daily death rates of several countries along with two epicenters, Hubei province and New York State. Figure 4 shows country wise confirmed cases in Italy, France, Spain, South Korea, the United States and China. As no universal procedure is used to measure the confirmed cases, reported cases between countries should not be compared for drawing any valid inferences.

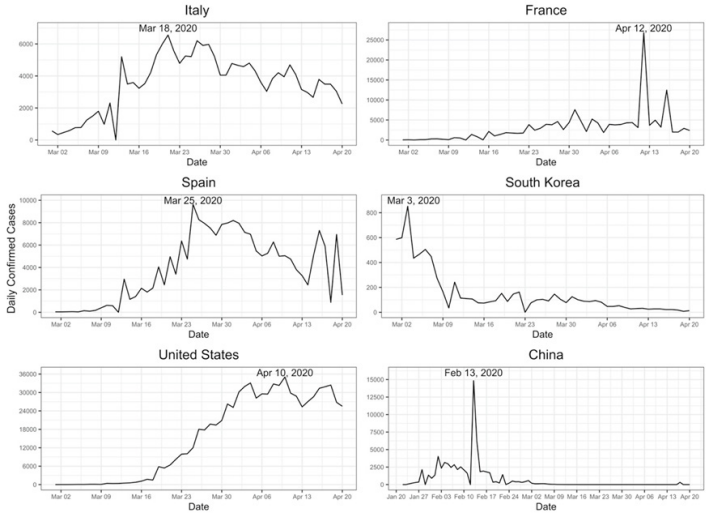

Figure 4: Confirmed COVID-19 cases in Italy, France, Spain, South Korea, the United States, and China.

We see in Figure 4 that South Korea controlled new confirmed cases within in a week, whereas the United States has not done so in two months. Italy and Spain have very similar patterns of confirmed cases, and eventually the curves of Italy, Spain, and France are flatting down. Spreading of infection in China started in January, and the spike on February 13 can be attributed to the change in the data collection policy. However, no reason is given for a drastic reduction of new confirmed cases within three days of the peak. Figure 5 displays daily death rates (i.e. proportion of daily deaths) in Italy, France, Spain, South Korea, the United States, and China. As no universal standard was maintained while reporting these death rates, a direct comparison among countries should not be appropriate to evaluate. However, we can discuss some common features and contrasts between these countries. (i) Except China, we see that it takes approximately three weeks from the peak point for the death rate to stabilize, and for China it takes about two weeks (January 28-February 10), (ii) China, Italy and France start with high daily death rates, whereas South Korea, Spain, and the United States start reporting with very low death rates.

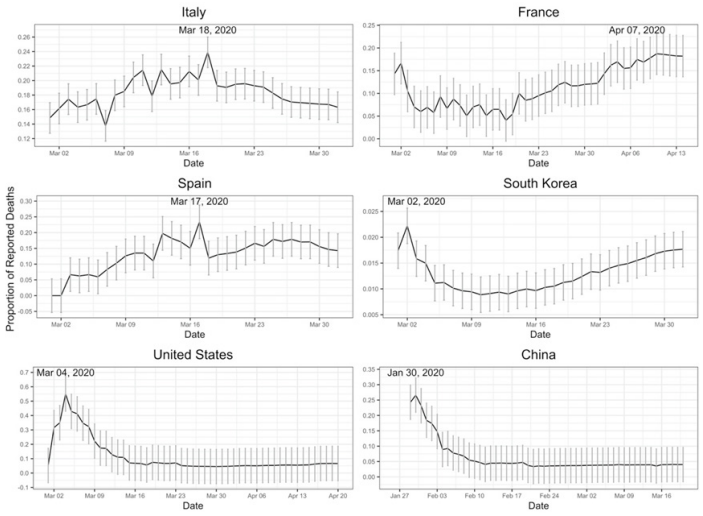

Figure 5: Proportion of Reported Deaths in Italy, Spain, South Korea, France, the United States, and China.

In Figure 6, we compare the proportion of daily deaths in two epicenters, Hubei Province in China, and New York State (NY) in the United States. There is a sharp contrast between these two figures. We see that in Hubei, the daily death rate starts with a high value, then decreases rapidly, and stabilizes in two weeks. Whereas in NY, the daily death rate starts with a low value, increases, and has not stabilized as of April 22. (i) More precisely, Hubei does not report data from the start of the infection, and NY does not have data to show stabilization yet. Based on the 18 thousand cumulative deaths in NY from March 14 - April 20, we can project that at least that many deaths would have been unreported in Hubei province. This argument shows that (i) reported deaths in Hubei are inconsistent and under reported. (ii) Stabilization of death rates in Hubei in two weeks compared to three weeks in all other countries under consideration raises another concern of underestimation.

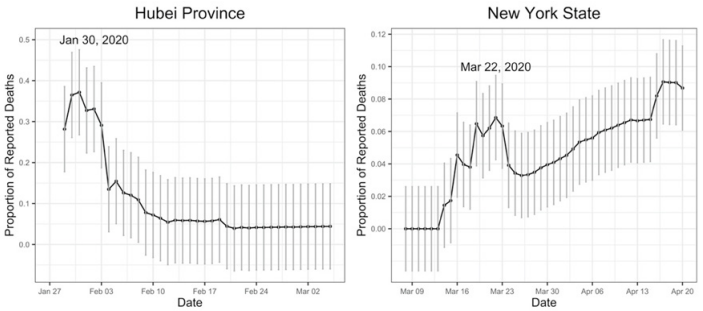

Figure 6: Proportion of Reported Deaths in Hubei Province and New York State.

In order to address the underestimation problem raised in (i) and (ii), we compare the population sizes of NY (19 million) and Hubei (59 million) and try to predict the number of deaths in Hubei before February 2020. The majority of COVID-19 cases in NY can be attributed to cases from New York City alone, and the majority of cases in Hubei province can be attributed to cases from the city of Wuhan. We note that the population sizes of New York City and Wuhan are respectively 8.4 million and 11.2 million, which are very much comparable. Wuhan started to treat its COVID-19 patients rigorously by expanding the medical facilities, admitting confirmed cases in medical centers, locking down the entire province, and isolating its citizens in February, not in January. The same procedures started in New York City from the second week of April. Both cities are epicenters of COVID-19 infections, unequipped to treat confirmed cases at the beginning, and as a result, observed very high death rates. In addition, both New York City and Wuhan have very dense populations.

In NY, there were a total of 2,373 deaths by April 2 approximately two weeks from the start of COVID-19 related deaths. Comparing population sizes of NY and Hubei and considering the cumulative deaths in NY by April 2, we predict about 7,000 cumulative deaths in Hubei alone before February 2020 due to COVID-19. Our prediction also matches with the number of patients admitted to treatment facilities in Hubei as well as in Wuhan. Around February 2, there were 27,000 beds available in different hospitals and medical facilities in Wuhan to treat COVID-19 patients. At a 4 percent conservative death rate for 27,000 patients over a period of two weeks (the last two weeks of January 2020), the total number of deaths in Wuhan alone becomes close to 16,000. Certainly, in Hubei Province the total estimated deaths reported in January alone should have been more than 16,000, and that number for all of China should have been drastically more than 16,000.

Massachusetts Data

As we have discussed the death rate in several countries, we then look into characteristics of patients who have died of COVID-19. In this section we discuss data that was available from the Massachusetts Department of Public Health, showing characteristics at the patient level. We would like to examine how the death rate varies over age and gender, as well as whether patients with comorbidities are more vulnerable, and whether the majority of patients were hospitalized. The proportion of reported deaths in Massachusetts is presented in Figure 7. The Massachusetts Department of Public Health provided patient level data for 146 patients that have died of COVID-19 from March 27 - April 19, 2020. As of April 20, Massachusetts has estimated a total of 1807 deaths. In Table 1, we see of those 146 deaths, almost 50 percent are male, 86 percent are above 60 years old, 53 percent were admitted to the hospital, and 98 percent have an underlying condition. This supports the idea that the most vulnerable patients are older and/or have comorbidities.

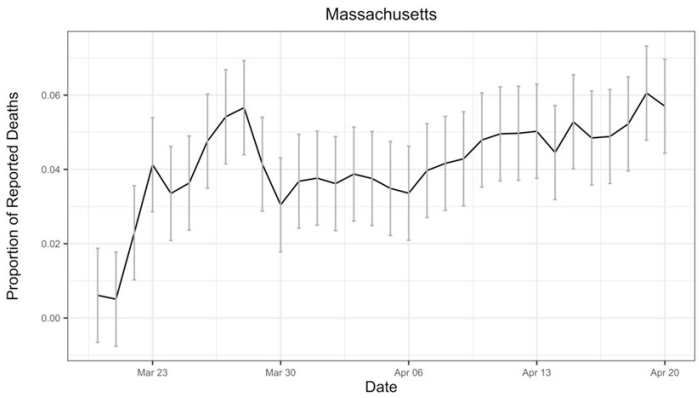

Figure 7: Proportion of Reported Deaths in Massachusetts.

Variable	Category	N (Percent)
Sex	Female	74 (50.7)
	Male	72 (49.3)
Age Category	30s	(0.68)
	40s	(0.68)
	50s	(1.37)
	60s	(11.64)
	70s	(18.49)
	80s	(41.10)
	90s	(25.34)
	100s	(0.68)
Hospitalization	No	11 (7.5)
	Yes	77 (52.7)
	Unknown	58 (39.7)
Preexisting Conditions	No	1 (1.7)
	Yes	59 (98.3)
	Unknown	86

Table 1: Characteristics of Deceased Patients in Massachusetts State.

Environmental Effect on Transmission

Next we address how the transmission of COVID-19 travels through air and how Relative Humidity (RH) and pH of the water in air affects such spread of infection. According to a study published by a group in Applied Environmental and Microbiology, 2012, many airborne viruses are sensitive to humidity and temperature [1]. A thorough understanding of this phenomenon may provide insight into the temporal and spatial distribution of diseases. For instance, multiple studies suggest atmospheric humidity is an important environmental determinant in the transmission of influenza in temperate regions [2-4]. Studies revealed that Influenza A (W.S. strain) under a pH (<5) has viability highest at 30-34 percent RH, lowest at 58-60 percent RH; whereas Influenza A (WSN strain) under pH (<5) is most viable at <40 percent RH, and lowest at 40-60 percent RH. SARS COV is stable at 40-50 percent RH and becomes rapidly ineffective at higher RH (>50 percent) [5]. A very recent study (March 9, 2020) reported by Wang, Tang, Feng, Lv clearly identifies that the COVID-19 virus is most active at lower temperature (< 36˚C) and has a negative gradient towards higher temperature (>36˚C) [6].

A similar regression is observed for RH from 35-85 percent. This paper investigates how air temperature and humidity influence the transmission of COVID-19 [6]. After estimating the serial interval of COVID-19 from 105 pairs of the virus carrier and infection, authors calculated the daily effective Reproductive Rate for each of all 100 Chinese cities with more than 40 cases. Using the daily reproductive rate values from January 21-23, 2020 as proxies of non-intervened transmission intensity, they found, under a linear regression framework for 100 Chinese cities, high temperature and high RH significantly reduce the transmission of COVID-19, even after controlling for population density and GDP per capita of those study cities. Increment of temperature by one degree Celsius and relative humidity by 10 percent yielded a lower effective production of the virus by 0.0383 and 0.0224 respectively. This result is consistent with the fact that high temperature and high humidity significantly reduce the transmission of influenza. It indicates that the arrival of summer and rainy season in the northern hemisphere may effectively reduce the transmission of the COVID-19.

Effects of Bacillus Calmette-Guerin (BCG) Vaccination

Infection and death rates due to COVID-19 have a large variation across countries; for example, the peak death rate in Italy at one time was close to 26 percent, whereas in India it is not even 0.01 percent up to this point. In addition to many factors, such as socioeconomic condition, varying percentages of older population, different health structures, scientists are asking whether the universal BCG vaccination at early childhood provides any protection for less infections and lower death rate among COVID-19 confirmed cases. BCG is a live attenuated strain derived from an isolate of Mycobacterium bovis. In many countries (e.g. India, China, Japan), it is primarily used as a vaccine given at early childhood for Tuberculosis (TB).

Miller et al. conducted a study using data from 28 countries where there is a universal BCG vaccination policy [7]. They

observed a positive significant correlation of ρ=0.44 (p-value = 0.02) between the age of subjects when universal BCG vaccination is implemented and the mortality rate. This result is very much consistent with the literature that the earlier (age) the policy is implemented; the larger fraction of the elderly population will be protected. For example, Iran established the universal BCG vaccination policy in 1984, and has a mortality with 19.7 deaths per million, whereas Japan implemented the universal BCG vaccination policy in 1947 and has mortality of only about 0.28 per million. This may justify the surprisingly lower death rate in India due to COVID-19.

Effect of Hydrocloroxychloroquine (HCQ)

A debate is still going on whether HCQ has any protective effect in terms of reducing the infection rate of COVID-19. HCQ is primarily prescribed to prevent malaria and is also an effective medication for autoimmune conditions (e.g. rheumatoid arthritis, lupus). A recent retrospective study of Veterans Affairs (VA), United States, suggests that HCQ with or without Azithromycin (AZ) cannot lower the risk of ventilation needed by patients hospitalized with COVID-19 (Hydroxychloroquine Ineffective for COVID-19, VA Study Suggests - Medscape - Apr 23, 2020) [8]. This study had three arms: HCQ + Standard Care (75), HCQ + AZ + Standard Care (n = 113) and Standard Supportive Care (n = 158). The study shows an increased risk of death among study patients treated with HCQ alone. The risk of death of the HCQ + Standard Care compared to the Standard Support Care is significantly higher (adjusted hazard ratio 2.61; 95 percent confidence interval: 1.10 - 6.17 with p-value = .03) but the HCQ+AZ group did not show such results when compared with the Standard Support Care (HR, 1.14; 95 percent CI, 0.56 - 2.32; p-value = 0.72). Another randomized trial with two equal (75) arms (HCQ vs Standard Care) performed in China also showed no benefit but more side effects of HCQ according to a preprint posted online on April 14 [8].

Discussion

Even after several months of the devastating outbreak of COVID-19, both confirmed daily new cases and deaths have not stabilized in the United States. The severe impact on human life, mental health, and the economy has disrupted normal life all over the country. Official statements that downplay the severity of COVID-19, scarcity of protective materials, and lack of proper guidance have compounded the problem even further. At this critical moment, a large segment of the population is trying to maintain social distancing, avoid travelling, and imposing self-quarantine, whereas others are asking to restore their personal freedoms curtailed due to COVID-19. It is now established that gatherings of more people without proper protection increases both the likelihood of spreading the infection and death rate with no limit.

As lockdown regulations are lifting, the situation is even more unclear. Personal protection equipment is not only expensive but also hard to obtain for the general public. Medicines to treat COVID-19 patients, or vaccines to develop immunity are yet to be expected by the end of this year. Under the current situation without any medication or vaccine, more research is needed to determine how to reduce the spreading of the infection and protect individuals. A better understanding of protective environmental factors will help us to maintain an optimal condition at our vicinity for reducing the spreading rate of the infection. As testing becomes more accessible, more data is collected, and a vaccine becomes available, we will unravel the mysteries of COVID-19 and find a way to return to normalcy.

References

1. Wan Wang and Linsey C Marr (2012) Mechanism by which ambient humidity may affect viruses in aerosol, Applied Environmental Molecular Biology 78: 6781-6788.

2. Wang J, Tang K, Feng K, Lin X, Lv W, et al. (2020) High Temperature and High Humidity Reduce the Transmission of COVID-19. SSRN.

3. Hemmes JH, Winkler KC, Kool SM (1960) Virus survival as a seasonal factor in influenza and poliomyelitis. Nature 188: 430-431.

4. Shaman J, Goldstein E, Lipsitch M (2011) Absolute humidity and pandemic versus epidemic influenza. Am J Epidemiol 173: 127-135.

5. Shaman J and Kohn M (2009) Absolute humidity modulates influenza survival, transmission, and seasonality. Proc Natl Acad Sci USA 106: 3243-3248.

6. Wang H, Yang P, Liu K, Guo F, Zhang Y, et al. (2008) SARS coronavirus entry into host cells through a novel clathrin- and caveolae-independent endocytic pathway. Cell Res 18: 290-301.

7. Miller A, Reandelar MJ, Fasciglione K, Roumenova V, Li Y, et al. (2020) MedRxiv. Correlation between universal BCG vaccination policy and reduced morbidity and mortality for COVID-19: an epidemiological study.

8. Marcia Frellick M (20200 Hydroxychloroquine ineffective for COVID-19, VA study suggests. The Hospitalist.