Comparison of COVID-19 patterns between severely-struck Italy and the rest of the worlds: an observational study

Tung-Hui Jen¹,², Wei-Chih Kan³, Yu-Tsen Yeh⁴, Jui-Chung John Lin⁵, Tsair-Wei Chien⁶*

¹ Department of Chinese Medicine, Chi-Mei Medical Center, Taiwan
² Bachelor Program of Senior Services, Southern Taiwan University of Science and Technology,
³ Nephrology Department, Chi-Mei Medical Center, Tainan, Taiwan
⁴ Medical School, St. George’s University of London, London, United Kingdom
⁵ USA Sports Medicine, Sherman oaks, CA
⁶ Department of Occupational Medicine, Chi-Mei Medical Center, Tainan, Taiwan

Running title: The Kano model applied to interpret COVID-19 situations

Research Department, Chi-Mei Medical Center, Tainan, Taiwan

Word count: 2,835, Table: 1, Figure: 7

*Corresponding Author

Tsair-Wei Chien
Chi-Mei Medical Center, 901 Chung Hwa Road, Yung Kung Dist., Tainan 710,
Taiwan

E-mail: ophkuo@gmail.com
Yu-Tsen Yeh(YT): jess97yeh@gmail.com
Wei-Chih Kan(WC): rockiekan@ntu.edu.tw
Tsair-Wei Chien(TW): smile@mail.chimei.org.tw
Tung-Hui Jen(THJ): renin9999@gmail.com
Jui-Chung John Lin (JCJ): juichung.81@gmail.com
Abstract

**Background:** The country of Italy was placed on lockdown when the number of cases of COVID-19 increased exponentially. It is of great concern which countries/regions would have the most similar patterns to Italy and the comparison of patterns, assuming homogeneity of variance within datasets, among regions worldwide, became necessary. There were no results in past literature regarding the identification of COVID-19 patterns among countries/regions. We were therefore motivated to propose an appropriate mathematic method, using a specific country/region (e.g., Italy), for example, in detecting similar patterns at the respective peak of the outbreak. A visual display highlighting COVID-19 patterns was proposed in this study using vector mathematics.

**Methods:** We downloaded COVID-19 outbreak numbers with confirmed cases in countries/regions on a daily basis from the Github website. The top peak point was identified for each country/region. Next, thirteen-time points were assigned before and after the inflection point. COVID-19 patterns were assessed by inspecting both similarity and distance using angle (or cosine theta in trigonometric function) and Chi-square statistics. Two sets of data on confirmed cases, one based on cases per 100,000 population and the other does not, were compared using rankings as well as
four quadrants divided by similarity and distance. An app was developed to display regions with similar COVID-19 patterns to the selected country.

**Results:** The top four countries presenting with the most similar COVID-19 patterns to Italy were Switzerland, Norway, Iceland, and Luxembourg with Chi-square statistics per one freedom degree at 0.12, 0.37, 0.58, and 0.80, respectively, based on a population of 100,000. Visualizations with four quadrants and world map were shown on Google Maps. Rank correlation in rankings was -0.03 between two sets of confirmed cases with and without using the basis of cases per population of 100,000.

**Conclusion:** We proposed and demonstrated a method using both criteria of similarity and distance from vectors in mathematics to identify countries that are most similar to Italy in terms of the pattern at the peak of the outbreak. The app was developed to display countries/regions with the most similar COVID-19 patterns to the targeted country/region using a dashboard laid on Google Maps. This method can be generalized and applied to study other countries/regions for the current pandemic and other past infectious disease in history.

**Keywords:** inflection point; COVID-19; four quadrants; world map; similarity and distance; Google Maps
Background

The novel coronavirus was named COVID-19 by the World Health Organization (WHO) based on the time when the first positive case was in 2019 [1]. Although the origin of this virus was traced back to a seafood market and said to be related to bats [2], COVID-19 has spread globally and was defined as a pandemic by the WHO in March 2020 [3]. As of April 14, 2020, COVID-19 has affected 185 countries and regions around the world, with more than 1,979,477 confirmed cases and 125,910 deaths [4]. Attention was drawn to the more severely affected countries defined by case counts and comparisons were frequently made, such as US(608,377 and 25,981), Spain(174,060 and 18,255), Italy(162,488 and 21,067), Germany(132,210 and 3,495), France(131,361 and 15,748), United Kingdom(94,845 and 12,129), China(83,321 and 3,345), and Iran(74,877 and 4,683), where the numbers of infected patients and deaths are shown in brackets.

1.1 Comparisons of COVID-19 epidemics and the limitations

Comparisons of COVID-19 epidemics among countries have been discussed in some articles[5-7]. The assumption of homogeneity of variance should be preliminarily ensured for the comparison of groups [8-10]. For instance, there are discrepancies in the transmission dynamics between the city of Wuhan(China) and elsewhere in the Hubei province, where the number of confirmed cases remains
Similarly, it is unrealistic to assume that Italy would follow the pattern of that happened in Hubei, or more directly, to compare the pattern shown in Greater Wuhan, home to 19 million people, and the region of Lombardy with a population of 9 million where it was most affected across the country of Italy[7]. Observation of discrepancies and similarities in outbreak patterns among countries and regions would be beneficial in understanding COVID-19 further.

1.2 Comparisons of confirmed cases different on a population of 100,000

We identified 5,054 articles by searching the keywords “case fatality rate” on PubMed Central (PMC)[11]. The use of incidence per 100,000 population to describe a disease has been suggested in the literature[12-16]. Countries most struck by the COVID-19 based on the infectious density can be identified in this way.

1.3 What reasons for Italy devastated by COVID-19 hit so hard

The country of Italy was placed on lockdown as the number of cases of COVID-19 soared. With 21,067 deaths, Italy’s death toll is significantly higher than in other countries apart from the US(25,981) on April 14, 2020. The reasons Italy and its people were at risk for more severe infections were identified as below[17]: (1) an older population with a greater percentage of adults over the age of 65: (2) more than 21% of Italians are smokers, compared to those less than 14% in the U.S; (3) local customs such as to greet friends and loved ones with a kiss on both cheeks made it
more difficult to adjust to the social distancing regulation as an effort fighting against COVID-19; (4) the high case fatality rate (CFR) in Italy might be due to the many asymptomatic or minimally symptomatic patients were not accounted for because of the lack of testing. It would be interesting to see if the countries with characteristics similar to Italy would follow the same outbreak pattern and is a vital focus of our study.

In this study, we developed an algorithm for (1) comparing the number of infected patients at the peak of COVID-19 outbreak, (2) visualizing patterns which are most similar to Italy, and (4) generalizing the similarity approach for observations of other targeted countries/regions on the COVID-19 epidemic.

**Methods**

**2.1 Data sources**

We downloaded COVID-19 outbreak data, including information on confirmed cases in infected countries/regions from Github on April 16, 2020[18]. All downloaded data were publicly deposited on the website[18]. Ethical approval was, therefore, not necessary for this study.

**3.2 An algorithm for identifying similarity and distance in data**

We applied a trigonometric function and vector mathematics to yield the angle and distance between two datasets, using the 13 elements before and after the peak
point (i.e., the maximal daily number of confirmed cases for two compared time
points deviated from the peak point). For simplicity, we illustrated three elements for
two data strings named vectors (e.g., \{0, 1, 0\} and \{0, 0, 2\} based on the origin point
at \{1,0,0\}) as below:

\[
PQ = \begin{vmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 2 & 0 & 2 \end{vmatrix}, \quad PR = \begin{vmatrix} 0 & 0 & 2 \\ 0 & 0 & 0 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 1 & 2 \end{vmatrix}
\]

According to the vector multiplication, \(PQ \cdot PR = |PQ||PR| \cos \theta\), the angle\((=71.5^\circ)\)
can be obtained by the following formula:

\[
\cos \theta = \frac{PQ \cdot PR}{|PQ||PR|} = \frac{-1 \times (-1) + 1 \times 0 \times 2}{\sqrt{(-1)^2 + 1^2 + 0^2} \times \sqrt{(-1)^2 + 0^2 + 2^2}} = \frac{1}{\sqrt{10}} \quad \text{and} \quad \theta = \cos^{-1} \left( \frac{1}{\sqrt{10}} \right) \approx 71.5^\circ
\]

Accordingly, the angle of the direction between two vectors (e.g., representing
countries Italy and Spain in outbreak) can be referred to as the quantified feature of
similarity. The elements can be expanded to numerous numbers. For example, using a
series of daily numbers of confirmed cases was applied to this study, see Equation (1).

\[
\cos \theta = \frac{\sum_{i=1}^{n} x_i^{(P)} x_i^{(Q)}}{\sqrt{\sum_{i=1}^{n} (x_i^{(P)})^2} \times \sqrt{\sum_{i=1}^{n} (x_i^{(Q)})^2}}, (1)
\]

Where, \(x^{(P)} = \begin{bmatrix} x_1^{(P)} \\ x_2^{(P)} \\ \vdots \\ x_n^{(P)} \end{bmatrix}\), \(x^{(Q)} = \begin{bmatrix} x_1^{(Q)} \\ x_2^{(Q)} \\ \vdots \\ x_n^{(Q)} \end{bmatrix}\)

Thus, The Eq1 can be simplified as \(\cos \theta = \frac{x^{(P)\text{T}} x^{(Q)}}{\sqrt{x^{(P)\text{T}} x^{(P)}} \times \sqrt{x^{(Q)\text{T}} x^{(Q)}}}, (2)\).
Therefore, the two directions denote the similarity by multiplying elements of two vectors mentioned in Eq.(1), see Figure 1.

\[ \text{Figure 1 inserted here} \]

The distance between two vectors describe the deviation based on Euclidean Distance (e.g.,
\[ S_{pq} = \sqrt{\sum_{i=1}^{n} (x_i^{(P)} - x_i^{(Q)})^2} \]), Manhattan Distance (e.g.,
\[ S_{pq} = \sum_{i=1}^{n} |x_i^{(P)} - x_i^{(Q)}| \]), or Chebyshev Distance (e.g.,
\[ S_{pq} = \max_{i} (|x_i^{(P)} - x_i^{(Q)}|) \]).

Similarly, the Chi-Square test (e.g.,
\[ S_{pq} = \sum_{i} \left( \frac{(x_i^{(P)} - x_i^{(Q)})^2}{x_i^{(Q)}} \right) \]) can be used in this study to represent the distance. The larger deviation means a substantial difference between the two areas in the context of a pandemic. For instance, five regions shared a similar outbreak pattern with Italy at the peak period, but distinct differences were found in the number of logarithmic confirmed case counts (denoted by the mean Chi-Square statistics in parentheses), see Figure 2, based on the raw data of ln(daily confirmed cases).

\[ \text{Figure 2 inserted here} \]

3.3 Common or unique in patterns with others using four quadrants on the Kano diagram

In order to compare the difference in patterns among countries/regions, we plotted a diagram with four quadrants divided by similarities on the Y-Axis and distances on the X-axis. Countries/regions with a more similar pattern and a closer
distance related to the target country (e.g., Italy) would appear in quadrant II. Otherwise, they located in quadrants I, III, or IV. The low number of countries/regions in quadrant II means a unique pattern compared to other countries/regions.

The Kano diagram [19] contains features in the top, middle, and bottom panels for entities [20]. The results of two domains (i.e., similarity on Axis Y and distance on Axis X) appeared suitable for classifying the features along with the similarity and distance in four quadrants.

3.4 A dashboard on Google Maps to present the results

A dashboard was developed for daily display of the results across various regions. An app was developed to display those similarity and distance related to the designated country (e.g., Italy or others).

In the current study, the Google Maps platform was applied to demonstrate the Kano diagram as well as the world map to display the features that are similar to the targeted country.

Top 20 countries/regions with the most number of daily confirmed cases at the peak point were drawn using the pyramid plot.

Results

There are 15 countries/regions (colored by yellow bubbles) with similar patterns and closer distance to Italy (colored by a black bubble in quadrant I), see
Figure 3.

The top four countries presenting with the most similar COVID-19 patterns with Italy were Switzerland, Norway, Iceland, and Luxembourg with Chi-square statistics per one freedom degree at 0.12, 0.37, 0.58, and 0.80, respectively, based on a population of 100,000 see Figure 4.

When the black bubble (i.e., Italy) is clicked, all bubbles will appear on a world map, see Figure 5. It can be seen that 15 yellow bubbles locate in Europe. Readers are invited to click the bubble of interest and examine the country with the most similar patterns and shorter distances shown in quadrant II.

Correlation in rankings was -0.03 between two datasets of confirmed cases, one was based on cases per 100,000 population, and the other was not, indicating that considerations with or without the 100,000 population basis yield distinct rankings upon the comparison of countries affected by COVID-19, see Table 1.

The top 20 countries/regions with the most number of daily confirmed cases at the peak point were drawn in Figure 6. We can see that France placed in the first place,
followed by Hubei (China) and Spain. The peak date in France is on April 5, 2020, see Figure 7.

Discussions

4.1 Principle findings

The top four countries presenting with the most similar COVID-19 patterns with Italy were Switzerland, Norway, Iceland, and Luxembourg with Chi-square statistics per one freedom degree at 0.12, 0.37, 0.58, and 0.80, respectively, based on a population of 100,000 (Figure 4) which was different from the basis without considering the population of 100,000 (Figure 2). Rank correlation in rankings was -0.03 between two sets of confirmed cases with and without using a population basis of 100,000 (Table 1).

4.2 What this finding adds to what we already knew

Numerous articles have discussed using CFRs in comparison to death tolls in various diseases [11]. The incidence rates and CFRs can be adjusted by a population of 100,000 to allow fair and reasonable comparison with one another [12-16]. We analyzed data using both confirmed case based on a population of 100,000 and without, and found a distinct difference in rankings (Table 1).
Similarly, Figures 2 and 4 present differences in countries with outbreak patterns at the peak and the adjacent data points due to different definitions of confirmed cases in data. Surprisingly, all those countries have a commonly weird phenomenon that the numbers of daily confirmed cases on the date ahead of the peak point were a few; see Figures 2 and 4, which is worth studying in the future.

**4.3. What it implies and what should be changed**

We illustrated two major visual representations with four quadrants and world maps in Figures 3 and 5, using Italy as an example on Google Maps, which can be examined in detail by observing different diagrams or maps from multiple perspectives. The dashboard-type visualizations were designed and displayed on Google Maps, which can help epidemiologists to easily understand COVID-19 situations, particularly for countries with a similar pattern at the peak of the outbreak.

It is worth mention that the four countries (Switzerland, Norway, Iceland, and Luxembourg) have a similar outbreak pattern to Italy at its peak. Whether the four features that put Italy and its people at higher risk for severe disease[17] were in these countries would be interesting to study in the future.

**4.4. Strengths of this study**

Three features are highlighted below: (1) a vector mathematic method was proposed to assess the most similar COVID-19 pattern at the peak points for a
designated country, which has never been seen before in literature but of importance when observing COVID-19 situations in epidemiology; (2) two sets of confirmed cases with and without using a population of 100,000 as basis were compared in terms of rankings using quadrants divided by similarity and distance. A significant difference was found in data providing evidence with the methods yielding distinct results; (3) two types of visualizations on the world map and four quadrants were provided to readers, using an app, which is unique and innovative in comparing two countries/regions with similar COVID-19 patterns based on the assumption of homogeneity of variance being ensured.

4.5. Limitations and future studies

Our study has some limitations. First, although the data were downloaded from Github[18] on a daily basis, we cannot be guaranteed that the difference in criteria of determining confirmed cases among regions would not affect the results. For instance, the confirmed cases are solely from clinically diagnosed cases in Hubei since Feb. 14, 2020[21].

Second, although we applied both similarity and distance combined on a dashboard(Figure 3). The cutoff points for these two axes affect the consequent classification in Table 1. Attempts to replicate this study in the future should take into account the determination of cutoff-points used in Figure 3.
Third, there might be some biases when quoting data from the US states. It is because the original data were only compiled in regions based on counties (e.g., such as Spokane County, WA(US), St. Louis County, MO(US), Suffolk County, NY(US), Ulster County, NY(US), Unassigned Location, VT(US), Volusia County, FL(US), Unknown Location, MA(US), etc.). All of which should be combined into the specific state in the US before computing the CFR100 and TCC100. There have been numerous additional input based on counties in the US in the data recently downloaded from Github [32], making arrangements and analysis more difficult than usual.

Fourth, although we recommend using a population base method, the traditionally confirmed case counts cannot be neglected because the general public is more familiar with it and would understand it easily, despite the fact that infectious density would provide better interpretation of outbreak situations.

Finally, each government has its own strategy and policy against COVID19. We used Italy as an example in this study, but the difference in policies would have an impact on outbreak patterns, especially in Asia and the US. Various measures have been applied to fight against COVID-19 by limiting human-to-human transmission, including to rapidly identify, isolate, and provide optimized care for patients as well as to accelerate the development of diagnostics, therapeutics, and vaccines[22]. The
need to balance measures to minimize social disruption and economic impact should not be ignored. International, collaborative, and multisectoral approaches would be vital. More effort would be worthwhile in studying patterns of disease outbreak for the current COVID-19 and any future infectious diseases.

**Conclusion**

We proposed and demonstrated a method using both criteria of similarity and distance from vectors in mathematics to identify countries that are most similar to Italy in terms of the pattern at the peak of the outbreak. The app was developed to display countries/regions with the most similar COVID-19 patterns to the targeted country/region using a dashboard laid on Google Maps. This method can be generalized and applied to study other countries/regions for the current pandemic and other past infectious disease in history.

**List of abbreviations:**

- CC= correlation coefficient
- CFR= case fatality rate
- PMC= Pubmed Central

**Declarations**

**Ethics approval and consent to participate**

Not applicable.

All data were downloaded from the website database at Github

**Consent to publish**
Not applicable.

**Availability of data and materials**

All data used in this study are available in Additional files.

**Competing interests**

The authors declare that they have no competing interests.

**Funding**

There are no sources of funding to be declared.

**Authors' Contributions**

THJ developed the study concept and design. JCJ, YT, and WC analyzed and interpreted the data. TWC monitored the process of this study and helped in responding to the reviewers’ advice and comments. TWC drafted the manuscript, and all authors provided critical revisions for important intellectual content. The study was supervised by TWC. All authors read and approved the final manuscript.

**Acknowledgments**

We thank Enago (www.enago.tw) for the English language review of this manuscript.

All authors declare no conflicts of interest.

**References**

1. Callaway E. Time to use the p-word? Coronavirus enters dangerous new phase.

   Nature. 2020;579:12. [PubMed] [Google Scholar]
2. The Economist Tourism flows and death rates suggest covid-19 is being under-reported. 2020.
https://www.economist.com/graphic-detail/2020/03/07/tourism-flows-and-death-rates-suggest-covid-19-is-being-under-reported

3. Leung G, Wu J. Real-time nowcast and forecast on the extent of the Wuhan CoV outbreak, domestic and international spread. 2020.
https://www.med.hku.hk/f/news/3549/7418/Wuhan-coronavirus-outbreak_AN-UP_DATE_20200127.pdf

4. JHC. Coronavirus disease 2019(COVID-19) outbreak.2020/2/14 available at http://www.healthup.org.tw/kpiall/covid19dashboard.asp

5. Lei Z, Cao H, Jie Y, et al. A cross-sectional comparison of epidemiological and clinical features of patients with coronavirus disease (COVID-19) in Wuhan and outside Wuhan, China [published online ahead of print, 2020 Apr 9]. Travel Med Infect Dis. 2020;101664. doi:10.1016/j.tmaid.2020.101664

6. Porcheddu R, Serra C, Kelvin D, Kelvin N, Rubino S. Similarity in Case Fatality Rates (CFR) of COVID-19/SARS-COV-2 in Italy and China. J Infect Dev Ctries. 2020;14(2):125–128. Published 2020 Feb 29. doi:10.3855/jidc.12600
7. Remuzzi A, Remuzzi G. COVID-19 and Italy: what next?. Lancet. 2020;395(10231):1225–1228. doi:10.1016/S0140-6736(20)30627-9

8. Böing-Messing F, Mulder J. Automatic Bayes Factors for Testing Equality- and Inequality-Constrained Hypotheses on Variances. Psychometrika. 2018;83(3):586–617. doi:10.1007/s11336-018-9615-z

9. Kim YJ, Cribbie RA. ANOVA and the variance homogeneity assumption: Exploring a better gatekeeper. Br J Math Stat Psychol. 2018;71(1):1–12. doi:10.1111/bmsp.12103

10. Wang Y, Rodriguez de Gil P, Chen YH, et al. Comparing the Performance of Approaches for Testing the Homogeneity of Variance Assumption in One-Factor ANOVA Models. Educ Psychol Meas. 2017;77(2):305–329. doi:10.1177/0013164416645162

11. PMC. Over 5,054 articles were searched by the keywords case fatality rate. 2020/4/12 available at https://pubmed.ncbi.nlm.nih.gov/?term=%22case+fatality+rate%22++&sort=pubdate

12. Eton V, Schroeter A, Kelly L, Kirlew M, Tsang RSW, Ulanova M. Epidemiology of invasive pneumococcal and Haemophilus influenzae diseases in Northwestern
Ontario, Canada, 2010-2015. Int J Infect Dis. 2017;65:27–33. doi:10.1016/j.ijid.2017.09.016

13. Ciruela P, Martínez A, Izquierdo C, et al. Epidemiology of vaccine-preventable invasive diseases in Catalonia in the era of conjugate vaccines. Hum Vaccin Immunother. 2013;9(3):681–691. doi:10.4161/hv.23266

14. Dalcin D, Sieswerda L, Dubois S, Ulanova M. Epidemiology of invasive pneumococcal disease in indigenous and non-indigenous adults in northwestern Ontario, Canada, 2006-2015. BMC Infect Dis. 2018;18(1):621. Published 2018 Dec 4. doi:10.1186/s12879-018-3531-9

15. Gyenwali D, Vaidya A, Tiwari S, Khatiwada P, Lamsal DR, Giri S. Pesticide poisoning in Chitwan, Nepal: a descriptive epidemiological study. BMC Public Health. 2017;17(1):619. Published 2017 Jul 3. doi:10.1186/s12889-017-4542-y

16. Wamala JF, Malimbo M, Okot CL, et al. Epidemiological and laboratory characterization of a yellow fever outbreak in northern Uganda, October 2010-January 2011. Int J Infect Dis. 2012;16(7):e536–e542. doi:10.1016/j.ijid.2012.03.004

17. Melanie RG, Angela NB. As the number of coronavirus cases rises, why is Italy being hit so hard? 2020/4/12 available at
18. Google Sheet. Novel Coronavirus (2019-nCoV) Cases, provided by JHU CSSE.

2020/4/9 available at https://github.com/CSSEGISandData/COVID-19

19. Kano N, Seraku N, Takahashi F, Tsuji S. Attractive Quality and Must-Be Quality.

Journal of the Japanese Society for Quality Control 1984; 41: 39-48.

20. Sauerwein E, Bailom F, Matzler K, Hinterhuber HH. The kano model: how to delight your customers. Int Work Sem Prod Econ. 1996;1(4):313–327.

21. BBC .Coronavirus: Sharp increase in deaths and cases in Hubei. 2020/2/14

available at https://www.bbc.com/news/world-asia-china-51482994

22. Xiao Y, Torok ME. Taking the right measures to control COVID-19 [published online ahead of print, 2020 Mar 5]. Lancet Infect Dis. 2020;
doi:10.1016/S1473-3099(20)30152-3

**Figure legends:**

Figure 1 Computation of the angle $\theta$ in two vectors $\mathbf{PR}$ and $\mathbf{PQ}$

Figure 2 The example of similarity and deviation in outbreak

Figure 3 Which countries have an identical outbreak pattern with Italy

Figure 4 Similarity analysis on patterns on a population per 100,000

Figure 5 Countries in outbreaks similar Italy shown on the world map
Figure 6 The top 20 countries/regions with the most number of daily confirmed cases at the peak point on COVID-19

Figure 7 The daily confirmed cases in France on COVID-19

**Additional files**

Additional File 1:

Xlsx file: study dataset