pagepress

Geospatial Health 2022; volume 17(s1):1054

A spatio-temporal study of state-wide case-fatality risks during the first wave of the COVID-19 pandemic in Mexico

Ricardo Ramírez-Aldana,[1] Juan Carlos Gomez-Verjan,[1] Omar Yaxmehen Bello-Chavolla,[1] Lizbeth Naranjo[2]

[1]Directorate of Research, National Institute of Geriatrics, Mexico City; [2]Department of Mathematics, Faculty of Sciences, National Autonomous University of Mexico, Mexico City, Mexico

Abstract

A spatio-temporal analysis of the first wave of the coronavirus (COVID-19) pandemic in Mexico (April to September 2020) was performed by state. Descriptive analyses through diagrams, mapping, animations and time series representations were carried out. Greater risks were observed at certain times in specific regions. Various trends and clusters were observed and analysed by fitting linear mixed models and time series clustering. The association of co-morbidities and other variables were studied by fitting a spatial panel data linear model (SPLM). On average, the greatest risks were observed in Baja California Norte, Chiapas and Sonora, while some other densely populated states, *e.g.*, Mexico City, had lower values. The trends varied by state and a four-order polynomial, including fixed and random effects, was necessary to model them. The most common risk development was observed in states belonging to two clusters and consisted of an initial increase followed by a decrease. Some states presented cluster configurations with a retarded risk increase before the decrease, while the risk increased throughout the time of study in others. A cyclic behaviour with a second increasing trend was also observed in some states. The SPLM approach revealed a positive significant association with respect to case fatality risk between certain groups, such as males and individuals aged 50 years and more, and the prevalence of chronic kidney disease, cardiovascular disease, asthma and hypertension. The analysis may provide valuable insight into COVID-19 dynamics applicable in future outbreaks, as well as identify determinants signifying certain trends at the state level. The combination of spatial and temporal information may provide a better understanding of the fatalities due to COVID-19.

Correspondence: Lizbeth Naranjo, Department of Mathematics, Faculty of Sciences, National Autonomous University of Mexico, Circuito Exterior s/n, Ciudad Universitaria, Coyoacán, C.P. 04510, Mexico City, Mexico. E-mail: lizbethna@ciencias.unam.mx

Key words: Clustering in time series; linear mixed models; SARS-CoV-2; COVID-19; spatial panel linear models; spatio-temporal analysis; Mexico.

Acknowledgements: we are grateful to anonymous reviewers and editor who provided insightful comments and suggestions that greatly improved both the readability and the content of this article.

Funding: this work was partly supported by UNAM-PAPIIT IN118720, Mexico.

Supplementary materials: supplementary material is available at the journal's webpage in an *ad-hoc* Appendix. The reader is referred to the online supplementary material for more figures and tables. A gif with maps corresponding to weekly case-fatality risks since January 2020 is also available.

Received for publication: 18 November 2021.
Revision received: 10 March 2022.
Accepted for publication: 10 March 2022.

Geospatial Health 2022; 17(s1):1054
doi:10.4081/gh.2022.1054

Introduction

Coronavirus disease 2019 (COVID-19), which is caused by the severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2 virus in Wuhan, China, was first reported to the World Health Organization (WHO) by the Chinese Government on December 31st, 2019 (WHOa, 2020) and on 11 March 2020, WHO declared the disease a pandemic (WHOb, 2020). On that date, the number of cases of COVID-19 outside China had increased 13-fold and the number of affected countries tripled; there were more than 118,000 cases in 114 countries and 4291 people lost their lives, with more than 90% of all cases in just four countries (WHOb, 2020). Two of these, China and the Republic of Korea, had already significantly declining epidemics at this time, whereas 81 countries had not reported any cases and 57 countries had reported 10 or less (WHOb, 2020). On 19 October 2020, more than 40.6 million cases of the disease had been reported in 220 countries and territories in the world, with the United States, India, Brazil, Russia, and France having reported the highest number of infected and more than 1.1 million deaths (WHOc, 2020). The United States, Brazil, India, Mexico, and the United Kingdom had the highest number of fatalities (WHOc, 2020).

According to the Pan American Health Organization (PAHO), advanced age, obesity (BMI ≥40) and smoking as well as underlying health conditions, such as cardiovascular disease, chronic

kidney disease, chronic respiratory disease, chronic liver disease, diabetes, cancer, HIV/AIDS, active tuberculosis, chronic neurological disorders, sickle cell anaemia, hypertension have been found to be associated with fatality risk (PAHO, 2020). In Mexico, the first reported case of COVID-19 was detected on 28 February 2020 and by 30 April, the number of patients had exponentially increased, reaching a total of 17,799 confirmed cases and 1732 deceased; as of 26 June 2021, Mexico had confirmed 2,503,408 cases and 232,521 deaths due to COVID-19 (INFOBAE, 2021). Moreover, the National Institute of Statistics (INEGI) reported in 2021 that, from January to August 2020, COVID-19 was the second most common cause of death in Mexico. The country also has a large population with diseases associated with increased COVID-19 fatality risk and ranks as one of the countries with the greatest number of obese people in the world according to the latest National Health and Nutrition Survey (ENSANUT) (Barquera *et al*., 2020).

Spatial analysis describes geographic variations for diseases with respect to demographic, environmental, behavioural, socioeconomic, genetic, and infectious risk factors, and they have proven useful to the study of the spread of infectious diseases (Elliott and Wartenberg, 2004). In this sense, they could help us highlight areas and communities which could be hotspots or coldspots, pinpoint cases, measure risks and map transmission in a cost-saving way, thus improving the targeting of limited resources (Tatem, 2018). Several examples of spatial analysis have demonstrated its utility for providing mathematical models, *e.g*., for the study of the behaviour of gases to understand the COVID-19 propagation in Mexico City (Salcido, 2021. Several studies in France (Levratto *et al*., 2020) and Italy (Bourdin *et al*., 2021) have indicated that the demographic composition (Sannigrahi *et al*., 2020), *e.g*., population density, influences the rate of fatalities due to COVID-19. Additionally, a recent publication of data from twelve European countries shows that the number of medical practitioners, hospital beds and the level of social trust are correlated with low COVID-19 death rates (Amdaoud *et al*., 2021). Moreover, a spatial analysis performed in the USA demonstrate that COVID-19 could move from less vulnerable counties to more vulnerable ones and back again over time (Neelon *et al*., 2021). Even so, it has been demonstrated that demographics explain the spatial heterogeneity in COVID-19 testing and infection rates in New York City neighbourhoods (Schmitt-Grohé *et al*., 2020). In a recent published study by Argoty-Pantoja *et al*. (2021) the authors found that COVID-19 adversely affects indigenous populations, particularly patients who received initial outpatient care.

We aimed here to perform a spatial epidemiological study of case-fatality risks due to the COVID-19 in Mexico from April to September 2020 based on spatio-temporal analyses concerning risk by state in Mexico, including descriptive analyses through mapping and time series representations. Additionally, we aimed to analyse trends through time by state by using linear mixed models and time series clustering. Finally, the association of the main comorbidities with the COVID-19 death risks were studied by fitting a spatial panel linear model (SPLM), a variant of a panel data regression including spatial effects.

Materials and methods

Data sources

We used open-source data of suspected COVID-19 cases collected by the General Directorate of Epidemiology (DGE) of the Ministry of Health, Mexico [*Dirección General de Vigilancia Epidemiológica*, *Secretaría de Salud* (Ministry of Health, 2020)]. The dataset was updated by 30 September 2020 and divided into one weekly and another monthly dataset. The starting date for the databases was 1 April 2020, the date at which all states had at least one death, which was used for most analyses, except for the SPLM that referred to 15 April 2020 since that was the first date when there were positive cases in each combination of time and state. Weekly deaths were obtained since 1 January 2020 to generate an animation (see Supplement). We used the date when the patient entered the respective hospital unit and, in case of death, the date when the patient died.

For the assembly of the weekly and monthly datasets, we considered the following: only the cases with positive test results for SARS-CoV-2 were selected and the data were grouped by state and date, weekly and monthly, respectively. For each level of grouping, we obtained the number of positive cases, number of deaths, number of men, number of people by age group (groups configurated for every 10 years but, as discussed below, we finally used the ≥50 years age group), and the number of pregnant women. In addition, the number of patients in the DGE dataset with a positive diagnosis for each of the following co-morbidities were counted: asthma, cardiovascular disease, chronic kidney disease, diabetes, chronic obstructive pulmonary disease (COPD), hypertension, immunosuppression, pneumonia and obesity. As response variable, we used the case fatality risk, calculated as the number of deaths over the number of positive cases grouped by entity and date. The remaining variables were used as explanatory in the SPLM.

Analyses

Descriptive space-time analyses

We obtained the weekly and monthly case-fatality risk associated for each time and state. The monthly risk was counted from April 2020, and the weekly one from January 2020 and also from April 2020. Maps corresponding to these risk data were obtained. Additionally, we obtained a space-time plot (Hovmöller diagram), which represents the weekly risk by a colour gradient for each combination of state and time. An animation (gif) associated with maps corresponding to the weekly risk since January was obtained, and the time series associated with the weekly risk by state from April 2020 were plotted. R packages *spacetime* and *magick* were used for mapping and gif analysis.

Analyses of the trends structure by state

To determine trends for each state, linear mixed models (LMM) (Verbeke and Molenberghs, 2000) were fitted for the weekly data from 15 April 2020; thus, associating random effects with each state. First, we fitted a LMM including constant and linear terms, both for fixed and random effects. A map of linear trends by state was obtained since all random and fixed effects were significant according to a likelihood ratio test (LRT), allowing comparison with a model without random effects and t-tests, respectively. For a better understanding of the type of trend, polynomic

trends were added. From the linear trend model, random effects corresponding to polynomials of different degree were sequentially added. For instance, a second-order random effect was added to the linear model, which was significant according to an LRT and thus added to the model. This new model was compared with one adding a third-order random effect and so on until a random effect associated with a certain order showed up as not significant. According to this process, a four-order polynomial was used, including each term as random and fixed that was all significant according to an LRT (comparing with a model without random effects) and t-test, respectively. The original and predicted time series by state were plotted. R package *nlme*, *sp, ggplot2,* and *lattice* were used to the trends structure by state analysis.

Spatial panel linear models

The SPLM were fitted as described by Pebesma (2012), Millo and Piras (2012) and Wikle *et al*. (2019). A spatial weights matrix is required for this process; thus, we first obtained a squared matrix with of dimension of 32 (the number of states in Mexico) corresponding to Queen contiguity with all entries equal to zero or one; the latter indicating that two states are neighbours. Weights are calculated by integrating this matrix in a row-standardized form. SPLM variants can be fitted according to the presence or not of different terms: a space lagged effect associated with the response; a space lagged effect associated with the error term; and the presence or not of random or fixed effects suitable for panel data analyses. The risk values were used as response variables and the remaining variables as explanatory. Each observation corresponded to a time-state combination. All explanatory variables were used as percentages with respect to the number of infected people. Age was first obtained according to 10-year age groups but was finally used as the percentage of individuals aged ≥50 years, since this aggregated age group was more associated with the risk levels, and all the age-groups forming it were positively associated.

The response variable was transformed into different scales, but we finally used a logit transformation since the normality assumption was better satisfied at this scale. We added a small constant (summing around 0.5 deaths by observation), chosen for each model as the value that improved the normality assumption, since in the logit transformation logarithms are used and there are zero values in some observations. The association between explanatory variables, and between them, with the response was analysed by using Pearson, Spearman and Kendall correlations. We fitted univariate models to determine which variables were significantly associated with risk, eliminating from the multivariable model those that were not significant. We analysed the linearity assumption of each input with the response by obtaining scatter plots including a curve derived through locally estimated scatterplot smoothing (LOESS). Since we observed that few important outliers in some variables were present, we replaced them for more appropriate values. The SPLM with the variables obtained after this process was fitted and model assumptions assessed. We determined whether the space-lagged effect and random effect were significant, jointly and conditionally by using the LM-H one-sided joint test and conditional LM tests (Baltagi *et al*., 2003, 2007), and whether random or fixed effects were preferable by using a Hausman test (Hausman, 1978; Mutl and Pfaffermayr, 2011) for spatial models. R packages *splm, spdp, rgdal, ggplot2,* and *corrplot* were used for the SPLM analyses.

Time series clustering

Time series clustering, a method used to create groups or clusters of dynamic data, as described by Sardá-Espinosa (2017, 2019) was used. The members of the same cluster are similar to each other, but dissimilar from objects in a different cluster. In this kind of analysis, the measure of similarity or distance, the function of prototype or centroid extraction (a time series derived from the data representative of a cluster), the clustering algorithm and the cluster evaluation should be considered. The most common clustering procedures are hierarchical and partitional. The clustering of time-series proposed by Aghabozorgi *et al*. (2015) may be shape-based, feature-based or model-based. For the former (shape-based), it is common to utilize dynamic time warping (DTW) distance as dissimilarity measure. In this paper, we used a partitional clustering, a DTW distance, a k-medoids method (partition around medoids or PAM), which means that a time series in the data set is used as a representative of a cluster and average linkage. All the time series were standardized. To evaluate the clusters, we variated the number of clusters (3 to 9) and compared them using the silhouette as an internal evaluation measure. R package *dtwclust* was used to cluster the data.

Results

Results of descriptive space-time analyses

The gif presented on the website corresponds to the evolution through time of case-fatality risks in Mexico by state and by week from 1 January to 30 September 2020. During the early spread of COVID-19 in Mexico, high case fatality risk levels were present in the north-western states (mainly Baja California and Chihuahua) and in the centre states (Mexico, Morelos, Hidalgo and Guerrero). Moreover, on average, the highest mortality rates were observed in July, followed by June and August, with higher case fatality risks observed mainly in Baja California Norte (bordering the United States), Chiapas (bordering Guatemala) and Sinaloa followed by Campeche, Hidalgo, Mexico and Morelos (the last three in the centre around Mexico City) and other states showing lower risk, such as Baja California Sur, Tamaulipas and Yucatán.

Based on the Hovmöller diagram (Figure 1A) and the trends by state and time (Figure 1B), we observed that there were states with a greater case fatality risk at the beginning that then decreased, as in Durango and Tabasco; other states in which the peak came earlier; and a few states, such as Chiapas, where this peak was reached much later. The monthly case fatality risks by state are presented in Figure 1C.

Results of analyses of trends structure by state

Considering the case fatality risk as the response variable and time as the explanatory variable, the fixed effects for both constant and linear terms were significant (t=29.912, P<0.001; and t=–3.420, P<0.001, respectively). The corresponding random effects were jointly significant when compared with the model including only the fixed effects (LR=295.056, P<0.001). A map corresponding to this linear trend (including both fixed and random effects) is shown in Figure 2A. As the states had different case fatality risk trends, we could group them by the trend levels as follows: Aguascalientes had the highest (in yellow, around 0.0035) followed by Chiapas, Jalisco, San Luis Potosí and Tamaulipas (in

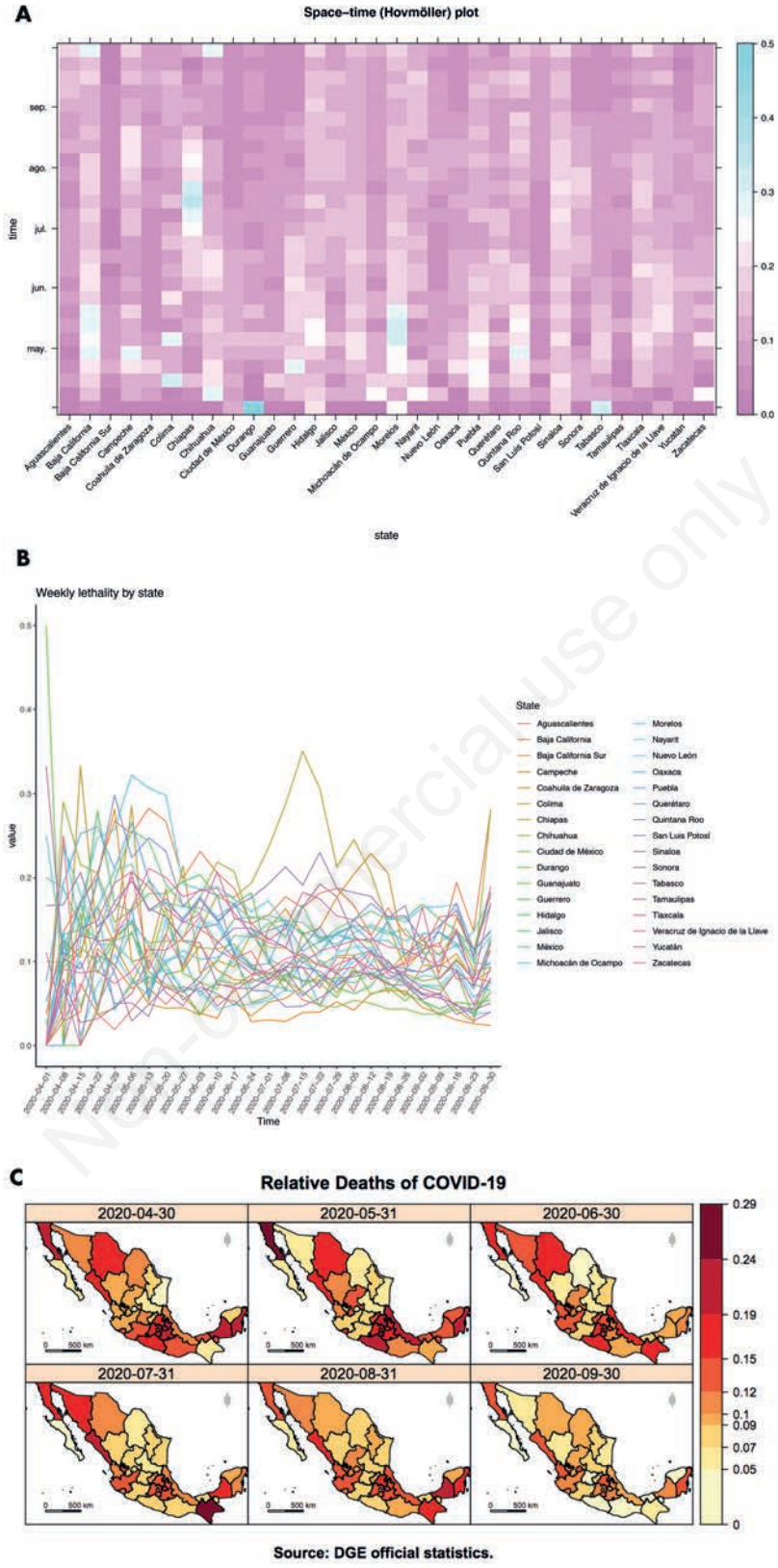

Figure 1. Representations of COVID-19 case-fatality risks in Mexico by state from 1 April to 30 September 2020. Space time (Hovmöller) plot representing weekly risks in a space-time cross section (a colour gradient is presented in which purple represents less risk and blue greater risk) (A), Weekly time series (trends) associated with each state (B) and the monthly risk (C).

orange, around 0.0015) and Campeche and Nayarit (in salmon, around 0.001); others states were around zero, for instance Coahuila de Zaragoza, Nuevo León and Querétaro (pink, around 0.0005), Sonora (magenta, around 0.0) and Baja California Sur, Tlaxcala, Veracruz de Ignacio de la Llave, Yucatán and Zacatecas (purple, around –0.0005). Finally, other states had the lowest negative trend levels, for instance, Baja California Norte, Colima, Guanajuato, Michoacán de Ocampo, Oaxaca, Sinaloa and Mexico (blue violet, around –0.0015), followed by Hidalgo, Mexico City and Quintana Roo (blue, around –0.002), Chihuahua and Durango (dark blue, around –0.0025) and Puebla (navy, around –0.0035) and finally Guerrero, Morelos and Tabasco (dark navy, around –0.0045).

We added some polynomial trends as random factors to the models including only linear effects. The quadratic trend was significant (LR=57.908, P<0.001) and thus added to the model. In the new model, the cubic trend was significant (LR=72.855, P<0.001) and thus added, resulting in another model in which the four-order trend was also significant (LR=16.044, P=0.007) and thus also added to the model. However, considering the last model, a five-order trend was not significant (LR=7.517, P=0.2756), thus, the model corresponded to a four-order polynomial. The corresponding fixed effects were all included and significant (t tests with P<0.001 for each fixed effect); in this, our final model, all random effects were jointly significant (LR=413.095, P>0.001).

The original and predicted time series by state are plotted in Figure 2B. The graphs show the estimated values, *i.e.* the marginal mean profiles, obtained from the LMM, in which a four-order polynomial with intercept, including both fixed and random effects by state, was fitted. We observed four different trend patterns. Thus, i) Aguascalientes, Jalisco and San Luis Potosí had trends that increased on average; ii) Baja California Norte, Chiapas, Chihuahua, Coahuila de Zaragoza, Morelos and Nuevo León showed a mixed patterns where the trends first increased, then decreased and, at the end of the study, increased again; iii) Baja California Sur, Campeche, Colima, Durango, Guanajuato,

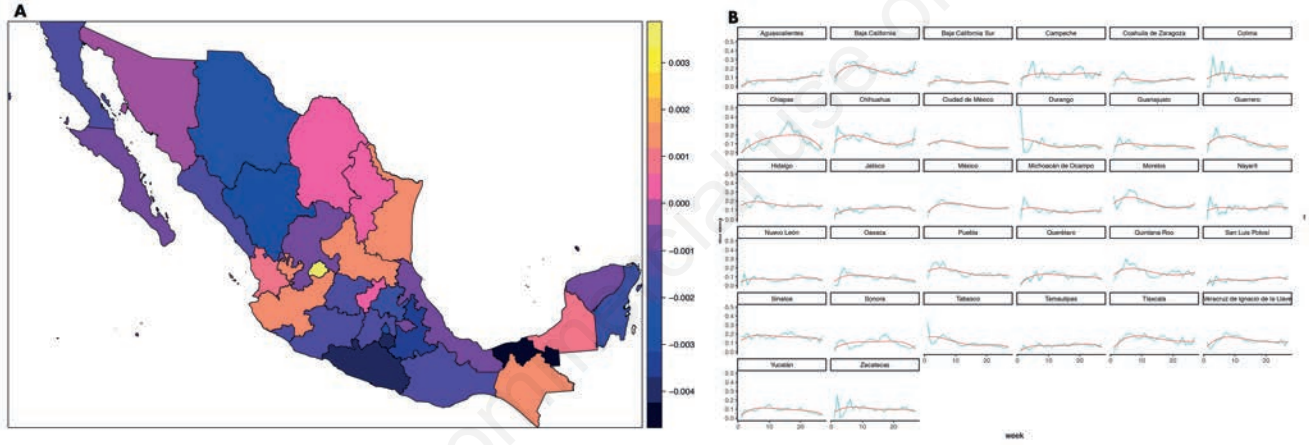

Figure 2. Representations of trends estimated through linear mixed models (LMM): A) Map of the linear trend by state obtained using an LMM with fixed and random effects for constant and linear terms; and B) Observed and predicted case fatality risks obtained from a LMM with fixed and random effects corresponding to a four-order polynomial.

Table 1. Univariate and multivariate spatial panel linear model results of fatality risk analysis.

Variable	Univariate SPLM			Multivariate SPLM		
	Estimate	Std. error	P-value	Estimate	Std. error	P-value
Intercept*	-	-	-	–4.94804	0.16184	<0.0001
Men (%)	0.01605	0.00320	<0.0001	0.01762	0.00257	<0.0001
Age group ≥50 years (%)	0.04286	0.00206	<0.0001	0.03909	0.00276	<0.0001
Asthma	0.03728	0.00798	<0.0001	0.02904	0.00966	0.00265
Cardiovascular diseases	0.02913	0.00803	0.0003	0.03105	0.01006	0.00203
Chronic kidney failure	0.09289	0.00987	<0.0001	0.06413	0.00965	<0.0001
Diabetes	0.03765	0.00313	<0.0001	–0.00113	0.00401	0.77688
COPD**	0.05130	0.01135	<0.0001	0.00552	0.01033	0.59302
Hypertension	0.03701	0.00311	<0.0001	0.00822	0.00382	0.03135
Immunosuppression	0.03688	0.01480	0.0127	0.00509	0.01375	0.71084
Pneumonia***	0.04296	0.00155	<0.0001	-	-	-
Obesity	0.01974	0.00314	<0.0001	–0.00238	0.00314	0.44835

*Estimated constant terms for the univariate models differ by model and are not presented; **Chronic obstructive pulmonary disease; ***Pneumonia was not included in the multivariate model. SPLM, spatial panel data linear model.

Hidalgo, Michoacán de Ocampo, Nayarit, Quintana Roo, Sinaloa, Sonora, Mexico, Tamaulipas and Zacatecas where little change was noted over the study period; and Guerrero, Mexico City, Oaxaca, Puebla, Querétaro, Tabasco, Tlaxcala, Veracruz de Ignacio de la Llave and Yucatán where we observed generally decreasing trends. The highest levels of the average trends of case fatality risk were in Baja California Norte, Chiapas, Morelos and Sinaloa.

Spatial panel linear models

In Figure 3, using the constant term of 0.5 in the logit transformation, we show the association (Pearson correlation) between the relative number of males, people in the age group of ≥50 years, pregnant women and the co-morbidities, whereas in Figure S1 (Appendix) we present a similar figure including the response *vis-à-vis* 10-year age groups, in which it can be seen that it was convenient to aggregate age. Similar results were obtained using other association measures. We observed high associations (>0.5) between people aged ≥50 years and diabetes, hypertension, pneumonia; the highest association between co-morbidities were diabetes with hypertension and diabetes with pneumonia. Pneumonia had the highest association with the logits; however, since an association in both directions between this variable and the response was suspected, and there was some confusion between them, we decided to eliminate it.

To fit the spatial lineal model, we first obtained the Queen contiguity matrix indicating the neighbours of each state (Table S1, Appendix). Then the corresponding weights matrix was built. The percentage of males and people aged ≥50 years showed as significant variables in the univariate spatial linear models and in terms of co-morbidities prevalence of asthma; cardiovascular diseases; diabetes; COPD; hypertension; immunosuppression; pneumonia; obesity; and chronic kidney disease (Table 1). The percentage of pregnant women was not significant and thus not included in the multivariable model. From the significant variables, we examined the linearity assumption and observed that two variables (asthma and cardiovascular diseases) had one outlier, each corresponding to earlier observations at the time when few positive cases were present; thus, we replaced them for the corresponding mean (Figure S2, Appendix).

Since the correlation matrix did not show a high correlation between the inputs (Figure 3) and thus multicollinearity was not suspected, we used all inputs in the multivariable model, except for pneumonia and pregnancy as discussed before. Hence, the multivariate spatial linear model included percentage of men and people aged ≥50 years as well as prevalence of asthma, cardiovascular diseases, chronic kidney disease, diabetes, COPD, hypertension, immunosuppression and obesity (Table 1). We determined that the random effects (μ) and spatial effects associated with the response (λ) were jointly significant (LM-H=566.71, P<0.001) and significant when conditioned to one another (LM*-μ=25.477, P<0.001 and LM*-λ = 6.471, P<0.001; for the random effects and spatially lagged effect, respectively), that random effects were necessary (Hausman Chisq=0.800, P=0.999), and that the spatial lagged effect for the error term (ρ) was also significant (t=6.187, P<0.0001). Some interaction effects were considered, but since they caused multicollinearity problems, they were omitted from the model. The residuals obtained from the LMM by using all covariates did neither show lack of normality nor lack of constant variance (Figure S3 in Appendix). Results associated with the multivariable model are given in Table 1. All variables, except diabetes, immunosuppression, COPD and obesity, were significantly associated with risk. The results show that the percentage of males and being in the ≥50 group years were positively associated with case fatality risk (OR=1.01778, P<0.001; OR=1.0399, P<0.001, respectively). The co-morbidities significantly associated with higher fatality risks were chronic kidney disease failure (OR=1.0662, P<0.001), cardiovascular disease (OR=1.0315, P=0.002), asthma (OR=1.0294, P=0.003), and hypertension (OR=1.008, P=0.030).

Time series clustering

According to the silhouette criterion, four groups of time series were necessary. The prototypes (centroids) of each cluster corresponded to the states of Baja California Sur (cluster 1), San Luis Potosí (cluster 2), Mexico City (cluster 3) and Sonora (cluster 4). The names of each state by group are displayed in Figure 4.

The groups were obtained according to the risk trends by state; we observed four available trajectories of the standardized risks along time (Figure 4). Group 1 included seven states which displayed a high increasing trend at the beginning and then a slow decreasing trend, having a constant trend on average at the middle of the time of observation. In general, these states kept their constant levels (Colima or Querétaro). Group 2 included five states with an increasing trend; they were those that maintained low risks at the beginning but started to grow with time (Aguascalientes or San Luis Potosí). Group 3 included ten states with a high fatality risk from the beginning then a slow decrease and at the end remaining more or less constant; these states had the highest fatality risks from the beginning (Baja California, Mexico City or Morelos). Finally, Group 4 included ten states with a high risk increase at the beginning, followed by a constant trend midway

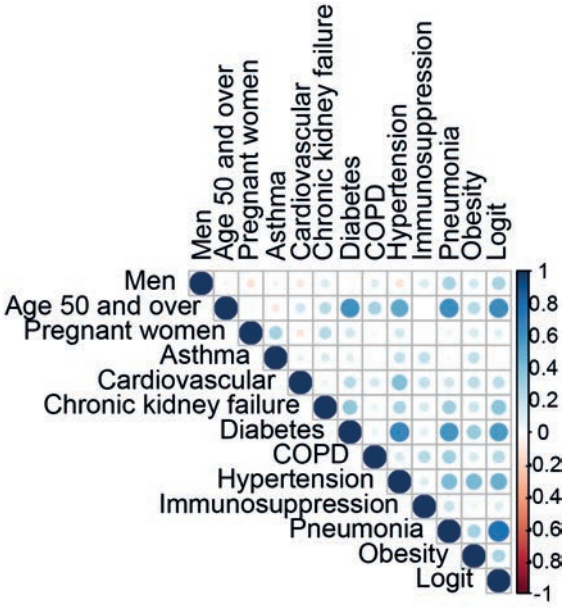

Figure 3. Association between inputs and between them and the output, the logit transformation of the case fatality risks: men (%), age group 50 years and over, pregnant women, and prevalence associated with the comorbidities.

and ending with a decreasing trend. The trends in these states were more dissimilar than in other groups, having a wider range of variation; this group included states such as Chiapas, which had a high risk from the beginning and Yucatán that had had low values all throughout the study.

Discussion

We applied different spatio-temporal analyses concerning COVID-19 case fatality risks. The descriptive analyses through maps and animations allowed us to understand how the COVID-19 case-fatality risk evolved with time in what can be considered as the first wave of COVID-19 and the beginning of the next wave in Mexico. We observed how the risk in the various states differentially increased, though this increase seemed to first be concentrated in the Northwest, the Centre and in the Yucatán Peninsula. Later on, in states such as Jalisco or Chiapas, the risk increased to rapidly reach the highest values, together with those corresponding to the same zones in which the risk increased from the beginning. Of notice is how Chiapas has some of the highest values, which may be related to the presence of indigenous population groups as some studies have indicated through survival analyses (Argoty-Pantoja *et al*., 2021). In this sense, it has been shown that the association with COVID-19 cases spatially varies according to a disadvantage measure and an indigenous composition (Huyser *et al*., 2021). Interestingly, Mexico City, despite having the greatest number of deaths and positive cases, did not reach the same high risk levels as the other states, which may be attributed to higher detection of mild and asymptomatic cases compared to the rest of the country. We also observed that the risk levels increased again at the end of the study period; again a greater increase was observed in some states in the North and in the Yucatan Peninsula, which might suggest a cyclic behaviour. However, attention should be paid with regard to the results associated with the first weeks, since few COVID-19 cases were identified at that time; and thus, the risk seemed to grow faster compared to the then future weeks. This was also observed when the spatial linear model was fitted in terms of other variables, and this was the reason why percentages associated with two variables had to be replaced by their means. Thus, we think that the risk assessments become more reliable after April or May.

The spatial linear model had a good fit and allowed us to use the longitudinal (through time) and spatial information (by state) improving the precision of the estimated associations when compared with a cross-sectional analysis. The dynamic of the risks varied with time, thus results of different cross-sectional analysis could provide different pictures. To improve the estimation process, we used a small constant term in each model, which allowed us to improve the normality assumption. Additionally, by considering that both time and space generate correlation between observations, we improved the estimations. For the multivariable model, we eliminated pneumonia as an input since it is not necessarily a precondition, as other co-morbidities, and can directly be caused by the COVID-19 infection. Thus, we think that an analysis of the direction of the association between COVID-19 and pneumonia could be relevant in future research. According to the SPLM results, the variables that were significantly, positively associated with the case fatality risk were the percentage of males, percentage of individuals in the ≥50 years age group, and prevalence of chronic kidney disease, hypertension, cardiovascular diseases and asth-

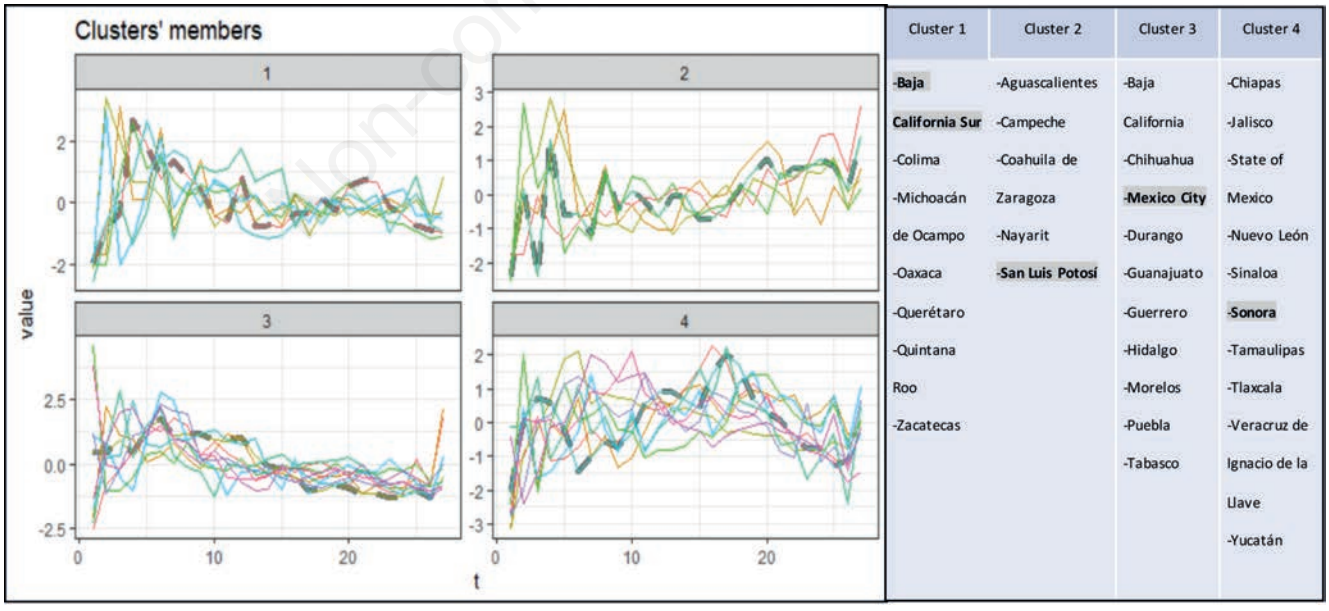

Figure 4. Groups of COVID-19 case-fatality risks trends (time series clustering) by state. They are obtained using a partitional clustering method, a dynamic time warping distance, k-medoids, and average linkage. Centroids are shown as grey dashed lines and correspond to the states of Baja California Sur (cluster 1, with seven elements), San Luis Potosí (cluster 2, with five elements), Mexico City (cluster 3, with ten elements), and Sonora (cluster 4, with ten elements). Cluster members are also shown, the centroids of each cluster are in bold.

ma. Notably, co-morbidities associated with the distribution of COVID-19 in Mexico had been identified previously. An older population structure and a high burden of cardio-metabolic co-morbidities predispose individuals to a more severe disease, thus increasing overall mortality (Bello-Chavolla *et al*., 2020, 2021). The northern states in Mexico are characterized by higher rates of chronic cardio-metabolic diseases, which may explain some of the higher case-fatality rates observed in this study. Whether the distribution of metabolic burden may be the primary influencer of case fatality or whether this could be attributable to intrinsic deficiencies in the healthcare system obstructing access and quality of medical care remains as an opportunity for future research.

When analysing clusters of the time series associated with the case fatality risks, we used both partitional and hierarchical clustering methods. We focused on the partitional method since we obtained one cluster with only one state with the hierarchical clustering approach. However, the structure of the time series pertaining to each cluster produced similar results independently of the method used. It is interesting to see the different patterns of risk developed with time. The most common pattern, as seen in Group 1 and 3, corresponded to states with an initial increase followed by a decrease of the case fatality risk with time and afterwards stabilizing into a constant trend. However, some states showed a retarded increase, while others (Group 2) generally increased with time. Probably these differences were related to mobility and how the disease propagated; perhaps much earlier in some states compared to others. In this sense, we can clearly see how the northern frontier around the Baja California Peninsula, Yucatan and the areas near Mexico City had an earlier risk increase, which remained high through all the time windows analysed. This can be explained by considering that there is a higher external and internal mobility, particularly around Mexico City, leading to higher population densities and a stronger economic activity, which may be the reason for the earlier risk increase (Ramírez-Aldana *et al*., 2021). Similarly, social vulnerability could also be behind the trend difference between the states as shown when comparing death rates in counties in the United States (Neelon *et al*., 2021) with high and low vulnerability and in marginalized municipalities within Mexico City (Antonio-Villa *et al*., 2021). The difference between death rates in Europe has also shown this differential of fatality by time and by country (Amdaoud *et al*., 2021) indicating that spatial clusters in some regions evolve with time; however, it is difficult to identify and define all the factors behind this behaviour.

The risk trends could also be examined through the fitted linear mixed models, in which we used random and fixed effects and identified that a four-order polynomial was necessary to properly model the trends. The estimated trends allowed us to identify that the most general structure is one in which the risk increases and then decreases; though, as we saw in the clusters, this trend is not present in all states. When considering only linear trends, we observed that they decreased or remained around zero in most states, which makes sense since the first wave showed a relatively rapid up and down character in most states. However, once again some states showed a generally increasing trend, possibly related to the difference in the propagation of the disease through the territory as a whole. It is important to understand the different trends by state to identify what the differences are. Perhaps, by identifying the reasons why some states had a more successful risk decrease, it might be possible to implement this successful scheme in other states and for future waves.

It is important to notice that the outcomes obtained are aggregated by state and time. In this sense, the results, particularly in terms of the spatial lineal model, cannot be inferred to an individual level (Pearce, 2000). For instance, we know that a higher percentage of individuals aged ≥50 years is associated with a greater risk, but this fact does not necessarily tell us that all individuals ≥50 years are associated with a higher risk. Thus, all conclusions in terms of the model must be taken with care. However, the aggregated analyses give us a better understanding of trends, clustering and risk association at the state level than a cross-sectional or retrospective analysis does. In terms of the factors associated with COVID-19 fatalities, both individual and aggregated analyses are relevant. From the latter, it is possible to identify policy actions that could improve specific conditions lowering the risk in groups of individuals in future COVID-19 waves and also for other infectious diseases. Additionally, we obtained associations that can be considered robust since we simultaneously considered time and different geographical units; thus, obtaining a better picture of which conditions are truly associated with the risks in the first wave of the COVID-19 pandemic. Indeed, it could be that all significant estimated coefficients have low values (or odds ratios close to one); however, in terms of deaths their effect is quite relevant. For instance, assuming a risk of 0.2, or odds of 0.25, after an increase in the percentage of males in one unit (being its estimated coefficient 0.017, considering fixed values for the other variables), the risk under the model would now have a value of 0.203, whereas the odds would be 0.254. Hence, considering for instance 1000 infected individuals per week (a common value in many states during the pandemic), the number of deaths would increase by approximately three. In addition, considering the large population size in Mexico and that the number of COVID-19 cases is with all probability much larger than the one reported, the presence of other significant variables could lead to effects that should be considered. Our analysis could be helpful in terms of policy making (Franch-Pardo *et al*., 2020; Ahasan and Hossain, 2021) since we are able to understand which zones are more vulnerable (with higher risk), what type of death curve can be expected globally and by state, and whether the death patterns will change with time. This could help us understand if some measures considered in some places could also be useful for other regions, or if similar patterns should be expected in the future. Thus, policy should perhaps vary according to the geographical conditions instead of applying the same approach all over the country. Additionally, from the linear model, we could identify which group of individuals and diseases should be looked at more closely in order to reduce mortality rates.

A spatio-temporal analysis of COVID-19 case-fatality in Mexico, as the one we present here, has not been previously performed. However, we must consider that the number of COVID-19 cases is probably larger than reported and that the number of deaths could vary as well. Thus, the risk is only indicative for the real values. In this sense, we used only the information from January to September in our analyses and not afterwards since after October the definition of the meaning of a positive case was modified. If further spatio-temporal analyses were performed, including dates after and before October, this matter should be considered, and they would most possibly lead to adjusted risk values. For future research, a similar analysis could be performed at a higher resolution, such as at the municipality level. Indeed, efforts have recently been made to provide aggregated data at this level using the same dataset we used for our analysis, but also including additional variables (Mas, 2021).

Conclusions

A heterogeneous profile of the distribution of case fatality risks across Mexico has been produced for the first wave of COVID-19 in 2020. State profiles point at spatially defined units, which may have influenced how COVID-19 mortality occurred during this first wave and may provide valuable insight into COVID-19 dynamics in future outbreaks. They may as well as identify additional determinants. By combining spatial and temporal information, a more in-depth understanding of COVID-19 case fatality may inform public policies for regional pandemic management.

References

Aghabozorgi S, Shirkhorshidi AS, Wah TY, 2015. Time-series clustering - a decade review. Inf Syst 53:16-38.

Ahasan R, Hossain MM, 2021. Leveraging GIS and spatial analysis for informed decision-making in COVID-19 pandemic. Health Policy Technol 10:7-9.

Amdaoud M, Arcuri G, Levratto N, 2021. Are regions equal in adversity? A spatial analysis of spread and dynamics of COVID-19 in Europe. Eur J Health Econ 22:629-42.

Antonio-Villa NE, Fernandez-Chirino L, Pisanty-Alatorre J, Mancilla-Galindo J, Kammar-García A, Vargas-Vázquez A, González-Díaz A, Fermín-Martínez CA, Márquez-Salinas A, Guerra EC, Bahena-López JP, Villanueva-Reza M, Márquez-Sánchez J, Jaramillo-Molina ME, Gutiérrez-Robledo LM, Bello-Chavolla OY, 2021. Comprehensive evaluation of the impact of sociodemographic inequalities on adverse outcomes and excess mortality during the coronavirus disease 2019 (COVID-19) pandemic in Mexico City. Clin Infec Dis 74:785-92.

Argoty-Pantoja AD, Robles-Rivera K, Rivera-Paredez B, Salmerón J, 2021. COVID-19 fatality in Mexico's indigenous populations. Public Health 193:69-75.

Baltagi BH, Song SH, Koh W, 2003. Testing panel data regression models with spatial error correlation. J Econom 117:123-50.

Baltagi BH, Song SH, Jung B, Koh W, 2007. Testing for serial correlation, spatial autocorrelation and random effects using panel data. J Econom 140:5-51.

Barquera S, Hernández-Barrera L, Trejo-Valdivia B, Shamah T, Campos-Nonato I, Rivera-Dommarco J, 2020. Obesity in Mexico, prevalence and trends in adults. Ensanut 2018-19. Salud Publica Mex 62:682-92.

Bello-Chavolla OY, Bahena-López JP, Antonio-Villa NE, Vargas-Vázquez A, González-Díaz A, Márquez-Salinas A, Fermín-Martínez CA, Naveja JJ, Aguilar-Salinas CA, 2020. Predicting Mortality Due to SARS-CoV-2: A Mechanistic Score Relating Obesity and Diabetes to COVID-19 Outcomes in Mexico. J Clin Endocrinol Metab 105:2752-61.

Bello-Chavolla OY, González-Díaz A, Antonio-Villa NE, Fermín-Martínez CA, Márquez-Salinas A, Vargas-Vázquez A, Bahena-López JP, García-Peña C, Aguilar-Salinas CA, Gutiérrez-Robledo LM, 2021. Unequal impact of structural health determinants and comorbidity on COVID-19 severity and lethality in older mexican adults: considerations beyond chronological aging. J Gerontol A Biol Sci Med Sci 76:e52-9.

Bourdin S, Jeanne L, Nadou F, Noiret G, 2021. Does lockdown work? A spatial analysis of the spread and concentration of COVID-19 in Italy. Regional Stud 55:1182-93.

Elliott P, Wartenberg D, 2004. Spatial epidemiology: current approaches and future challenges. Environ Health Perspect 112:998-1006.

Franch-Pardo I, Napoletano BM, Rosete-Verges F, Billa L, 2020. Spatial analysis and GIS in the study of COVID-19. A review. Sci Total Environ 739:140033.

Gobierno de la Ciudad de México, 2020. Casos a nivel nacional asociados a COVID-19 para la CDMX. Retrieved from Datos Abiertos Ciudad de México. Available from: https://datos.cdmx.gob.mx/explore/dataset/casos-asociados-a-covid-19/table/

Hausman JA, 1978. Specification tests in econometrics. Econometrica 46:1251-71.

Huyser KR, Yang TC, Yellow-Horse AJ, 2021. Indigenous Peoples, concentrated disadvantage, and income inequality in New Mexico: a ZIP code-level investigation of spatially varying associations between socioeconomic disadvantages and confirmed COVID-19 cases. J Epidemiol Community Health 75:1044-9.

INEGI, 2020. Características de las Defunciones Registradas en México durante 2019. Retrieved from Instituto Nacional de Estadística y Geografía (INEGI), Mexico. Available from: https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2020/EstSociodemo/DefuncionesRegistradas2019.pdf Accessed: October 29, 2020.

INEGI, 2021. Características de las Defunciones Registradas en México durante Enero a Agosto de 2020. Instituto Nacional de Estadística y Geogragía (INEGI), Mexico. Accessed: January 27, 2021.

INFOBAE, 2021. Coronavirus en México al 26 de junio: suman más de 2 millones y medio de contagios. Retrieved from INFOBAE. Available from: https://www.infobae.com/america/mexico/2021/06/26/coronavirus-en-mexico-al-26-de-junio-mas-de-2-millones-y-medio-de-contagios/ Accessed: June 26, 2021.

Levratto N, Amdaoud M, Arcuri G, 2020. Covid-19: analyse spatiale de l'influence des facteurs socio-économiques sur la prévalence et les conséquences de l'épidémiedans les départementsfrançais. EconomiX Working Papers, University of Paris Nanterre, EconomiX 2020-4. Available from: https://economix.fr/pdf/dt/2020/WP_EcoX_2020-4.pdf

Mas JF, 2021. Spatio-temporal dataset of COVID-19 outbreak in Mexico. Data Brief 35:106843.

Mercado CEG, Lawpoolsri S, Sudathip P, Kaewkungwal J, Khamsiriwatchara A, Pan-ngum W, Yimsamran S, Lawawirojwong S, Ho K, Ekapirat N, Maude RR, Wiladphaingern J, Carrara VI, Day NPJ, Dondorp AM, Maude RJ, 2019. Spatiotemporal epidemiology, environmental correlates, and demography of malaria in Tak Province, Thailand (2012-2015). Malar J 18:240.

Millo G, Piras G, 2012. splm: Spatial Panel Data Models in R. J Stat Soft 47:1-38.

Mizumoto K, Tariq A, Roosa K, Kong J, Yan P, Chowell G, 2019. Spatial variability in the reproduction number of Ebola virus disease, Democratic Republic of the Congo, January-September 2019. Euro Surveill 24:1900588.

Mutl J, Pfaffermayr M, 2011. The Hausman test in a Cliff and Ord panel model. Econom J 14:48-76.

Neelon B, Mutiso F, Mueller NT, Pearce JL, Benjamin-Neelon SE, 2021. Spatial and temporal trends in social vulnerability and COVID-19 incidence and death rates in the United States. PLoS One 16:1-17.

PAHO, 2020. COVID-19 and comorbidities in the Americas: Background information. Retrieved from Pan American Health Organization (PAHO). Accessed: July 29, 2020. Available from: https://www.paho.org/es/documentos/covid-19-comorbilidades-americas-antecedentes

Pearce N, 2000. The ecological fallacy strikes back. J Epidemiol Community Health 54:326-7.

Pebesma E, 2012. spacetime: Spatio-Temporal Data in R. J Stat Soft 51:1-30.

Ramírez-Aldana R, Gomez-Verjan JC, Bello-Chavolla OY, García-Peña C, 2021. Spatial epidemiological study of the distribution, clustering, and risks factors associated with early COVID-19 mortality in Mexico. PLoS One 16:e0254884.

Salcido A, 2021. A lattice gas model for infection spreading: Application to the COVID-19 pandemic in the Mexico City Metropolitan Area. Results Physics 20:103758.

Sannigrahi S, Pilla F, Basu B, Basu AS, Molter A, 2020. Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach. Sustain Cities Soc 62:102418.

Sardá-Espinosa A, 2017. Comparing time-series clustering algorithms in R using the dtwclust package. Semantic Scholar, 1-45. Retrieved from Semantic. Available from: https://www.semanticscholar.org/paper/Comparing-Time-Series-Clustering-Algorithms-in-R-Sarda-Espinosa/a46ec863bbf3e179de4e7ccedd205a96ab1ca64f

Sardá-Espinosa A, 2019. Time-series clustering in R using the dtwclust package. The R Journal 11:1-22.

Schmitt-Grohé S, Teoh K, Uribe M, 2020. Covid-19: Testing Inequality in New York City. National Bureau of Economic Research Inc., 27019.

Ministry of Health, 2020. Datos Abiertos Bases Históricas. Dirección General de Epidemiología. (Secretaría de Salud). Available from: https://www.gob.mx/salud/documentos/datos-abiertos-bases-historicas-direccion-general-de-epidemiologia

Shaweno D, Karmakar M, Alene KA, Ragonnet R, Clements ACA, Trauer JM, Denholm JT, McBryde ES, 2018. Methods used in the spatial analysis of tuberculosis epidemiology: a systematic review. BMC Med 16:193.

Tatem AJ, 2018. Innovation to impact in spatial epidemiology. BMC Med 16:209.

Verbeke G, Molenberghs G, 2000. Linear mixed models for longitudinal data. Springer, Berlin, Germany.

Wikle CK, Zammit-Mangion A, Cressie N, 2019. Spatio-temporal statistics with R. Chapman and Hall, CRC Press, Boca Raton, FL, USA.

WHO, 2020a. Coronavirus disease (COVID-19) pandemic. World Health Organization (WHO), Geneva, Switzerland. Available from: https://www.who.int/es/emergencies/diseases/novel-coronavirus-2019

WHO, 2020b. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization (WHO), Geneva, Switzerland. Available from: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—-11-march-2020 Accessed: March 11, 2020.

WHO, 2020c. COVID-19 Weekly epidemiology update. World Health Organization (WHO), Geneva, Switzerland. Available