ORIGINAL ARTICLE PHYSIOLOGY

NORTH CLIN ISTANB 2021;8(4):321–331
doi: 10.14744/nci.2021.42492

Global evaluation of coronavirus disease 2019 cases and clustering of similar countries

Handan Ankarali,[1] Seyit Ankarali,[2] Erkan Kilinc[3]

[1]Department of Biostatistics, Istanbul Medeniyet University Faculty of Medicine, Istanbul, Turkey

[2]Department of Physiology, Istanbul Medeniyet University Faculty of Medicine, Istanbul, Turkey

[3]Department of Physiology, Bolu Abant Izzet Baysal University Faculty of Medicine, Bolu, Turkey

ABSTRACT

OBJECTIVE: It was aimed to be obtained descriptive values with respect to the outbreak time course, demographic structure, and symptom distribution by the help of case-based data, and to be compared countries by being grouped according to their similarities of outbreak indicators.

METHODS: The data were obtained from open-access database. Univariate tests and cluster analysis were used to analyze the data.

RESULTS: After the symptoms onset, the prolonged admission to the hospital significantly increases the risk of death. The average age and percentage of the male gender of the deceased cases were found to be significantly higher. In addition, the symptoms including fever, throat complaints, and dyspnea were determined in 70%. Countries were divided into four clusters according to their similarities in terms of three outbreak indicators. The differences among the clusters with regard to mean age, urban rate, and average of the outbreak indicators were found significant.

CONCLUSION: Delaying treatment from the moment the symptoms appear will increase the risk of death and the average time to recovery or death was 2.5 weeks. It can be stated that the most important measure is to focus on methods that can detect the cases before symptoms. The indicators that have a very important role in defining the pandemic are also related to each other. Therefore, multivariate methods, which take these relationships into account, are able to produce more accurate information in determining the similarities of countries.

Keywords: Coronavirus; coronavirus disease 2019; expectation maximization clustering; outbreak indicators; pandemic.

Cite this article as: *Ankarali H, Ankarali S, Kilinc E. Global evaluation of coronavirus disease 2019 cases and clustering of similar countries. North Clin Istanb 2021;8(4):321–331.*

As of mid-April, 2020, there are coronavirus disease 2019 (COVID-19) cases in almost every country in the world. The first cases with regard to COVID-19 have appeared in the city of Wuhan in Hubei Province, China, in December 2019, but it turned into a pandemic that surrounds the world in the past 3 months [1, 2]. The infection ranks first on the world agenda, because of the spreading speed of the disease is very high and its effects lead to serious consequences [3].

When the biological characteristics of the outbreak are examined, despite the fact that the outbreak transumes to a stationary process due to a rapid increase in the number of cases in a certain period and then a decrease in new cases, both the spreading speed and its effects exhibit diversity with respect to countries. There could be many reasons for this difference. Investigation of the mentioned reasons is of great importance in terms of taking necessary precautions and developing appropriate policies.

Received: December 04, 2020 *Accepted:* January 30, 2021 *Online:* August 26, 2021

Correspondence: Erkan KILINC, MD. Bolu Abant Izzet Baysal Universitesi Tip Fakultesi, Fizyoloji Anabilim Dali, Bolu, Turkey.
Tel: +90 533 138 04 83 e-mail: e_kilinc_27@hotmail.com

To the best of our knowledge, as of mid-April 2020, there is no clear information on the disease process that defines the time from the first exposure to the onset of symptoms and admission to the hospital. In addition, there is limited numerical information about air temperature, density in country, and in cities, age-sex ratio, and lung cancer prevalence which may be associated with coronavirus prevalence but has not yet given absolute results. Furthermore, there is a need to investigate the effects of precautions for the pandemic such as school closure policy, public space restrictions, and staying at home. Although various comparisons are performed through the help of indicators, analyzing these indicators together will result in more accurate results. For a comparative analysis of countries, the duration of the fight against the pandemic should also be taken into account.

For these reasons, the primary objective of this study was to define the disease process with the help of a case-based data set and to investigate its relationship with the country and individual characteristics. The second aim was to cluster the countries according to their similarities using the outbreak indicators which are total death in one million, totally recovered in 1 million, and total active cases in 1 million with together in April 2020, and to compare these clusters.

MATERIALS AND METHODS

Considering the time of data collection, the present study is a cross-sectional study type covering the time between the onset of the outbreak and mid-April, 2020.

Two different data sets were used in the study. The data set used to investigate the first aim of the study was obtained from the Kaggle database, and the data set used to examine the second aim of the study was obtained from World Bank, Kaggle and Worldometers databases. All recorded data were used in the evaluation. The indicator results of the countries consist of daily records, and country characteristics are also obtained from the Worldometers/population website.

Data for First Objective and Variables

For evaluation of the first aim, the data set obtained from the Kaggle database containing information of 1085 cases from the 32 different countries was used [4]. It was recorded genders, ages, first exposure days to COVID-19, symptoms onset day, 1st consultation days to the hospital, and recovery/death days in the finalized outcomes. Furthermore, the symptoms of the disease were defined. Taking into account these data, COVID-19 cases were described from various perspectives.

Highlight key points

- After the symptoms onset, the prolonged admission to the hospital significantly increases the risk of death.
- The average recovery or death time of the cases varies between 2 and 3 weeks.
- The average age and percentage of the male gender of the deceased cases are significantly higher.
- Countries are divided into four clusters according to their similarities in terms of three outbreak indicators including total numbers of the deceased, the recovered and active cases per 1 million people.

Data for Second Objective and Indicators

For evaluation of the second aim, the data set obtained from the World Bank, Kaggle, and Worldometers databases on April 15, 2020, was used [5–9]. All countries with a total number of cases over 100 were evaluated. The data set has the total number of cases, the daily number of new cases, the total number of deaths, the total number of recovered cases, the number of active cases, and the number of serious critical cases of 144 countries.

Statistical Analysis

In the first data set, the age and gender distribution of the cases resulting in death due to the COVID or the recovered cases were analyzed by independent samples t-test and Pearson Chi-square test. The cases that died and recovered, in terms of the time (days) from the exposure of cases to the hospital admission, were compared by the Mann–Whitney U-test. In the second data set, 144 countries were clustered according to their similarities by expectation maximization (EM) cluster analysis in terms of the total numbers of the death, the recovered and the active cases per 1 million people on April 15, 2020. The clusters were compared by the univariate tests in terms of demographic characteristics of countries, the restrictions applied due to the outbreak, the elapsed time since the first case (duration of the fight against the outbreak), and the outbreak indicators.

Cluster Analysis

Cluster analysis is an unsupervised learning algorithms. Unsupervised systems are not provided by any training examples. In these algorithms, all variables are treated in

TABLE 1. Descriptive statistics of the age of the cases and important periods for COVID-19

	Age	Days			
		Time from first exposure to onset of symptoms	Time from onset of symptoms to hospital visit	Time from onset of symptoms to death	Time from onset of symptoms to recovered
n	843	85	443	7	99
Mean	49.48	7.96	2.90	17.86	19.14
Mode	55.0	4	0	–	17
SD	18.25	6.75	3.33	7.45	6.4
Minimum	0.25	0	0	7	5
Maximum	96.00	34	19	27	35
Percentiles					
25th	35.0	4.0	0.00	9.0	15.0
Median	51.0	6.0	2.0	20.0	19.0
75th	64.0	9.0	4.0	24.0	23.0

COVID-19: Coronavirus disease 2019; SD: Standard deviation.

the same way. There is no distinction between explanatory and dependent variables. Cluster analysis is organizing data into clusters or groups such that they have high intracluster similarity and low intercluster similarity. EM algorithm was used for the clustering of the countries according to total death cases, total recovered cases, and total active cases algorithm, and it applies by calculating the similarity based on the distance measurement method. The algorithm functions by alternating EM and split operations. Bayesian Information Criterion (BIC) and Schwarz Criterion were used for the selection of the number of clusters. The inputs of the algorithm are cluster number k, data set, and stopping tolerance. Its outputs are a set of k-clusters with a weight that maximizes log-likelihood function and minimizes information criteria such as BIC and Schwarz information criterion (SBIC) [10]. WEKA software (ver. 3.8.4) was used for the analysis.

RESULTS

Evaluation Results of the First Objective

Of the 1085 cases evaluated, 197 (18.2%) belonged to China, 190 (17.5%) to Japan, 114 (10.5%) to South Korea, 94 (8.7%) to Hong Kong, 93 (8.6%) to Singapore, and 54 (5%) to Germany and the remaining 343 (31.6%) cases belonged to 32 countries in various continents of the world. Gender information of 802 (74%) of the cases was reached and 520 (64.8%) of them were male. In addition, the age information of 843 individuals were reached and the average age is 49.5±18.25, and the ages vary from 0 to 96 years. However, value of the most frequently recurrent age was 55 and the median age value was 51 (Table 1). Descriptive values of some important duration concerning the disease are given in Table 1, Figure 1A and B. According to Table 1, the mean value of the time from first exposure to onset of symptoms was 7.96 days. Furthermore, the most often day was 4 and median value was 6 days. When the time from onset of symptoms to hospital visit was calculated using 443 case, it was observed that the mean value was obtained as 2.9±3.33 days. However, the most observed value of this time was found to be 0 day. The date of onset of symptoms and the date of death were recorded in 7 out of 63 cases reported to result in death. The mean value of the time from onset of symptoms to death was calculated as 17.86±7.45. Out of 143 cases, who were definitely recovered from the disease, 99 of the healing dates and the first symptom date were recorded. The mean value of the difference between two time points was obtained as 19.14±6.4 days, and also, it varies from 5 to 35 days. The mean age of the deceased cases (68.6±13.6) was significantly older than recovered cases (41.4±16.5) ($p<0.001$). In addition, the mean day from onset of symptoms to hospital visit in the deceased cases was 5.44±3.5, and it was found that this time was significantly longer than that in the recovered cases (2.18±2.65) ($p<0.001$). When the gender distribution of the cases, who definitely recovered or died from

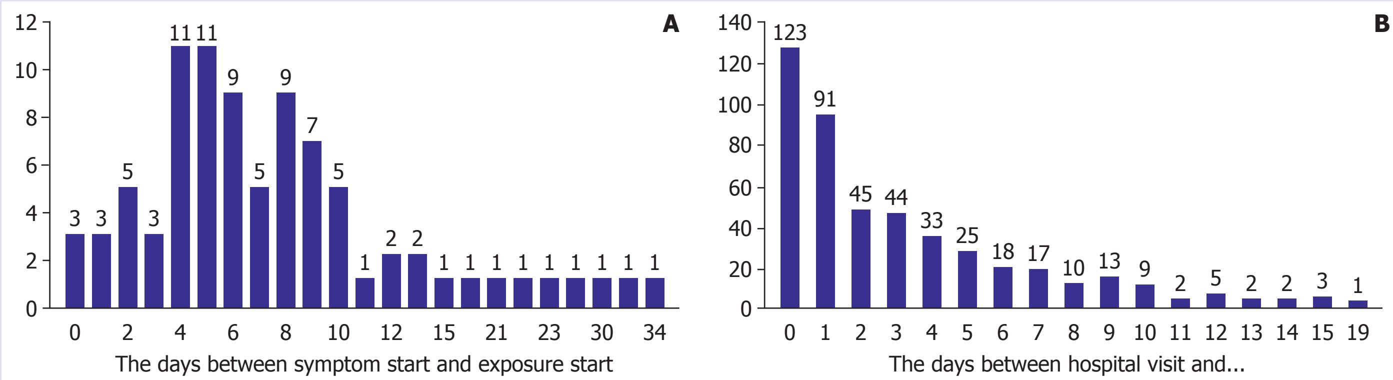

Figure 1. **(A, B)** Distributions of the days between symptom onset day and the first exposure day, and also hospital visit day and symptom onset day.

Table 2. Countries in each cluster

Cluster No.	Countries included in clusters
1	Austria, Bahrain, Belgium, Brunei, Canada, Denmark, France, Germany, Iran, Ireland, Israel, Italy, Luxembourg, Netherlands, Norway, Portugal, Qatar, Spain, Sweden, the United Kingdom, the USA
2	Iceland, Switzerland
3	Afghanistan, Argentina, Bangladesh, Bolivia, Cambodia, Colombia, Congo, Costa Rica, Cuba, Egypt, El Salvador, Georgia, Ghana, Honduras, India, Indonesia, Iraq, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Kyrgyzstan, Lebanon, Madagascar, Mali, Mexico, Morocco, Niger, Nigeria, Pakistan, Paraguay, Philippines, Rwanda, Senegal, South Africa, Sri Lanka, Taiwan, Thailand, Tunisia, Ukraine, Uzbekistan, Venezuela, Vietnam
4	Albania, Argentina, Armenia, Australia, Azerbaijan, Belarus, Bosnia and Herzegovina, Brazil, Bulgaria, Chile, China, Croatia, Cyprus, Czechia, Djibouti, Dominican Republic, Ecuador, Estonia, Finland, Greece, Hungary, South Korea, Kuwait, Latvia, Lithuania, Malaysia, Malta, Mauritius, Moldova, Montenegro, New Zealand, North Macedonia, Oman, Panama, Peru, Poland, Romania, Russia, Saudi Arabia, Serbia, Singapore, Slovakia, Slovenia, Trinidad and Tobago, Turkey, the United Arab Emirates, Uruguay

the disease, was examined, it was determined that the death rate was 8.5% for males and 3.7% for females and this difference between them was significant ($p<0.001$).

The data set contained the symptoms of 271 cases. The fever has been observed 76% (n=206) times and throat complaints has been found 56% (n=152) times. The distribution of symptoms is presented in Figure 2.

Evaluation Results of the Second Objective

In the second data set, 114 countries were clustered using the three outbreaks indicators. The three indicators used in this study for clustering are the most important criteria used in tracking the outbreak progress. These indicators are the total number of deaths in 1 million, the total number of recovered in 1 million, and the total number of active cases in 1 million, the sum of which gives the total number of cases. However, each has a different meaning and importance in itself while watching the course of the epidemic. Despite this, evaluating the similarities and/or differences of countries by considering each indicator separately leads to a loss of information. Therefore, three indicators were analyzed together. The cluster number with the smallest information criterion was determined as 4. After clustering, the clusters were compared in terms of preventive anti-epidemic measures and some demographic characteristics of the countries.

In the first stage of the cluster analysis, the numbers of total deaths, the numbers of total recovered, and the numbers of total active cases were standardized by considering the population sizes of the countries. As a result, four clusters of countries were obtained that were similar in itself but different among themselves, according to three indicators (Table 2, Fig. 3).

When the four clusters are compared with regard to the country characteristics, it was found that (i) the number of countries implementing quarantine policy

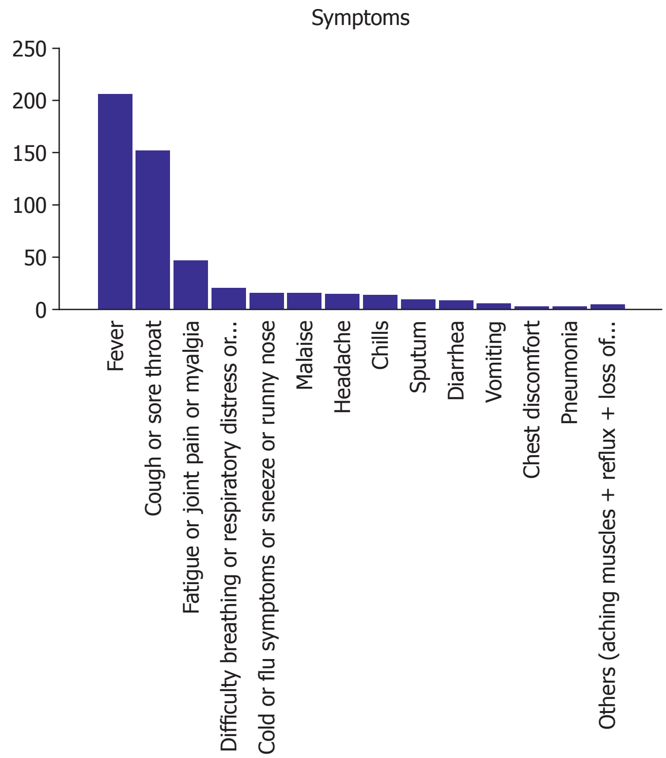

FIGURE 2. Distributions of the symptoms of 271 cases with coronavirus disease 2019 (usually one case has more than 1 complaint).

was found significantly higher in the first cluster than the other three clusters (n=12, 57.1% and p=0.001), (ii) two countries in the second cluster did not implement this policy, (iii) 6 of the 44 countries in the third cluster applied the quarantine (13.6%), and (iv) 7 of the 47 countries in the fourth cluster (14.9%) also applied the quarantine. In addition, it was found that (i) countries implementing a school closure policy were significantly higher (n=14, 66.7%) in the first cluster (p=0.001), (ii) one of the two countries in the second cluster implemented this policy, (iii) six countries (13.6%) in the third cluster took a decision in respect to the school closure, and (iv) four countries (8.5%) in the fourth cluster also took a decision in respect to the school closure.

Moreover, when countries that apply the initial date of public places restrictions rule are examined, it was found that this rate was 33.3% in the first cluster, 0% in the second cluster, 4.5% in the third cluster, and 8.5% in the fourth cluster. According to these results, it is seen that the number of countries in the first cluster that implemented this restriction was significantly higher than the others (p=0.012). The initial date of non-gathering restriction has been implemented by 23.8% of countries in the first cluster, 50% of countries in the second cluster, 9.4% of countries in the third cluster, and 4.3% of the countries in the fourth cluster. When the result obtained due to the fact that there are two countries in the second cluster which is not included in the interpretation, it is seen that the countries included in the first cluster further complied with this restriction significantly (p=0.026). When the demographic structures of the countries included in the clusters are compared, the results are given in Table 3. There was no significant difference among the clusters in terms of the amount of citizens per square

TABLE 3. Descriptive statistics of the other characteristics of the countries in each cluster

	Cluster 1 (n=21)		Cluster 2 (n=2)		Cluster 3 (n=44)		Cluster 4 (n=47)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
The amount of citizens per square meter	270.9	474.9	111.0	152.7	185.9	238.1	454.9	1574.9
Median age	39.5	5.14	40.0	4.24	27.9	7.45	37.98	5.47
The % of people who lives in urban locations	79.2	13.23	84.0	14.14	55.9	20.6	73.23	15.05
Amount of males per female	1.10	0.54	0.99	0.03	0.99	0.04	1.00	0.20
Death rate from lung diseases per 100 k people	20.63	6.81	17.47	3.17	29.41	21.14	18.07	6.87
Average temperature in Celsius between January and April	12.41	7.80	5.07	6.19	22.95	9.71	15.58	9.41
Average percentage of humidity between January and April	73.14	10.17	76.68	2.76	62.65	14.08	69.25	12.71
Days after 1st case(s)	56.14	12.43	48.50	2.12	45.52	15.59	46.94	12.23

SD: Standard deviation.

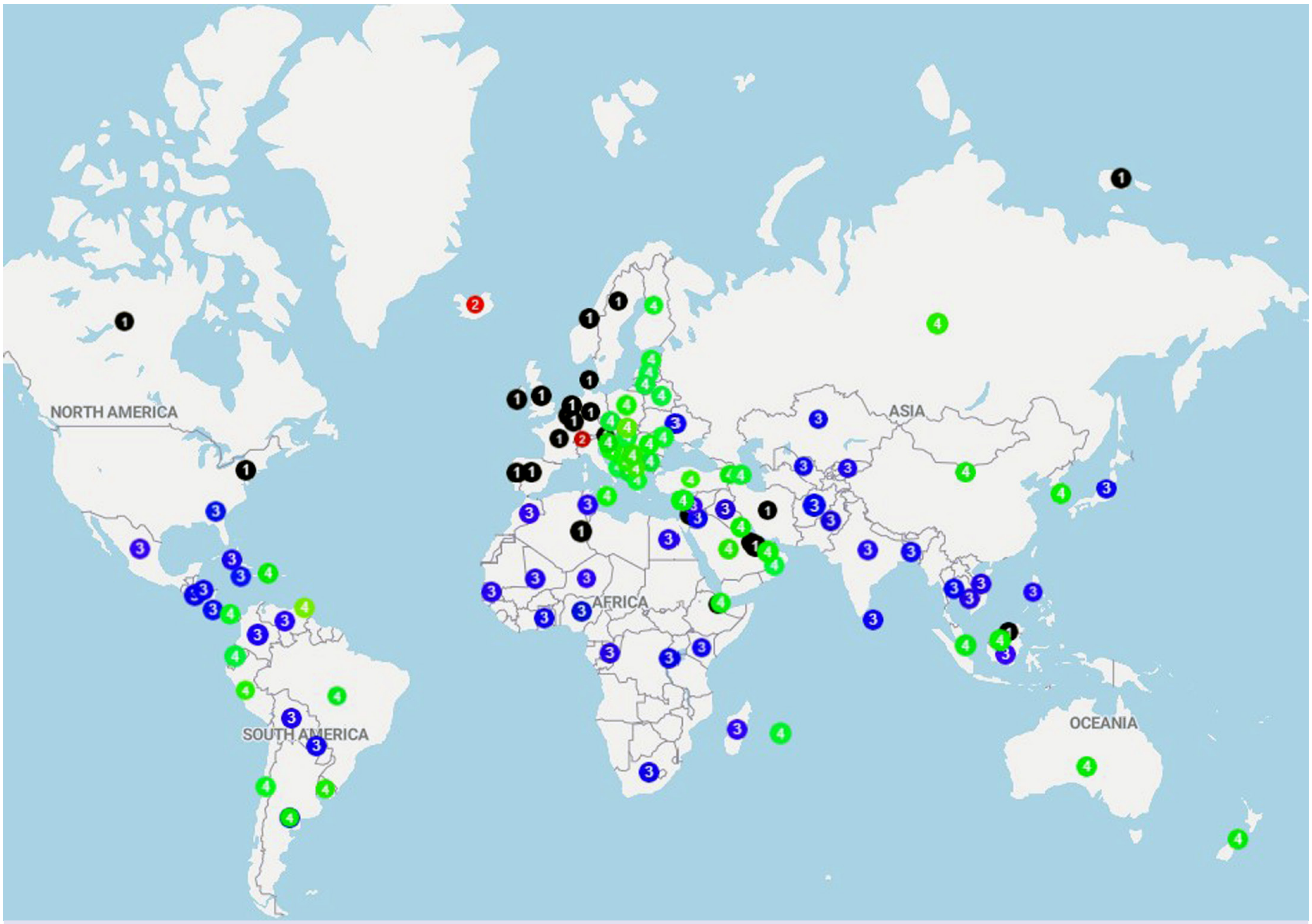

Figure 3. Countries in each cluster (Black: First cluster, Red: Second cluster, Blue: Third cluster, and Green: Forth cluster).

meter (p=0.351). However, the median age in the countries in cluster number 3 was significantly lower than that in countries in other clusters (p=0.001). However, no significant difference was found among the other clusters. The % of people who lives in urban locations in countries in cluster number 3 were significantly less than that in countries in the other clusters (p=0.001). On the other side, no significant difference was found among the four clusters with regard to amount of males per female (p=0.183). When the clusters are compared in terms of death rate from lung diseases per 100 k people, it was determined that this value was significantly higher in countries in cluster number 3 (p=0.006), but there was no significant difference among the other clusters. When the clusters are evaluated in terms of average temperature in Celsius between January and April, the mean value in cluster number 3 was found to be significantly higher than that in the other clusters (p=0.001). Average percentage of humidity between January and April was found to be significantly higher in cluster number 1 than that in cluster 3 and 4 (p=0.006). In addition, when days after 1st case (s) are examined, the time to fight the outbreak of the countries in cluster 1 was significantly longer. When the clusters are compared in terms of outbreak indicators, it was determined that the total number of cases per 1 million people exhibited a significant difference among the four clusters, and this was the highest in cluster 2 and the lowest was in cluster 3. The comparison results of the clusters in terms of outbreak indicators are given in Table 4. It was found that the daily number of new cases per 1 million people was the lowest in cluster 3 and the highest in cluster 1. However, it was found to be similar in cluster 2 and 4. The total numbers of deaths and recovered cases per 1 million people were found to be the lowest in cluster 3, and followed by cluster 4, 2, and 1. It was found that the total numbers of active cases and of critical cases were the lowest in cluster 3, and followed by cluster 4. While there was no significant difference be-

TABLE 4. The comparison results of the clusters in terms of outbreak indicators

	Cluster 1 (n=21)	Cluster 2 (n=2)	Cluster 3 (n=44)	Cluster 4 (n=47)	p*
Total case in 1 M pop					0.001
Mean	1856.12[c]	4051.95[d]	36.54[a]	603.64[b]	
Median	1593.31	4051.95	31.45	353.09	
SD	1149.32	1426.88	28.20	1562.94	
New case in 1 M pop					0.001
Mean	71.17[c]	33.37[b]	1.98[a]	18.11[b]	
Median	52.57	33.37	1.51	13.59	
SD	58.37	18.18	2.14	18.13	
Total death in 1 M pop					0.001
Mean	120.03[d]	83.30[c]	1.30[a]	18.12[b]	
Median	58.74	83.30	0.76	7.08	
SD	128.43	84.65	1.45	64.16	
Total recovered in 1 M pop					0.001
Mean	403.72[c]	2467.75[d]	6.91[a]	282.65[b]	
Median	296.11	2467.75	4.40	54.14	
SD	394.81	973.48	6.33	1505.34	
Active cases in 1 M pop					0.001
Mean	1332.12[c]	1500.90[c]	28.33[a]	302.88[b]	
Median	1220.20	1500.90	18.68	274.22	
SD	939.10	538.05	25.09	219.78	
Serious/critical in 1 M pop					0.001
Mean	45.31[c]	34.02[c]	0.69[a]	6.90[b]	
Median	32.00	34.02	0.10	5.14	
SD	40.12	14.96	1.36	6.36	
(Total death in 1 M pop/total case in 1 M pop) × 100					0.002
Mean	5.7[b]	2.58[a]	3.63[a]	2.5[a]	
Median	3.56	2.58	3.07	2.12	
SD	4.67	3.00	3.09	1.97	

*: When the mean of any cluster has different letter from the other clusters, the difference between them is statistically significant; SD: Standard deviation.

tween cluster 1 and 2, it was found significantly higher than the other two clusters.

When the clusters are compared by calculating the total death rates in the total cases of the countries in the clusters appeared, it was found that this rate was significantly higher in countries in cluster 1 than that in the other clusters. Thus, this result shows that the death in cases in the cluster 1 is higher than that in the other clusters. The number of new cases is one of the most important indicators to follow the course of the outbreak. The decrease in the number of new cases and tending to zero of the course is an indicator that the outbreak is under control and that the occurrence of new cases ends due to various reasons, and therefore, it can be said that the outbreak ended. However, the course in the number of new cases may also be related to the duration of the outbreak. For this reason, the distribution according to countries of the relationship between the elapsed time after the first coronavirus case report and the number of new cases per 1 million people is summarized in Figures 4 and 5.

DISCUSSION

In the first data set used in the study, descriptive statistics of demographic information, distribution of symptoms associated with coronavirus disease, determination of important periods defining the disease duration, and the relationship between these durations and the

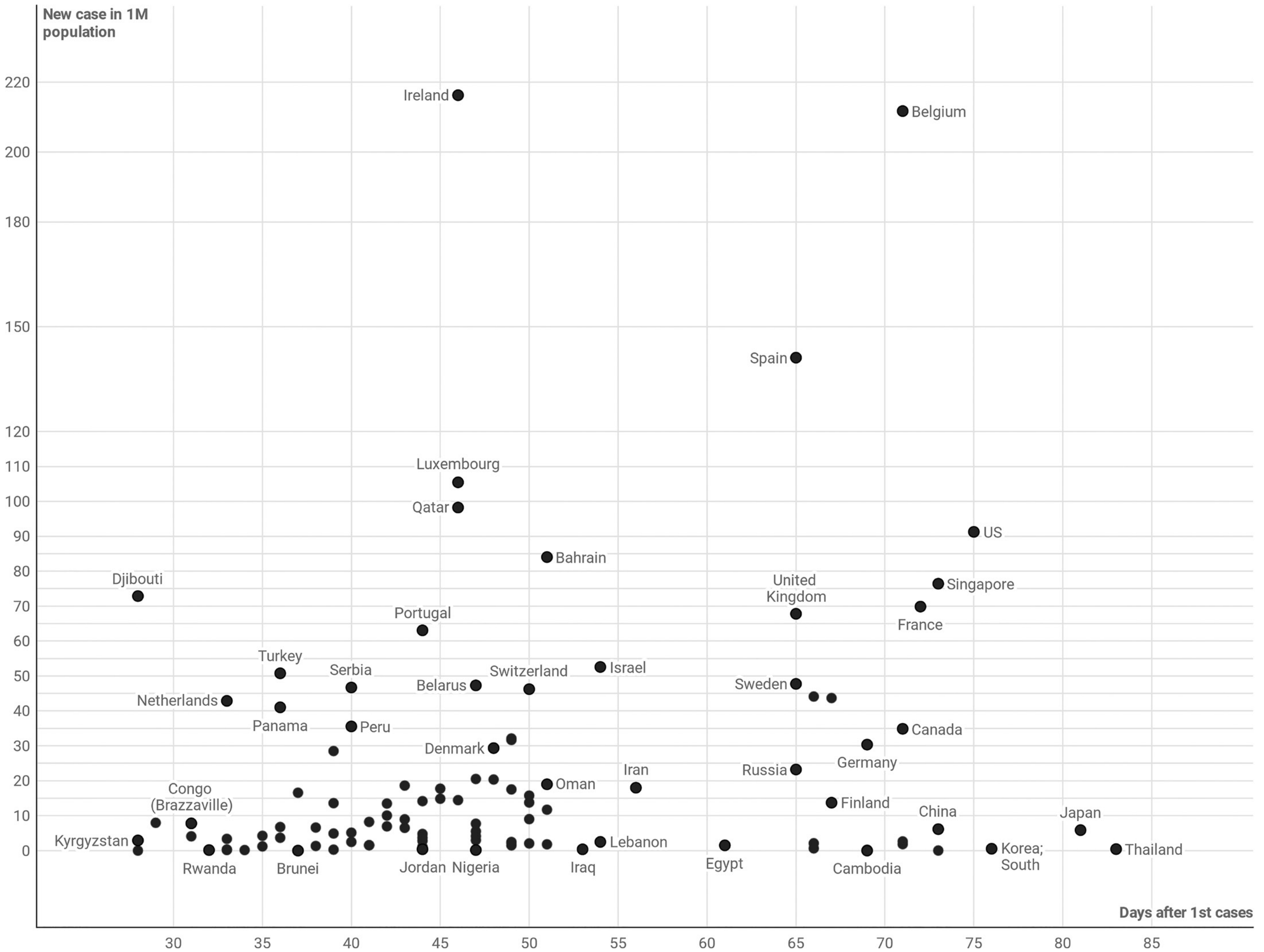

Figure 4. Distribution of countries according to the duration of the epidemic and number of new cases per 1 million people (X-axis: Days after first case(s), Y-axis: New cases in 1 M people).

outcome were investigated using case information. It was determined that approximately 65% of the cases, which were predominantly from the Far East Asian countries, were male and the average age was approximately 50. While the average value of the time from the first exposure to the onset of symptoms was found as 8 days, it was observed that the most frequently observed value of this period was 4.

Since this time course is important with regard to the infectiousness, people may be recommended to quarantine themselves for at least 4–6 days after exposure for any reason, even if a symptom does not appear. In addition, it was determined that admission to the hospital was performed within approximately 3 days following the onset of symptoms. It was calculated that an average of 18 days passed after the symptoms appeared in cases of death. The recovery period was found to be 19 days on average.

In line with this information, it is seen that the average period for the result to be clear as death or recovery is 2.5 weeks. In addition, it was determined that the risk of death increased significantly as the time from the onset of symptoms to the moment of admission to hospital prolonged. This finding shows the effect of early first intervention on the outcome. From the beginning of the pandemic until the middle of April 2020, in approximately 3/4 of the people exposed to the disease, complaints as fever came into prominence, however, in approximately half of them, complaints associated with the throat came into prominence. The rate of deaths and the mean age were found to be significantly higher in men.

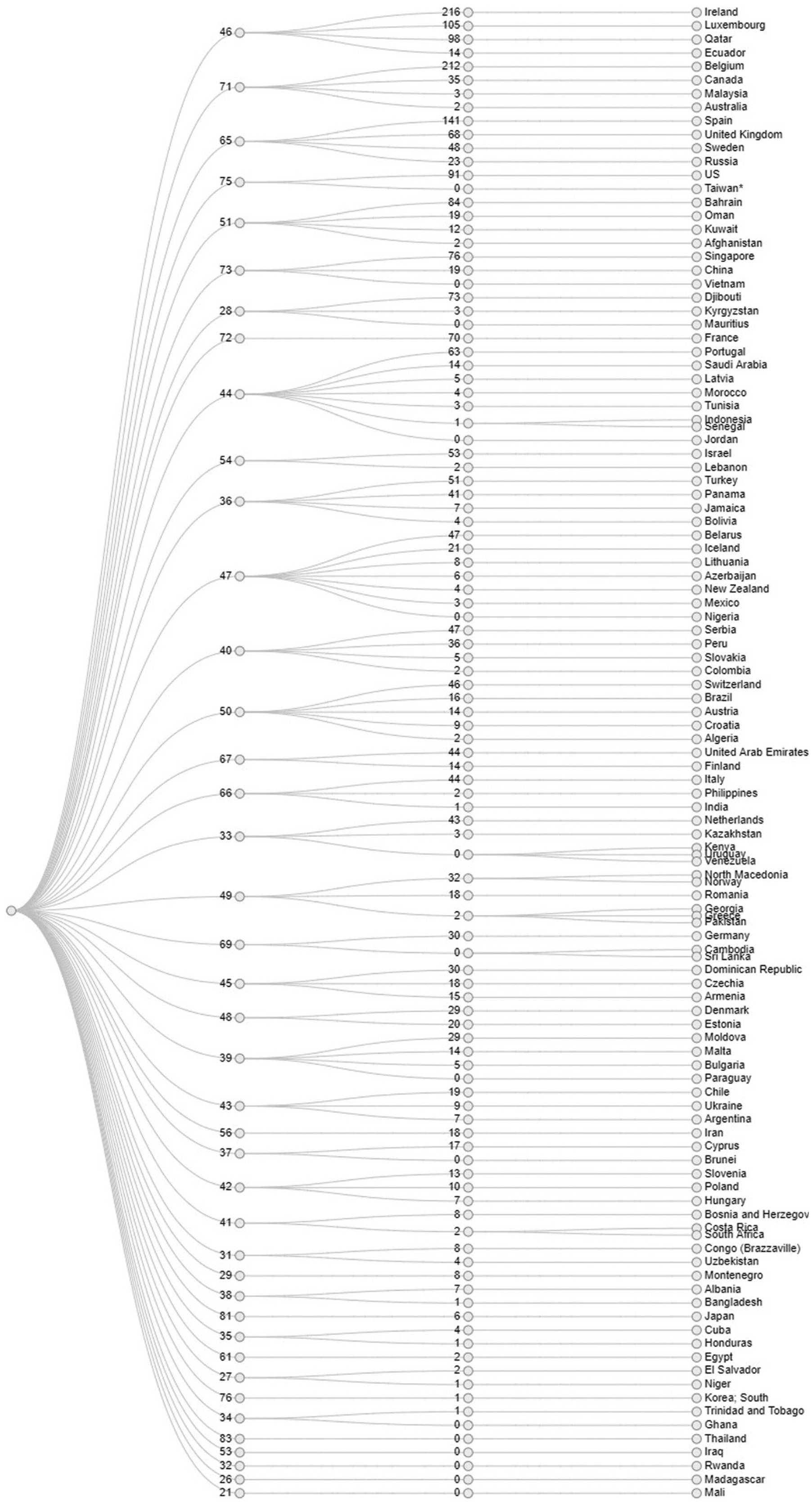

FIGURE 5. Number of new cases according to the duration of the epidemic in countries (First column: Days after first case(s), second column: New cases in 1 M people).

The negative effect of age on the death has been shown in some studies [1, 2]. However, although there are no detailed studies on the gender effects, several evaluations have shown that the death is higher in males [11]. This information obtained shows that the outbreak has spread to three quarters of the world countries and has turned into a pandemic. Therefore, this information will be useful for the measures to be taken to combat the pandemic in the mid-spring period. In particular, the numerical results defining the course of the disease as the duration are of great importance in determining the need for hospitals, beds, and necessary medical devices of countries and the measures to be taken [12]. In the second data set analyzed, 144 countries for which six outbreak indicators were defined, and where the number of cases has reached a certain level, were investigated. In addition, eight demographic characteristics of these countries, five pandemic measures, and the duration of combating the epidemic in days were evaluated. With the help of this data, it was aimed to group countries according to their similarities in terms of pandemic indicators and to examine the differences of these groups with regard to demographic characteristics, pandemic measures, and duration of combating the epidemic. Cluster analysis, which is a multivariate analysis method, was used to group the countries, and similarities of countries in terms of indicators were evaluated in this analysis. Identifying country similarities and differences are important in determining the measures to be taken. In line with this purpose, after the number of deaths, the number of recovered cases, and the number of active cases were standardized according to the population sizes of the countries, four separate clusters were obtained with similar characteristics in terms of these indicators.

The countries in the first cluster were generally the longest fighting countries against the outbreak, however, it was found that the time of struggle against the outbreak of countries in the other three clusters was similar. Since there are only two countries in the second cluster, even if the results obtained cannot be generalized to compare with the other clusters, it was determined that the measures taken against the outbreak in general were higher in the countries in clusters 1 and 2. It was observed that the rate of countries applying the quarantine policy in the first cluster was higher than the other three clusters and the majority of the countries in this cluster were European and North American countries. The result that pandemic prevention measures are more intensely implemented in the countries in the first cluster may be due to the longer time these countries struggle with the pandemic. However, the timing of the measures is also of great importance, and it reduces the spreading speed of the outbreak and allows to control the capacity of service of country [3].

The average values of the three indicators considered in countries of the first cluster were found to be higher than the other clusters when the second cluster containing only two countries was ignored. Since a significant portion of the countries in cluster number three started fighting the pandemic late, the average values of the epidemic indicators are lower than the other clusters. Moreover, the average duration of the fight against the epidemic in countries included in cluster 1 was significantly higher than the countries in other clusters. In addition, the number of critical cases was significantly higher in cluster 1 and 2. When the number of deaths per 1 million people is divided by the total number of cases per 1 million people, this rate is significantly higher in the first cluster. This result showed that the treatment measures in relevant countries in the first cluster are weaker.

In this study, it can be explained as a limitation that other possible factors such as demographic structure, epidemic measures, disease distribution, and geographical structure that may affect the epidemic course in countries are not included in the cluster analysis. In addition, it is of great importance to plan prospective cohort studies to examine the effects of the measures taken and other country characteristics on the epidemic course.

Conclusion

Delaying treatment from the moment the symptoms appear will increase the risk of death and the average time to recovery or death was 2.5 weeks. In addition, the average age and percentage of the male gender of the deceased cases are found significantly higher. It can be stated that the most important measure is to focus on methods that can detect the cases before symptoms. Otherwise, it is observed that the measures taken late will not be preventive at the desired level in the course of the outbreak.

Taken together, it is of great importance to analyze the factors together to minimize the negative impacts of the outbreak, which is an equation with multiple variables exclusive of its natural course, and to ensure control of the outbreak as soon as possible. The indicators that have a very important role in defining the pandemic are also related to each other. Therefore, multivariate methods, which take these relationships into account, are able to produce more accurate information in determining the similarities of countries.

Ethics Committee Approval: Ethics committee approval is not required as the study data were obtained from online databases.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study has received no financial support.

Authorship Contributions: Concept – HA; Design – HA, SA; Supervision – HA; Data collection and/or processing – EK; Analysis and/or interpretation – HA; Literature review – EK, SA; Writing – HA, SA, EK; Critical review – HA, SA, EK.

REFERENCES

1. Lee PI, Hu YL, Chen PY, Huang YC, Hsueh PR. Are children less susceptible to COVID-19? J Microbiol Immunol Infect 2020;53:371–2.
2. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA 2020;323:1239–42.
3. Wilder-Smith A, Freedman DO. Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. J Travel Med 2020;27:taaa020.
4. Kaggle database. Available at: https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset#COVID19_line_list_data.csv; 2020. Accessed Apr 15, 2020.
5. Worldbank database. Available at: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD; 2020. Accessed April 15, 2020.
6. Worldbank database. Available at: https://data.worldbank.org/indicator/SP.POP.TOTL.FE.ZS2020; 2020. Accessed April 15, 2020.
7. Worldometers database. Available at: https://www.worldometers.info/coronavirus/; 2020. Accessed April 15, 2020.
8. Median Age. Available at: https://worldpopulationreview.com/countries/median-age/; 2020. Accessed April 15, 2020.
9. Kaggle database. Available at: https://www.kaggle.com/koryto/countryinfo; 2020. Accessed April 15, 2020.
10. Gupta UD, Menon V, Babbar U. Detecting the number of clusters during expectation-maximization clustering using information criterion. In: 2010 Second International Conference on Machine Learning and Computing. IEEE: 2010. p. 169–73.
11. The Lancet. The gendered dimensions of COVID-19. Lancet 2020;395:1168.
12. Yonar H, Yonar A, Tekindal M, Tekindal M. Modeling and forecasting for the number of cases of the COVID-19 pandemic with the curve estimation models, the Box-Jenkins and Exponential Smoothing Methods. EJMO 2020;4:160–5.