DISTRIBUTION OF COVID-19 CASES AND DEATHS IN EUROPE DURING THE FIRST 12 PEAK WEEKS OF OUTBREAK

Hana Zach[1], Martina Hanová[1], Mária Letkovičová[2]
[1]Department of Statistics and Operations Research, Faculty of Economics and Management, Slovak University of Agriculture in Nitra, Nitra, Slovak Republic
[2]Environment a.s., Centre for Biostatistics and Environment, Nitra, Slovak Republic

SUMMARY

Objective: The aim of the study was to identify similar WHO European countries in COVID-19 incidence and mortality rate during the first 12 peak weeks of pandemic outbreak to find out whether exact coherent parts of Europe were more affected than others, and to set relationship between age and higher COVID-19 mortality rate.

Methods: COVID-19 cases and deaths from 28 February to 21 May 2020 of 37 WHO European countries were aggregated into 12 consecutive weeks. The fuzzy C-means clustering was performed to identify similar countries in COVID-19 incidence and mortality rate. Pearson product-moment correlation coefficient and log-log linear regression analyses were performed to set up relation between COVID-19 mortality rate and age. Mann-Whitney (Wilcoxon) test was used to explore differences between countries possessing higher mortality rate and age.

Results: Based on the highest value of the coefficient of overall separation five clusters of similar countries were identified for incidence rate, mortality rate and in total. Analysis according to weeks offered trends where progress of COVID-19 incidence and mortality rate was visible. Pearson coefficient (0.69) suggested moderately strong connection between mortality rate and age, Mann-Whitney (Wilcoxon) test proved statistically significant differences between countries experiencing higher mortality rate and age vs. countries having both indicators lower ($p < 0.001$). Log-log linear regression analysis defined every increase in life expectancy at birth in total by 1% meant growth in mortality rate by 22% ($p < 0.001$).

Conclusion: Spain, Belgium and Ireland, closely followed by Sweden and Great Britain were identified as the worst countries in terms of incidence and mortality rate in the monitored period. Luxembourg, Belarus and Moldova accompanied the group of the worst countries in terms of incidence rate and Italy, France and the Netherland in terms of mortality rate. Correlation analysis and the Mann-Whitney (Wilcoxon) test proved statistically significant positive relationship between mortality rate and age. Log-log linear regression analysis proved that higher age accelerated the growth of mortality rate.

Key words: COVID-19, Europe, age related deaths, fuzzy C-means clustering

Address for correspondence: H. Zach, Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 949 76 Nitra, Slovak Republic. E-mail: hana.zach@uniag.sk

https://doi.org/10.21101/cejph.a6394

INTRODUCTION

On 31st December 2019 China reported discovery of a novel coronavirus (SARS-CoV-2) infected pneumonia (COVID-19) which has caused serious illness and death (1). The World Health Organization (WHO) on 13th March 2020 declared that COVID-19 has become a global health concern where Europe has become the epicentre of the pandemic (2). The virus hit the continent harder than China (3). COVID-19 was defined as a new type of virus that spreads rapidly from person to person via droplets, contaminated hands or surfaces with incubation times of 2–14 days, had very dynamic structure and became a major epidemy that caused a great tragedy (3–5). Relevant features of the first cases in Europe of confirmed infection together with the first patient diagnosed with the disease on 24th January 2020 were discussed in Lancet Infectious Diseases (6). The virus caused fatal effects, especially in the elderly and those with chronic diseases (5, 7, 8). Given the rapid increase in COVID-19 in Europe, it was urgent to analyse the situation of the COVID-19 epidemy, trends, case increase and deaths in order to guide the implementation of prioritized prevention and control measures (9, 10). However, Europe responses to COVID-19 crisis revealed a lack of unity experiencing diverse national approaches bringing high diversity in number of cases and deaths (11). The transmission of pneumonia associated with SARS-CoV-2 has not yet been eliminated, thus, it is necessary to strengthen the monitoring of COVID-19 (5, 12).

MATERIALS AND METHODS

Data

Data were obtained from the European Union Open Data Portal crosschecked by WHO Situation Reports (13, 14). It has to be noticed that although WHO periodically updated surveillance definitions of COVID-19 cases and deaths, differences in reported data among countries were expected. The main reason was due to using different inclusion criteria and different data cut-off times, way of case detection, testing strategies or reporting practice. These factors, amongst others, influence the counts presented, with variable underestimation of true case and death counts (14). Absolute cases and deaths were aggregated into 12 consecutive weeks starting from 28 February to 21 May 2020 covering the very first peak weeks of pandemic outbreak. All data were presented in relative form per 100,000 inhabitants. The latest indicator for life expectancy at birth (LE), in total and years, was for 2018 and was used for identifying countries having older population. This indicator was obtained from the World Bank Data (15). The analyses were performed for 37 European countries, whose number of inhabitants was higher than 500,000. European islands were also excluded as these WHO EURO countries were identified as outliers possessing extreme values (incidence and mortality rate felt outside of cut-off 3 standard deviations).

Methods

Fuzzy C-means cluster analysis was used to identify similar European countries in COVID-19 incidence and mortality rate in total, and according to 12 separate weeks (16, 17). Optimum number of clusters was based on the highest value of coefficient of overall separation. Correlation (Pearson product-moment correlation coefficient) and log-log linear regression analyses were performed to set up relation between COVID-19 mortality rate (MR) and LE with 95% confidence. Due to not normal distribution of MR (tested by Kolmogorov Smirnov test) and differences in measures and scales, both indicators were transformed by logarithmic transformation (common logarithm) (18). Mann-Whitney (Wilcoxon) test was used to explore differences between countries possessing higher MR and LE vs. countries having both indicators lower. Analyses were calculated in Statgraphics Centurion version 18.1.12 and SYNTAX software.

RESULTS

Optimum number of clusters of total incidence rate (IR) and MR for 37 European countries* involved in analysis were 5, indicating 5 groups of countries possessing similar values of monitored indicators occurred during the monitored period (Fig. 1). The first cluster defined by the lowest values of monitored indicators with centroids of fuzzy clusters equal to 55.20 (total IR per 100,000 inhabitants) and 2.68 (total MR per 100,000 inhabitants) involved 18 countries predominantly situated in central and eastern Europe. The second, the most extreme cluster with centroids of 653.10 for IR, but not the worst in MR (18.15) was formed by Luxembourg. The third cluster of countries with centroids of IR 337.40 and MR 29.08 included Belarus, Italy, the Netherlands, Portugal, Sweden, Switzerland, and Great Britain. The fourth cluster of similar countries with centroids of IR 179.70 and MR 10.87 consisted of Austria, Denmark, Estonia, France, Germany, Moldova, Norway, and Serbia. The worst cluster in terms of deaths was the fifth one with centroids of IR 196.80 and MR 58.17, cluster formed Belgium, Ireland and Spain.

Figure 2 represented results of fuzzy C cluster analysis in IR for monitored countries in 12 consecutive weeks from 28 February to 21 May 2020. Also, in this case the optimum number of clusters was 5. From Figure 2 trends for 5 clusters of similar countries were visible. The first cluster of similar countries was formed

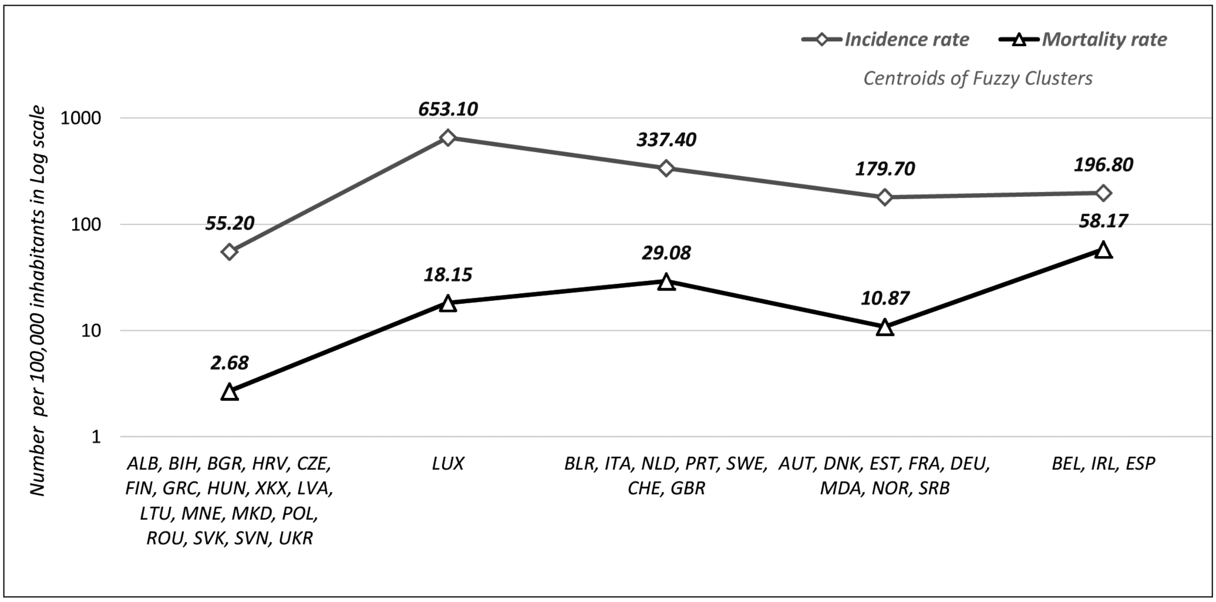

Fig. 1. *COVID-19 total incidence and mortality rate in 37 European countries, February 28 – May 21, 2020.*

*Albania, Austria, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Great Britain, Greece, Hungary, Ireland, Italy, Kosovo, Latvia, Lithuania, Luxembourg, Moldova, Montenegro, Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Serbia, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Ukraine

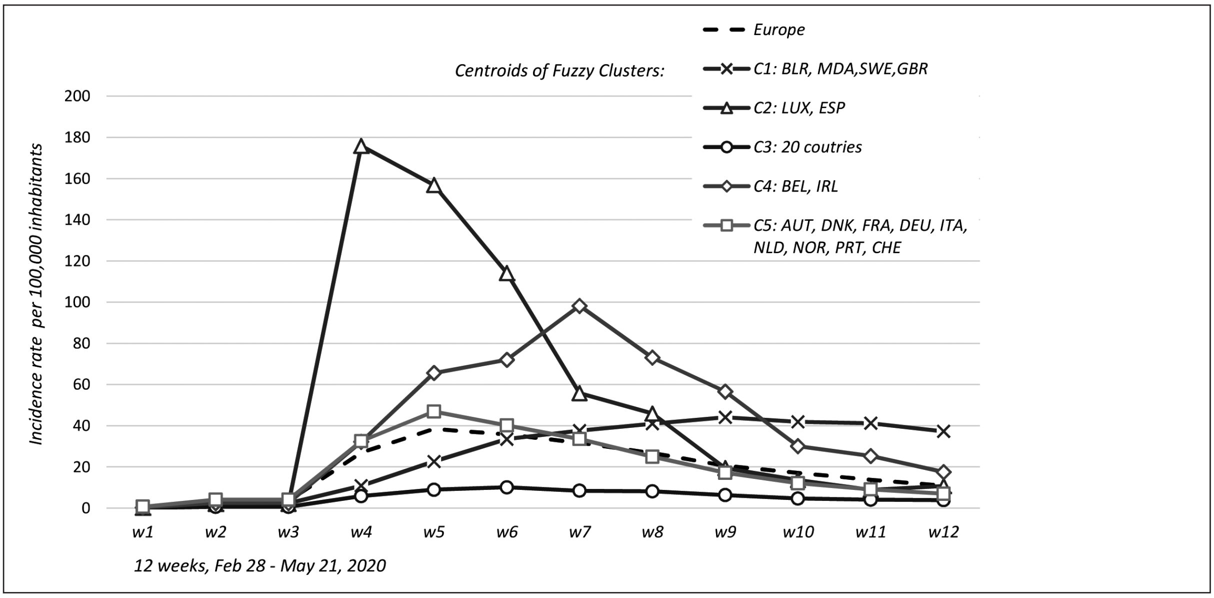

Fig. 2. *COVID-19 incidence rate per week in 37 European countries, February 28 – May 21, 2020.*

w1: Feb28–Mar05; w2: Mar06–Mar12; w3: Mar13–Mar19; w4: Mar20–Mar26; w5: Mar27–Apr02; w6: Apr03–Apr09; w7: Apr10–Apr16; w8: Apr17–Apr23; w9: Apr24–Apr30; w10: May01–May07; w11: May08–May14; w12: May15–May21

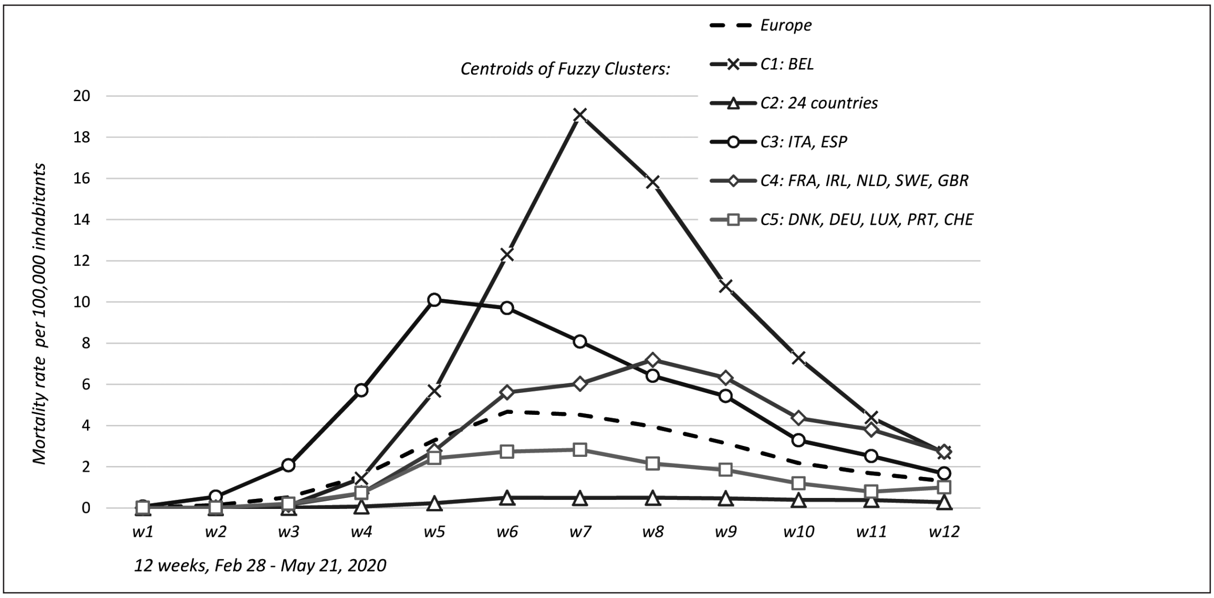

Fig. 3. *COVID-19 mortality rate per week in 37 European countries, February 28 – May 21, 2020.*

w1: Feb28–Mar05; w2: Mar06–Mar12; w3: Mar13–Mar19; w4: Mar20–Mar26; w5: Mar27–Apr02; w6: Apr03–Apr09; w7: Apr10–Apr16; w8: Apr17–Apr23; w9: Apr24–Apr30; w10: May01–May07; w11: May08–May14; w12: May15–May21

by Belarus, Moldova, Sweden, and Great Britain, the second one, with the highest cases of COVID-19 IR peaked in week 4 consisted of Luxembourg and Spain only. The third cluster, the cluster with flatter trend, consisted of 24 countries mainly situated in central and eastern Europe. The fourth cluster of similar countries in COVID-19 IR involved Belgium and Ireland. The last identified cluster closest to European trend (depicted on picture, Europe represented average for monitored countries) was comprised of Austria, Denmark, Germany, Italy, the Netherlands, Norway, Portugal, and Switzerland.

Cluster analysis for MR per 12 following weeks in monitored countries once again indicated optimum of 5 clusters (Fig. 3). Countries in those clusters were slightly different to previous analysis, however, after examining all previous calculations (Fig. 1, 2 and 3) separation was reasonable. In the first cluster Belgium only occurred having the highest MR peaking in week 7. The second, the biggest cluster, consisted of 24 countries with the lowest MR during the whole period. The third cluster of similar countries in COVID-19 MR consisted of Italy and Spain, the fourth of France, Ireland, the Netherlands, Sweden, and Great Britain. The fifth cluster had lower trend than the Europe depicted in Figure 3, and Denmark, Germany, Luxembourg, Portugal, and Switzerland created this cluster.

Kolmogorov Smirnov Test confirmed MR did not come from the normal distribution with 95% confidence ($p < 0.001$). Even though LE ranging from 71.58 to 83.55 years confirmed normality (p = 0.07), due to different measures and scales both indicators were transformed by logarithmic transformation. Pearson's correlation coefficient (r = 0.69, p < 0.001) verified positive relationship between COVID-19 MR and higher age indicating moderately strong relation. Relationship between MR and LE was modelled by log-log regression model in the form log MR = a + b*log LE, estimated as $MR = -37.71 + LE^{20.31}$, where b was the elasticity equal to 20.31 (R squared = 47.66%, standard error (SE) = 0.99, p < 0.001). Since the elasticity was greater than 1, the nonlinear effect of dependent variable MR increased with its increasing value. Exponent to LE 20.31 indicated that with increasing LE the growth of MR accelerated. A 1% increase in LE multiplied MR by $10^{20.31*\log(1.01)} = 1.22$, so a 1% increase in LE increased MR by 22%. Mann-Whitney (Wilcoxon) test proved statistically significant difference between countries experiencing higher MR and higher LE (clusters 1, 3 and 4, Figure 3, where MR was in the interval 0.51–19.13 and LE ranged from 71.58 to 83.55 years) vs. countries with both indicators lower (clusters 2 and 5, where MR was in the interval 32.37–80.11, and LE ranged from 81.36 to 83.33 years) (W = 208, p < 0.001 and W = 232, p < 0.001).

DISCUSSION

Clustering is the common way of exploring data in epidemiology, however, up till now this method was not used to identify European countries similar in COVID-19 incidence and mortality rate (16, 17). Published papers were strongly focused on clinical aspects and estimations of future numbers of cases and deaths, nonetheless, spatial distribution of pandemic in WHO European region was not described (5, 7–9). Analysing total COVID-19 incidence and mortality rate in 37 European countries in monitored period identified 5 clusters of similar countries (Fig. 1). According to fuzzy C cluster results countries in the first cluster reported the lowest level of IR (centroid 55.20 per 100,000 inhabitants) and MR (centroid 2.68 per 100,000 inhabitants), contrarily, cluster of countries possessing the highest incidence and mortality rate did not occur. Countries in this cluster recorded the best progress of COVID-19 incidence and mortality rate during the first 12 peak weeks of pandemic outbreak. Luxembourg created own cluster reaching the highest number of IR in Europe (centroid 653.10 IR), however, in terms of MR the second cluster was the third worst. Even this was the cluster with the highest IR, also countries grouped in the third cluster belonged to the most affected (centroid 337.40 IR), where MR was even higher than in case of Luxembourg. Belgium, Ireland and Spain were grouped in the last cluster although not having the highest numbers of IR, the MR was more than 21 times higher than in the best (the first) cluster.

Analysis of IR in the first 12 peak weeks allowed to compare trends in Europe (Fig. 2). The first three weeks all European countries reported low number of IR, however, after this period Luxembourg and Spain sharply peaked in week 4 reaching the highest IR in the whole monitored period (centroid 175.90), Belgium and Ireland in week 7 (centroid 98.21). Those countries recovered afterwards, leaving countries in cluster 1 (Belarus, Moldova, Sweden, and Great Britain) reporting dramatically higher IR in week 12 (centroid 37.28). In general, clusters of countries 1, 2 and 4 performed worse comparing to countries in clusters 3 and 5.

The last fuzzy C cluster analysis distinguished 5 groups of similar countries reporting comparable trends in MR in the first 12 weeks (Fig. 3). Three clusters (1, 3 and 4) observed dramatical increase in MR, pointing especially to Belgium, Italy and Spain. For example, Belgium reached the peak in week 7 (centroid 19.09), even though declined afterward, the MR remained the highest in Europe. On the other hand, two other clusters of countries (2 and 5) reported much lower MR than the European average. The lag in trends comparing to IR was evident.

Connection between higher MR and higher age in 37 European countries was firstly explored by correlation coefficient indicating countries having older population significantly experienced higher MR. Splitting countries into two groups of countries one experiencing lower MR and the second indicating higher MR also proved connection between age and deaths, when countries with higher MR were countries with significantly higher LE. Secondly, the relationship between MR and age was described by log-log regression verifying increase in LE multiplied MR. A 1% increase in LE increased MR by 22%.

CONCLUSIONS

Based on given analyses the worst countries in terms of incidence and mortality rate in total in the period from 28 February to 21 May 2020 were Spain, Belgium and Ireland, closely followed by Sweden and Great Britain. These countries struggled with IR (average for those 5 countries was 434.88 per 100,000 European inhabitants) and MR (average for those 5 countries was 52.81 per 100,000 European inhabitants) even at the end of May 2020. Luxembourg, Belarus and Moldova accompanied the group of the worst countries in terms of IR (average 381.72 per 100,000 inhabitants in 12 peak weeks) and Italy, France and the Netherlands in terms of MR (average 42.57 per 100,000 inhabitants in 12 peak weeks). All these countries displayed dramatic trends of incidence and mortality rate in monitored period that was still higher than the European average. On the other hand, Albania, Estonia, Finland, Bulgaria, Croatia, Czech Republic, Greece, Hungary, Kosovo, Latvia, Lithuania, Poland, Ukraine, Serbia, Slovakia, Slovenia, Romania, Montenegro, North Macedonia, and Bosnia and Herzegovina were countries that experienced the best results in IR (average 67.16 per 100,000 European inhabitants) and MR (average 2.99 per 100,000 European inhabitants). Worse countries, in terms of IR, MR or both, created coherent area in Europe, concluding western European countries (south west and north west, Great Britain and Ireland being north west included) and scattered Sweden, Belarus and Moldova in eastern European part were more severely affected by COVID-19 pandemic during the first 12 peak weeks of outbreak than better countries located in the central and eastern Europe. Correlation analysis proved statistically significant positive relationship between MR and LE (0.69) that was further explored by log-log regression analysis concluding every increase in LE by 1% meant growth in MR by 22%. Thus, countries having older population were more severely affected by COVID-19 mortality, significantly more than countries with lower LE. This conclusion was verified by non-parametric test (W = 208, p < 0.001; W = 232, p < 0.001) when European countries were split into two groups, one experiencing higher MR and LE vs. both indicators being lower.

Conflict of Interests

None declared

REFERENCES

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China. N Engl J Med. 2020 Feb 20;382(8):727-33.
2. World Health Organization. Coronavirus disease 2019 (COVID-19). Situation report 54 [Internet]. Geneva: WHO; 2020 March 14 [cited 2021 Jan 12]. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.
3. Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. Sci Total Environ. 2020 Aug 10;729:138817. doi: 10.1016/j.scitotenv.2020.138817.
4. Chan JF, Yuan S, Kin-Hang K, To KK, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancent. 2020 Feb;395(10223):514-23.
5. Zhai P, Ding Y, Wu X, Long J, Zhong Y, Li Y. The epidemiology, diagnosis and treatment of COVID-19. Int J Antimicrob Agents. 2020 May;55(5):105955. doi: 10.1016/j.ijantimicag.2020.105955.
6. Lescure FX, Bouadma L, Nguyen D, Parisey M, Wicky PH, Behillil S, et al. Clinical and virological data of the first cases of COVID-19 in Europe: a case series. Lancet Infect Dis Internet]. 2020 Jun;20(6):697-706. doi:10.1016/S1473-3099(20)30200-0.
7. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancent. 2020 Feb 15;395(10233):507-513.
8. Wang L, Li J, Guo S, Xie N, Yao L, Cao Y, et al. Real-time estimation and prediction of mortality caused by COVID-19 with patient information based algorithm. Sci Total Environ. 2020 Jul 20;727:138394. doi:10.1016/j.scitotenv.2020.138394.
9. Yuan J, Li M, Lv G, Lu ZK. Monitoring transmissibility and mortality of COVID-19 in Europe. Int J Infect Dis. 2020 Jun;95:311-15.
10. Karadag E. Increase in COVID-19 cases and case-fatality and case-recovery rates in Europe: A cross-temporal meta-analysis. J Med Virol. 2020 Sep;92(9):1511-7. doi:10.1002/jmv.26035.
11. Pacces AM, Weimer M. From diversity to coordination: a European approach to COVID-19. Eur J Risk Regul. 2020 Apr 16:1-14.
12. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancent. 2020 Feb 15;395(10223):470-473.
13. European Union Open Data Portal. COVID-19 coronavirus data - daily [Internet]. Solna: European Centre for Disease Prevention and Control; 2020 [cited 2021 Jan 12]. Available from: https://data.europa.eu/euodp/en/data/dataset/covid-19-coronavirus-data
14. World Health Organization. Coronavirus disease 2019 (COVID-19). Situation Report 39-122 [internet]. Geneva: WHO; 2020 March 14[cited 2021 Jan 12]. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports
15. The World Bank Data. Life expectancy at birth, total (years) - European Union [Internet]. Washington DC: The World Bank Group; 2020 [cited 2021 Jan 12]. Available from: https://data.worldbank.org/indicator/SP.DYN.LE00.IN?locations=EU
16. Wu Y, Duan H, Du S. Multiple fuzzy c-means clustering algorithm in medical diagnosis. Technol Health Care. 2015;23 Suppl 2:S519-27. Available from: https://content.iospress.com/articles/technology-and-health-care/thc989
17. Amirkhani A, Papageorgiou EI, Mohseni A, Mosavi MR. A review of fuzzy cognitive maps in medicine: taxonomy, methods, and applications. Comput Methods Programs Biomed. 2017 April;142:129-45. Available from: https://www.sciencedirect.com/science/article/pii/S0169260716307246
18. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999 Jun;8(2):135-60. Available from: https://journals.sagepub.com/doi/abs/10.1177/096228029900800204

Received July 3, 2020
Accepted in revised form January 12, 2021