Title

Spatial analysis of COVID-19 spread in Iran: Insights into geographical and structural transmission determinants at a province level

Authors

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

Institutions

[1]Research Division, Instituto Nacional de Geriatría, Mexico City, Mexico

[2]Department of Physiology, Facultad de Medicina, Universidad Nacional Autónoma de México, Mexico City, Mexico

***Corresponding Author**: Ramírez-Aldana R. **e-mail:** ricardoramirezaldana@gmail.com, Research Division, Instituto Nacional de Geriatría (INGER), Anillo Perif. 2767, San Jerónimo Lídice, La Magdalena Contreras, 10200, Mexico City, Mexico.

ABSTRACT

The Islamic Republic of Iran reported its first COVID-19 cases by 19th February 2020, since then it has become one of the most affected countries, with more than 73,000 cases and 4,585 deaths to this date. Spatial modeling could be used to approach an understanding of structural and sociodemographic factors that have impacted COVID-19 spread at a province-level in Iran. Therefore, in the present paper, we developed a spatial statistical approach to describe how COVID-19 cases are spatially distributed and to identify significant spatial clusters of cases and how socioeconomic and climatic features of Iranian provinces might predict the number of cases. The analyses are applied to cumulative cases of the disease from February 19th to March 18th. They correspond to obtaining maps associated with quartiles for rates of COVID-19 cases smoothed through a Bayesian technique and relative risks, the calculation of global (Moran's I) and local indicators of spatial autocorrelation (LISA), both univariate and bivariate, to derive significant clustering, and the fit of a multivariate spatial lag model considering a set of variables potentially affecting the presence of the disease. We identified a cluster of provinces with significantly higher rates of COVID-19 cases around Tehran (p-value< 0.05), indicating that the COVID-19 spread within Iran was spatially correlated. Urbanized, highly connected provinces with older population structures and higher average temperatures were the most susceptible to present a higher number of COVID-19 cases (p-value < 0.05). Interestingly, literacy is a factor that is associated with a decrease in the number of cases (p-value < 0.05), which might be directly related to health literacy and compliance with public health measures. These features indicate that social distancing, protecting older adults, and vulnerable populations, as well as promoting health literacy, might be useful to reduce SARS-CoV-2 spread in Iran. One limitation of our analysis is that the most updated information we found concerning socioeconomic and climatic features is

not for 2020, or even for a same year, so that the obtained associations should be interpreted with caution. Our approach could be applied to model COVID-19 outbreaks in other countries with similar characteristics or in case of an upturn in COVID-19 within Iran.

Keywords: COVID-19, Iran, SARS-CoV-2, spatial clusters, spatial epidemiology, spatial statistics.

Author Summary

Iran was among the first countries reporting a rapid increase in the number of COVID-19 cases. Spatial epidemiology is useful to study the spatial distribution of a disease and to identify factors associated with the number of cases of such disease. By applying these methods, we aimed to identify whether there are clusters of regions in Iran with high or low number of COVID-19 cases and the association of different factors with these numbers, considering spatial relationships and maps representing these associations. Interestingly, we found regions of high number of cases and that more COVID-19 cases were present in provinces with more urbanization, aging population, number of physicians, efficient communications, and greater average temperatures, whereas less COVID-19 cases were present in provinces with more literacy. This study allowed us to understand the spatial behavior of the disease and the importance of having adequate health policies, literacy campaigns, and disseminating health information to the population.

INTRODUCTION

On 11th March 2020, the General Director of the World Health Organization (WHO), Dr. Tedros Adhanom Ghebreyesus, declared the new infectious respiratory disease COVID-19, caused by the infection of novel coronavirus SARS-CoV-2 as a pandemic, due to the rate of growth of new cases, the number of affected people, and the number of deaths (1). As of the time of this writing (April 15th, 2020), the number of infected cases world-wide corresponded to more than 1 million, being the most affected countries: Italy (16,523 deaths), Spain (13,341 deaths), USA (10,792 deaths), France (8,911 deaths), United Kingdom (5,373 deaths), and Iran (3,739 deaths)(2,3).

Iran was among the first countries outside of China to report a rapid increase in the number of COVID-19 cases and associated deaths; its first confirmed cases were reported on 19th February 2020 in the province of Qom imported from Wuhan, China(4). Nevertheless, some reports suggest that the outbreak may have happened two or six weeks before the government official announcement (5). Iran had one of the highest COVID-19 mortality rates early in the pandemic, and its rate of spread has been amongst the highest. However, as with other countries, it may be a sub-estimation of cases, and there may be other cases not officially reported (6).

The large count of COVID-19 cases and mortality in Iran are multifactorial. Iran's response to the epidemic has been highly affected by several imposed economic sanctions and armed conflicts within the last 20 years. Moreover, its difficult economic situation due to a recession, having inflation rates that are among the highest in the region, has taken a toll on its public health system (7,8). Although there are approximately 184,000 hospitals and primary health-care staff, limitations in the availability of COVID-19 testing kits, protective equipment, and ventilators are quite important. On the other hand, over the last years, Iran has slowed the rate of mortality associated with infectious and maternal diseases. It is currently undergoing an epidemiological transition where infectious diseases interact with

chronic conditions (2). In this sense, Iran may represent other similar developing world countries with poor health systems and an increased prevalence of chronic diseases.

Spatial statistics have emerged as a useful tool for the analysis of spatial epidemiology, concerning mapping and statistical analyses of spatial and spatial-temporal incidences of different pathogens. The aim of this paper is to perform spatial analyses which allow us to better understand the COVID-19 outbreak in Iran, not only in terms of the strength of its presence and socioeconomic and structural factors which facilitate the disease spread within Iranian provinces, but also in terms of how the disease is spatially distributed and which variables are spatially related with it considering the spatial effect to obtain adequate inferences. Given the role of climate and socio-economic factors in determining the distribution of cases and its impact world-wide, we also aimed to incorporate said factors as predictors of SARS-CoV-2 spread (9,10).This could aid to understand the burden of COVID-19, its distribution in the country, and its implication on public health within Iran and similar countries (11) and could contribute to public health measures by providing insight to inform the implementation of interventions or to understand socio-demographic factors associated with the SARS-CoV-2 spread and COVID-19 heterogeneity as it has been applied to previous infectious diseases (12–16).

MATERIAL AND METHODS

Data sources

We obtained province-specific data considering 31 provinces or polygons in Iran **(Table 1)**. From the Statistical Centre of Iran (17), we extracted information concerning: 1) people settled in urban areas in 2016 (%), calculated from the population and household of Iran by province and sub-province information of the census, 2) people aged ≥60 years in 2016 calculated from the population disaggregated by age groups, sex, and province information of the census, 3) population density (people per km^2) in 2016, 4) literacy rate of population aged ≥6 years in 2016, obtained from the document of selected results from

the 2016 census, 5) the Consumer Price Index percent changes on March 2020 for the national households in contrast to the corresponding month of the previous year (point-to-point inflation), and 6) the average temperature (°C) of provincial capitals and 7) annual precipitation levels (mm) in 2015, both part of the climate and environment information. From the Iran data portal (18), we obtained, from the health section: 1) the number of physicians employed by the ministry of health and medical education in 2006 and 2) the number of beds in operating medical establishments in 2006 and from the government finance section: 3) the province contribution to gross domestic product (GDP) in 2004. We used a Transportation Efficiency Index (TEI)(19), constructed through Data Evelopment Analysis, being an indicator of the extent in which each province efficiently utilize their transportation infrastructure. The TEI has values between zero and one, values near to one indicate provinces better communicated, but we standardized it (values of each province minus its mean divided by the associated standard deviation) to allow interpretations in a better scale in terms of how the increase in a certain number of standard deviations of the TEI is associated with the number of COVID-19 cases. The cumulative number of cases with confirmed COVID-19 by Province from February 19th to March 18th, 2020, was also obtained (20). It is important to notice that, in order to obtain more accurate rates of cases with COVID-19, population size in 2020 by province was derived by using mathematical projection methods (arithmetic, geometric, exponential, and logistic or saturation methods) using information contained in the population and housing censuses from 2006, 2011, and 2016. Since all methods provided similar results, we show here only those associated with the arithmetic method. Shapefiles were obtained from the Stanford Libraries Earthworks: https://earthworks.stanford.edu/catalog/stanford-dv126wm3595, in which files are freely available for academic use and other non-commercial use (21).

COVID-19 rate estimation by Iranian provinces

We obtained quantile maps associated with raw rates of COVID-19 cases, as well as smoothed case rates by province using an empirical Bayes estimator, which is a biased estimator that improves variance instability proper of rates estimated in small-sized spatial units (22) (i.e. provinces with a larger population size have lower variance than provinces with a smaller population size). Since raw and smoothed rates were surprisingly similar, only results of smoothed rates are reported. We also obtained maps concerning excess or relative risk, serving as a comparison of the observed number of cases by the province to a national standard. For the variable concerning the number of people aged ≥60 years by province, raw, and smoothed rates were obtained using empirical Bayes, and the latter were used in all analyses.

Spatial weight estimation and spatial autocorrelation

Since all spatial analyses require spatial weights, we obtained queen contiguity weights (23). Provinces were considered as neighbors when they share at least a point or vertex in common, obtaining a squared matrix of dimension 31 (31x31 matrix) with all entries equal to zero or one, the latter value indicating that two provinces are neighbors. From these neighbors, weights are calculated by integrating a matrix in a row-standardized form, i.e., equal weights for each neighbor and summing one for each row. Moran's I statistic was obtained as an indicator of global spatial autocorrelation (24), and its significance was assessed through a random permutation inference technique based on randomly permuting the observed values over the spatial units (25). Local indicators of spatial autocorrelation (LISA) were obtained, being these a decomposition of Moran's I used to identify the contribution of each province in the statistic (26). LISA was used to derive significant spatial clustering through four cluster types: High-High, Low-Low, High-Low, and Low-High. For instance, the High-High cluster indicates provinces with high values of a variable that are significantly surrounded by regions with similarly high values. Analogous to Moran's I and LISA, estimates can be calculated to identify the spatial

correlation between two variables and to identify bivariate clustering (27). For instance, to identify provinces with high values in a first variable surrounded by provinces with high values for a second variable (cluster High-High). Bivariate clustering and quartile maps were obtained for each of the significant variables in a linear spatial model, to have a better understanding of the individual spatial effect of each of these variables over the smoothed rates associated with the disease.

Spatial multivariate linear models

Spatial multivariate linear models were fitted to identify variables that significantly impact the number of log-transformed COVID-19 cases (28). This response variable was chosen since the corresponding model better satisfies all statistical assumptions, the other variables introduced in the Data section were simultaneously introduced as explanatory, first removing from the model all variables generating multicollinearity. Ordinary Least Squares (OLS) estimation was used to identify whether a linear spatial model was necessary by using a Lagrange Multiplier (LM) and a robust LM statistics to compare the non-spatial model with spatial models (29). Two kinds of spatial models were compared; the spatial-lag model considers the spatially lagged response as an additional explanatory variable, whereas the spatial-error model considers that the error is a linear function of a spatially lagged error plus another error term. Another model was obtained by performing a backward selection process, considering the elimination of the most non-significant variable in each step and the minimization of the Akaike Information Criterion (AIC). This process allowed us to identify whether the associations obtained through this model were similar as those obtained through the model including all variables. For significant variables in the linear spatial models, interpretations in the original scale (i.e., as counts) were derived and we performed bivariate LISA significant clustering between each of

these significant variables and the rate of cases with COVID-19, as explained above. All statistical analyses were conducted using GeoDa version 1.14.0. A two-tailed p-value<0.05 was considered as the significance threshold.

RESULTS

Rates description and spatial autocorrelation of COVID-19 case rates between provinces

Maps for quartiles corresponding to the smoothed rates and excess risk of COVID-19 cases are shown in **Figure 1**. We observed that the highest rates of COVID-19 and excess risk values were located in the Northern region of Iran corresponding to the provinces of Qom, Marzaki, Mazandaran, and Semnan. There were also high rates associated with the provinces of Alborz, Gilan, Qazvin, and Yazd (last quartile). We observed significant spatial autocorrelation (Moran's I=0.426, p=0.002), indicating that COVID-19 rates between provinces are significantly spatially related. From the heat and significance maps corresponding to significant clusters using an empirical Bayes spatial technique, we delimited a High-High cluster in red, indicating a northern zone around Tehran and Alborz with significant high COVID-19 rates surrounded by areas with similarly high rates. Conversely, we delimited a Low-Low cluster in blue indicating southern provinces with small rates surrounded by areas with similarly lower rates, which includes the provinces of Bushehr, Homozgan, Sistan, and Baluschestan. Interestingly, Golestan showed in light purple, has significantly lower COVID-19 rates despite being surrounded by a cluster of provinces with higher rates (**Figure 2**).

Selection of multivariate linear spatial model for COVID-19 spread

Since the variable hospital beds is strongly associated with variables GDP and number of physicians (Kendall correlation coefficients above 0.55), we eliminated it from all models to avoid multicollinearity (Supplementary material). We confirmed that a spatial model was necessary since the error term from the OLS fitting showed significant spatial autocorrelation (Moran's I= 0.134, p-value=0.025). Additionally, the LM and Robust LM

statistics indicated that a spatial lag model was required since the spatial parameter (ρ) associated with the spatially lagged response was significant (LM=10.669, p-value=0.001; Robust LM=13.557, p-value < 0.001), which did not occur with the spatial error model since the corresponding spatial parameter was not consistently significant (LM=1.256, p-value=0.262; Robust LM=4.144, p-value=0.042). Thus, only the spatial lag model was fitted obtaining a significant spatial parameter (ρ = 0.723, p-value<0.001), which indicated that the rate of an area in the linear model is affected by COVID-19 rates in neighboring areas (R^2=0.877, σ^2=0.146). Normality and homoscedasticity assumptions were reasonably satisfied.

Predictors of COVID-19 spatial spread in Iranian provinces

The significant variables associated with the model obtained through the selection scheme were the same as those associated with the model including all variables. This simplified model excluded population density, Consumer Price Index, and annual precipitation. The estimated coefficients were similar for both models; however, we analyzed the estimations associated with the model including all variables to consider effects controlled for these three variables. Hence, the variables that significantly impact the log-transformed number of COVID-19 cases include: the percentage of people settled in urban areas (p-value = 0.019), smoothed rate of people aged ≥60 years (p-value < 0.001), literacy rate (p-value = 0.006), average temperature (p-value < 0.001), number of physicians employed (p-value < 0.001), and the TEI (p-value = 0.035) (**Table 2**). A 10% increase in urban population or a 1% increase in the population aged ≥60 years has a percentage increase of 29.29% (95%CI 26.55 – 32.10%) and 46.65% (95%CI 26.54% -69.95%), respectively, on the number of COVID-19 cases. Moreover, an increase of 1°C in the temperature levels, an increase of 1 physician, or an increase of one deviation over the standardized TEI also have a percentage increase of 11.98% (95%CI 5.54-18.80%), 0.08% (95%CI 0.06 - 0.11%), and 16.98% (95%CI 1.14-35.30%), respectively, over the number of COVID-19

cases. Finally, a 1% increase in the literacy rate showed a percentage decrease of 10.44% (95%CI 3.18-17.16%) on the number of cases.

Spatial lag predictors and province clusters

Quartile maps for each of the significant variables in the spatial lag model are shown in **Figure 3.** Finally, concerning bivariate LISA significant clustering, we observed a positive spatial relationship (Moran's I=0.341, p-value=0.002) between urban population and COVID-19 rates; provinces with high urban rates surrounded by areas with high COVID-19 rates are the same as the ones in the High-High cluster for COVID-19, except for Mazandaran, and similarly for the Low-Low cluster, except for Bushehr. There is also a positive spatial relationship (Moran's I=0.279, p-value=0.002) between the population aged ≥60 and COVID-19 rates. Both High-High and Low-Low clusters include similar provinces as the ones in the clusters for COVID-19, except for Qom and Alborz, which have significantly lower rates of people aged ≥60 years but are spatially surrounded by areas with high disease rates. Concerning literacy rates, we also identified a positive spatial relationship between literacy and disease rates (Moran's I=0.362, p-value=0.005). The associated High-High cluster and that obtained for COVID-19 rates are formed by the same provinces, whereas in the south, Hormozgan and Bushehr have high literacy rates but are surrounded by areas with low disease rates.

Concerning average temperature levels, the global spatial autocorrelation is negative (Moran's I=-0.107, p-value=0.103). The High-High clusters for temperature and COVID-19 rates are similar, except for Marzaki and Alborz, where there is significantly lower temperature surrounded by areas with high COVID-19 rates. In the south, there is a significantly high temperature with spatially lower disease rates.

There is a positive spatial relationship (Moran's I=0.302, p-value=0.003) between the number of physicians and the COVID-19 rate. There is a High-High cluster in the north with a High-Low zone between formed by Marzaki, Qom, and Semnan, with a significantly

lower number of physicians; however, they are spatially surrounded by areas with higher disease rates. Concerning the TEI, the spatial correlation is close to zero (Moran's I = -0.096, p-value=0.112), indicating a particular random global spatial relationship between TEI and the disease rate. The High-High cluster is the same as the High-High cluster for the disease, except for Mazandaran and Alborz, which have significantly low TEI but are surrounded by areas with high disease rates. In the south, two provinces, which formed a Low-Low cluster for COVID-19 cases, are now areas with high TEI spatially associated with areas with low disease rates.

DISCUSSION

Here, we demonstrate that the rates of COVID-19 cases within Iranian provinces are spatially correlated. This could be due to the origin of the outbreak, which started on the north of Iran, and can be seen through an important province cluster with the highest number of COVID-19 cases that we found around Tehran and Qom. Several mathematical models have been used to model the COVID-19 outbreak, mostly focused on forecasting the number of cases and assessing the capacity of country-level healthcare systems to manage disease burden (30–32). In the present report, we demonstrate that the spatial relationship and socio-demographic factors associated with the provinces must be considered to model the disease adequately, and this report also highlights structural factors that may lead to inequities in COVID-19 spread. Of relevance, we highlight the role of social determinants of health in sustaining SARS-CoV-2 transmission and provide additional evidence that human mobility or province interconnectedness might be associated in favoring disease spread (33).

Importantly, our approach demonstrates that urbanization, aging population, education, average temperatures, number of physicians, and inter-province communications are associated with the case numbers amongst Iranian provinces. The obtained results do not consider the spatial effect, which is accumulated since the spatially lagged response is

part of the explanatory variables, and they consider fixed values for all variables except the one being interpreted. Overall, these variables spatially correlate with the COVID-19 province clustering indicating a consistent association with the observed variables. The greatest increase in the number of COVID-19 cases is associated with people aged ≥60 years, urban population, and how well the provinces are communicated, with age having one of the most important associations, an increase of 1% in the corresponding rate implies a percentage increase of 46.65% over the number of cases. Of relevance, mortality attributable to COVID-19 complications is higher in this age group, and age increases the likelihood of developing the symptomatic disease and increased disease severity (34,35). Nevertheless, the association with older age could have different meanings depending on the number of comorbidities, with some reports labeling COVID-19 as an age-related disease (36). Our data demonstrate that the spatial spread of COVID-19 has a relationship with population aging structures, a concept that must be explored in this setting to obtain population-specific estimates and lethality and which could represent a significant structural inequality related to COVID-19 burden (37).

Urbanization rates also are associated with a percentage increase over the number of COVID-19 cases; we observed a similar association regarding province interconnectedness, which goes in line with recent information on human mobility and its effect in decreasing disease spread through social distancing (33). Urbanization, as a demographic phenomenon, leads to increased interconnectedness and human mobility as well as increased population density; these two factors facilitate disease spread. Emerging zoonotic diseases similar to SARS-CoV-2 have been linked to major structural factors that have been reported in other studies, including population growth, climate change, urbanization, and pollution (38,39). Thus, communication and the degree of urbanity, and what this implies in terms of pollution, overcrowding, among other factors, seem to be relevant to determine the number of COVID-19 cases and should imply geographical

targets for public health interventions to monitor disease spread and disease containment (40).

The only effect associated with a decrease in the number of COVID19 cases in our study was attributed to literacy, which might reflect several factors that ultimately influence disease spread. Data from several countries, including Iran, identified that higher health literacy was associated with a lower number of COVID-19 cases, probably reflecting attitudes towards public health measures including social distancing, early disease detection, and hand hygiene (41,42). Interestingly, this poses a potential public health intervention given that individuals with reduced health literacy, not only might have higher rates of COVID-19, but also increased likelihood for depression and impaired quality of life in suspected cases. Literacy's protective effect on disease spread also indicates a strong influence on social inequity and vulnerability as risk factors for COVID-19 spread, particularly on the influence of health equity, which will likely define the long-term impact of COVID-19 in many developing countries (43).

Concerning average temperature levels, we were able to obtain information associated only with the capitals and not the provinces, being a limitation of the analyzed information, obtaining some inconclusive results. On one hand, the global spatial autocorrelation was negative, though not statistically significant, indicating that global areas with higher temperatures are spatially related to areas with lower disease rates. On the other hand, on the spatial linear model, we derived that more temperature is associated with more cases. However; the former result does not contradict the latter since the direct effect in each province of a variable over the response is different from the spatial relationship between two variables. The latter considers one of the variables as spatially lagged (COVID-19), and thus the direct effect between variables in the same province is not included. In fact, this problem occurs in all the bivariate analysis, so care should be taken in all the interpretations. Notably, our results are consistent with previous analyses which have

analyzed the impact of climate on SARS-CoV-2 stability and spread (44). However, these results should be further studied considering the climate data limitations, that we obtained mixed results, and that some studies suggest there is no evidence that spread rates of the disease decline with higher temperatures (45).

Our study had some strengths and limitations. We approached COVID-19 using spatial analysis, which allowed us to identify province-level factors that are associated with the disease spread and which may be shared by other countries with similar socioeconomic or geographic structures by potentially identifying targets for country-wide public health interventions. This approach considers disease spread beyond individual-specific factors and could also be used to monitor areas of a potentially high number of undiagnosed cases that could facilitate disease spread and the surge of delayed waves of COVID-19 after initial mitigation (46,47). Methodologically, all our analyses consider the spatial nature of the data. We identified significant spatial clustering and in terms of the spatial multivariate linear model, by including a spatial effect, we consider that the number of cases in an area is affected by those in neighboring areas. In this way, a lack of independency between spatial units is considered, being independence assumed in a usual linear model, thus obtaining more precise estimations. Of course, other statistical methods are available for this task, as generalized linear mixed models or geographically weighted regression; however, they do not use spatial weights, making our results more comparable with the Moran's I or spatial clustering, which are based on such weights. A limitation of our approach is that most of the variables used to explain COVID-19 disease rates were taken from previous years and not updates, given the unavailability of recent estimates. Furthermore, smoothed COVID-19 rates were calculated using a projection of the population in 2020 since the most recent census corresponds to 2016, thus rates could have slightly different values. In this sense, the explanatory variables were not projected since information of previous years was not always available; however, precise projections

for each variable were out of the scope of this work; and, it is also possible that some variables have a time lagged effect over the response. However; the time lagged effect we included was unintentional and dependent on the information available and not considering a lagged time effect as defined by experts; for instance, for GDP we used a time lag of 16 years, when perhaps it should have been of fewer years. When obtaining estimates using both projected and population size in 2016, we observed no significant changes in the results, which confirms the robustness of our approach. In fact, with all the mathematical projection methods similar results were obtained. However, the projections by province could be improved by considering a demographic balance equation and probabilistic projection methods as the ones obtained by country by the UN (48). In this sense, we suspect similar results would still be obtained since our projected values by country are similar to those obtained by the UN. We also observed that the smoothed and raw rates of COVID-19 cases were similar, with an absolute difference between them of at most 0.607 (considering rates for every 1000 individuals), this was probably due to Iran not having provinces with extremely small or large population size. Future work could be focused on evaluating spatio-temporal modeling, which could be useful to monitor disease spread and identify additional factors relating not only to transmission rates but also to transmission dynamics. Since COVID-19 is currently challenging health systems all over the world, science-centered public health decisions could benefit from spatial modeling to investigate larger factors targeted for public health interventions.

In conclusion, COVID-19 spread within Iranian provinces is spatially correlated. The main factors associated with a high number of cases are older age, high degrees of urbanization, province interconnectedness, higher average temperatures, lower literacy rates, and the number of physicians. Structural determinants for the spread of emerging zoonotic diseases, including SARS-CoV-2, must be understood in order to implement

411 evidence-based regional public health policies aimed at improving mitigation policies and
412 diminishing the likelihood of disease re-emergence.

413 **ORCID**

414 Ricardo Ramírez-Aldana: 0000-0003-4344-2928

415 Juan Carlos Gómez-Verján: 0000-0001-7186-8067

416 Omar Yaxmehen Bello-Chavolla: 0000-0003-3093-937X

417

418 **Conflict of Interest:**

419 Nothing to disclose.

420 **Funding**:

421 This project was supported by a grant from the Secretaría de Educación, Ciencia,
422 Tecnología e Innovación de la Ciudad de México CM-SECTEI/041/2020 “Red
423 Colaborativa de Investigación Traslacional para el Envejecimiento Saludable de la Ciudad
424 de México (RECITES)”

425 **Author contributions**

426 Research idea and study design: RRA, JCGV, OYBC; data acquisition: RRA; data
427 analysis/interpretation: RRA, JCVG, OYBC; statistical analysis: RRA; manuscript drafting:
428 RRA, JCGV, OYBC; supervision or mentorship: RRA. Each author contributed important
429 intellectual content during manuscript drafting or revision and accepted accountability for
430 the overall work by ensuring that questions about the accuracy or integrity of any portion of
431 the work are appropriately investigated and resolved.

432

433 **REFERENCES**

434 1. Sohrabi C, Alsafi Z, O’Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health
435 Organization declares global emergency: A review of the 2019 novel coronavirus
436 (COVID-19). Int J Surg. 2020 Apr;76:71–76.

2. Forouzanfar MH, Sepanlou SG, Shahraz S, Dicker D, Naghavi P, Pourmalek F, et al. Evaluating causes of death and morbidity in Iran, global burden of diseases, injuries, and risk factors study 2010. Arch Iran Med. 2014 May;17(5):304–320.

3. Johns Hopkins Coronavirus Resource Center - Johns Hopkins Coronavirus Resource Center [Internet]. [cited 2020 Apr 17]. Available from: https://coronavirus.jhu.edu/map.html

4. Eden J-S, Rockett R, Carter I, Rahman H, de Ligt J, Hadfield J, et al. An emergent clade of SARS-CoV-2 linked to returned travellers from Iran. Virus Evol. 2020 Jan;6(1):veaa027.

5. How Iran Became a New Epicenter of the Coronavirus Outbreak | The New Yorker [Internet]. [cited 2020 Apr 17]. Available from: https://www.newyorker.com/news/our-columnists/how-iran-became-a-new-epicenter-of-the-coronavirus-outbreak

6. Lachmann A. Correcting under-reported COVID-19 case numbers. medRxiv. 2020 Mar 18;

7. COVID-19: Iran records 4,585 coronavirus deaths as restrictions eased | Mena – Gulf News [Internet]. [cited 2020 Apr 17]. Available from: https://gulfnews.com/world/mena/covid-19-iran-records-4585-coronavirus-deaths-as-restrictions-eased-1.70954074

8. GHO | By country | Iran (Islamic Republic of) - statistics summary (2002 - present) [Internet]. [cited 2020 Apr 17]. Available from: https://apps.who.int/gho/data/node.country.country-IRN

9. Bello-Chavolla OY, González-Díaz A, Antonio-Villa NE, Fermín-Martínez CA, Márquez-Salinas A, Vargas-Vázquez A, et al. 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. 2020 Jun 29;

10. Armitage R, Nellums LB. Water, climate change, and COVID-19: prioritising those in water-stressed settings. Lancet Planet Health. 2020;4(5):e175.

11. Gross B, Zheng Z, Liu S, Chen X, Sela A, Li J, et al. Spatio-temporal propagation of COVID-19 pandemics. medRxiv. 2020 Mar 27;

12. Chipeta MG, Giorgi E, Mategula D, Macharia PM, Ligomba C, Munyenyembe A, et al. Geostatistical analysis of Malawi's changing malaria transmission from 2010 to 2017. [version 2; peer review: 2 approved, 1 approved with reservations]. Wellcome Open Res. 2019 Jul 4;4:57.

13. Sanna M, Wu J, Zhu Y, Yang Z, Lu J, Hsieh Y-H. Spatial and temporal characteristics of 2014 dengue outbreak in guangdong, china. Sci Rep. 2018 Feb 5;8(1):2344.

14. Carmo RF, Nunes BEBR, Machado MF, Armstrong AC, Souza CDF. Expansion of COVID-19 within Brazil: the importance of highways. J Travel Med. 2020 Jun 27;

15. Deutsch-Feldman M, Brazeau NF, Parr JB, Thwai KL, Muwonga J, Kashamuka M, et al. Spatial and epidemiological drivers of Plasmodium falciparum malaria among adults in the Democratic Republic of the Congo. BMJ Glob Health. 2020 Jun;5(6).

16. Lin D, Cui Z, Chongsuvivatwong V, Palittapongarnpim P, Chaiprasert A, Ruangchai W, et al. The geno-spatio analysis of Mycobacterium tuberculosis complex in hot and cold spots of Guangxi, China. BMC Infect Dis. 2020 Jul 1;20(1):462.

17. Presidency of the I.R.I. Statistical Center of Iran [Internet]. Statistical Center of Iran. 2020 [cited 2020 Jul 22]. Available from: https://www.amar.org.ir/english/

18. Iran Data Portal [Internet]. [cited 2020 Apr 17]. Available from: https://irandataportal.syr.edu/

19. Performance Evaluation of the Provinces of Iran Reading To the Measures of Freight and Passenger Transportation [Internet]. [cited 2020 Apr 17]. Available from: http://www.ijte.ir/article_69681.html

20. Islamic Republic News Agency. Iran's coronavirus toll update March 22, 2020 [Internet]. 2020 [cited 2020 Jul 21]. Available from: http://archive.vn/nVmn5

21. First-level Administrative Divisions, Iran, 2015 | Stanford Digital Repository [Internet]. [cited 2020 Sep 17]. Available from: http://purl.stanford.edu/dv126wm3595

22. Rate Transformations and Smoothing Luc Anselin | Semantic Scholar [Internet]. [cited 2020 Apr 17]. Available from: https://www.semanticscholar.org/paper/Rate-Transformations-and-Smoothing-Luc-Anselin-Lozano-Koschinsky/88d8b02de84f97f556cfe0ef5a91a7df229cf363#paper-header

23. Cressie, Noel. Statistics For Spatial Data (wiley Classics Library). Revised. Hoboken, NJ: Wiley-interscience; 1993.

24. Moran PAP. Notes on continuous stochastic phenomena. Biometrika. 1950 Jun;37(1/2):17.

25. Wrigley N, Cliff AD, Ord JK. Spatial processes: models and applications. Geogr J. 1982 Nov;148(3):383.

26. Anselin L. Local Indicators of Spatial Association-LISA. Geogr Anal. 2010 Sep 3;27(2):93–115.

27. Lopes D, Assunção R. Visualizing Marked Spatial and Origin-Destination Point Patterns With Dynamically Linked Windows. Journal of Computational and Graphical Statistics. 2012 Jan;21(1):134–154.

28. Ward M, Gleditsch K. Spatial Regression Models. 2455 Teller Road, Thousand Oaks California 91320 United States of America : SAGE Publications, Inc.; 2008.

29. Silvey SD. The lagrangian multiplier test. Ann Math Statist. 1959 Jun;30(2):389–407.

30. Sahafizadeh E, Sartoli S. Estimating the reproduction number of COVID-19 in Iran using epidemic modeling. medRxiv. 2020 Mar 23;

31. He J, Chen G, Jiang Y, Jin R, He M, Shortridge A, et al. Comparative Analysis of COVID-19 Transmission Patterns in Three Chinese Regions vs. South Korea,Italy and Iran. medRxiv. 2020 Apr 14;

32. Muniz-Rodriguez K, Fung IC-H, Ferdosi SR, Ofori SK, Lee Y, Tariq A, et al. Transmission potential of COVID-19 in Iran. medRxiv. 2020 Mar 10;

33. Kraemer MUG, Yang C-H, Gutierrez B, Wu C-H, Klein B, Pigott DM, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science. 2020 Mar 25;

34. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet. 2020 Mar 28;395(10229):1054–1062.

35. Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nat Med. 2020 Apr;26(4):506–510.

36. Santesmasses D, Castro JP, Zenin AA, Shindyapina AV, Gerashchenko MV, Zhang B, et al. COVID-19 is an emergent disease of aging. medRxiv. 2020 Apr 15;

37. Kashnitsky I, Aburto JM. COVID-19 in unequally ageing European regions. 2020 Mar 18;

38. Liu Q, Cao L, Zhu X-Q. Major emerging and re-emerging zoonoses in China: a matter of global health and socioeconomic development for 1.3 billion. Int J Infect Dis. 2014 Aug;25:65–72.

39. Wu T, Perrings C, Kinzig A, Collins JP, Minteer BA, Daszak P. Economic growth, urbanization, globalization, and the risks of emerging infectious diseases in China: A review. Ambio. 2017 Feb;46(1):18–29.

40. Kost GJ. Geospatial Hotspots Need Point-of-Care Strategies to Stop Highly Infectious Outbreaks: Ebola and Coronavirus. Arch Pathol Lab Med. 2020 Apr 16;

41. Lin Y-H, Liu C-H, Chiu Y-C. Google searches for the keywords of “wash hands” predict the speed of national spread of COVID-19 outbreak among 21 countries. Brain Behav Immun. 2020 Apr 10;

42. Chung RY-N, Dong D, Li MM. Socioeconomic gradient in health and the covid-19 outbreak. BMJ. 2020 Apr 1;369:m1329.

43. Wang Z, Tang K. Combating COVID-19: health equity matters. Nat Med. 2020 Apr;26(4):458.

44. Aboubakr HA, Sharafeldin TA, Goyal SM. Stability of SARS-CoV-2 and other coronaviruses in the environment and on common touch surfaces and the influence of climatic conditions: a review. Transbound Emerg Dis. 2020 Jun 30;

45. Jamil T, Alam IS, Gojobori T, Duarte C. No Evidence for Temperature-Dependence of the COVID-19 Epidemic. medRxiv. 2020 Mar 31;

46. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science. 2020 Apr 14;

47. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). Science. 2020 Mar 16;

48. World Population Prospects - Population Division - United Nations [Internet]. [cited 2020 Jul 23]. Available from: https://population.un.org/wpp/Download/Probabilistic/Population/

TABLES

Table 1. Features extracted for spatial analyses disaggregated by Iranian provinces to predict the spread of COVID-19 cases.

Abbreviations: GDP, Gross Domestic Product; TEI, Transportation Efficiency Index

Province	Cases[+]	Urban population (%)*	>60 years (%)*	Area (km^2)*	Density*	Literacy (%)*	Average temperature (°C)*	Annual precipitation (mm)*	Physicians**	GDP** (2004)	Hospital beds** (2006)	Inflation*	TEI***	Population**** (2020)
Alborz	906	92.639	8.914	5122	529.559	92.2	16.7	220.5	2632.5	12.577	15327.5	28	0.524	2952309.6
Ardebil	213	68.169	9.371	17800	71.372	83.1	10.9	296.5	354	1.127	1654	23.4	0.46	1287965.6
Bushehr	46	71.854	6.841	22743	51.154	89.2	26.5	272.5	429	3.227	1345	24.4	0.301	1267760.8
Chahar Mahall and Bakhtiari	58	64.092	8.691	16328	58.045	84.7	11.8	309.7	499	0.727	1234	22.7	0.754	989763
East Azarbaijan	571	71.859	10.732	45651	85.642	84.7	14	286.9	1104	3.927	5964	21	0.56	4057677.6
Esfahan	1538	88.019	10.643	107018	47.850	89.9	17.7	96.3	2109	6.527	8261	24.2	0.696	5314080.4
Fars	386	70.119	9.456	122608	39.567	88.8	18.9	271.5	1661	4.527	7154	22.3	0.591	5054966.8
Gilan	924	63.343	13.250	14042	180.223	87.3	17.3	1388.3	1211	2.327	3716	24	0.472	2570553.6
Golestan	351	53.275	7.796	20367	91.757	86.1	18.8	477.8	998	1.527	1769	25.4	0.372	1942263
Hamadan	155	63.123	10.801	19368	89.748	85	13.1	215.7	688	1.627	3089	22.9	0.429	1722206.8
Hormozgan	124	54.707	6.046	70697	25.127	87.8	27.8	152.2	492	2.227	1686	32.1	1	1935000.6
Ilam	120	68.130	8.508	20133	28.816	84.9	18	842.4	145	0.827	875	27.9	1	598205.2
Kerman	127	58.728	7.811	180726	17.511	81.5	17.2	109.8	955	2.527	3325	25.7	0.47	3345302
Kermanshah	152	75.220	10.023	25009	78.069	85.4	16.5	512.8	755	1.627	2922	22.6	0.536	1958199.6
Khuzestan	359	75.453	7.052	64055	73.539	86.3	27.3	269.7	1599	14.627	7511	22.3	0.95	4853540.2
Kohgiluyeh and Buyer Ahmad	45	55.741	7.139	15504	45.991	84.4	15.7	611.1	232	4.027	573	24.1	1	756590.4
Kordestan	189	70.756	9.304	29137	55.016	84.5	15.4	444.4	605	1.127	2155	18.8	0.818	1690503.8
Lorestan	363	64.460	8.830	28294	62.227	83	17.9	535.6	616	1.327	2153	26.7	0.963	1765773.8
Markazi	782	76.935	10.892	29127	49.077	87	15.1	284.8	514	2.327	1866	23.4	0.689	1441887.8
Mazandaran	1494	57.780	11.414	23842	137.723	88.7	18.6	724.7	1585	3.527	4475	25.9	0.269	3451293.2

North Khorasan	100	56.118	8.500	28434	30.354	83.3	14.8	227.4	288	0.727	730	24.7	0.533	859384
Qazvin	526	74.751	8.925	15567	81.824	88.6	15.7	313.7	429	1.427	1403	25.4	0.544	1331517.8
Qom	1074	95.178	7.696	11526	112.119	88.7	19.6	111.6	319	1.127	1493	24.6	1	1404771.8
Razavi Khorasan	661	73.058	8.478	118851	54.139	89.1	17.2	183.4	3328	5.027	9131	20.5	0.658	6786580.2
Semnan	577	79.803	9.978	97491	7.204	91.5	19.5	107.5	493	0.927	1269	22.7	0.868	759273.6
Sistan and Baluchestan	88	48.491	4.886	181785	15.265	76	19.8	103.7	657	1.127	2117	26.5	1	2967563.6
South Khorasan	100	59.023	9.757	95385	8.061	86.8	17.4	144.3	512	0.527	660	24.5	0.605	853989.2
Tehran	4260	93.854	10.443	13692	969.007	92.9	19.1	209.3	2632.5	12.577	15327.5	28	1	14135033.8
West Azarbaijan	300	65.423	8.562	37411	87.280	82	12.5	277.3	993	2.027	3630	23.3	0.644	3412933.4
Yazd	471	85.316	8.788	129285	8.806	90.9	21.3	38.4	610	1.227	2395	23.1	0.941	1189817
Zanjan	261	67.253	9.783	21773	48.568	84.8	14	283.1	492	1.027	1264	22.4	0.651	1090842.6

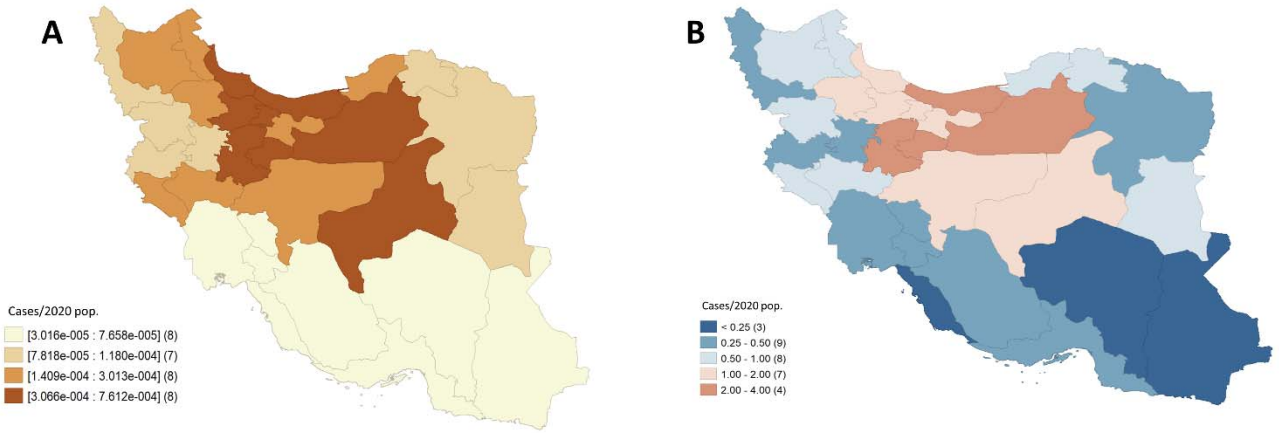
564 * Statistical Centre of Iran
565 ** Iran data portal
566 *** Obtained from reference (12)
567 **** Population projected by using the population census 2011 and 2016 and an arithmetic method
568 [+]Cases obtained from John Hopkins Database (https://coronavirus.jhu.edu/map.html)

Table 2. Spatial lag models estimated via maximum likelihood to predict log-transformed COVID-19 case distribution between Iranian provinces (model including all variables and model obtained through a selection scheme).

Variable	Coefficient		SE		z-value		p-value	
	All variables*	Selection scheme**	All variables	Selection scheme	All variables	Selection scheme	All variables	Selection scheme
Spatial parameter (ρ)	0.723	0.737	0.107	0.104	6.734	7.069	<0.001	<0.001
Model constant	2.510	2.853	2.550	2.425	0.984	1.176	0.325	0.239
Urban population (%)	0.026	0.026	0.011	0.010	2.345	2.653	0.019	0.008
Population aged ≥60	0.383	0.331	0.075	0.062	5.089	5.324	<0.001	<0.001
Population density	-0.0002		0.0007		-0.258		0.797	
Literacy	-0.110	-0.103	0.040	0.040	-2.771	-2.591	0.006	0.010
Average temperature (°C)	0.113	0.114	0.030	0.028	3.748	4.105	<0.001	<0.001
Precipitation levels (mm)	-0.0003		0.0003		-0.98		0.327	
Physician distribution	0.0008	0.0008	0.0001	0.0001	5.740	5.746	<0.001	<0.001
GDP	-0.051	-0.057	0.038	0.034	-1.343	-1.699	0.179	0.089
Consumer Price Index	0.032		0.039		0.812		0.417	
TEI	0.157	0.155	0.074	0.074	2.112	2.081	0.035	0.038

***Likelihood Ratio Test=15.628, p< 0.001 (no spatial vs spatial model); R2=0.877; AIC=57.165; σ^2= 0.146**
**** Likelihood Ratio Test=18.682, p< 0.001 (no spatial vs spatial model); R2=0.872; AIC=52.704; σ^2= 0.152**

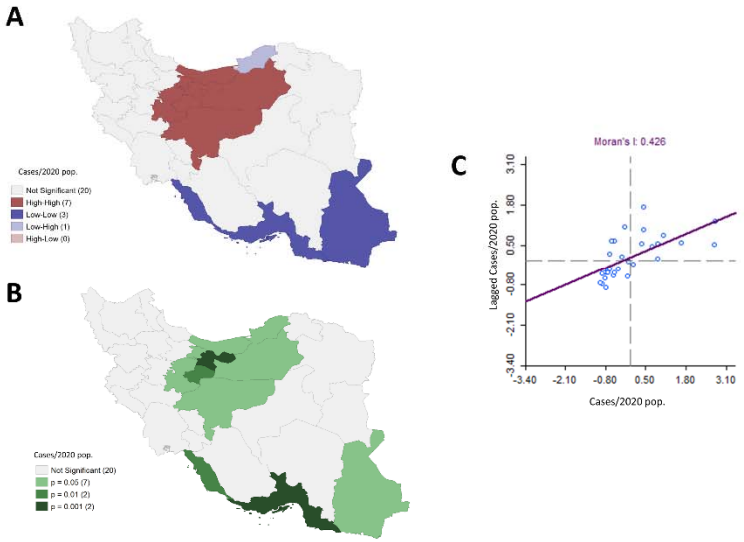
574 **FIGURE LEGENDS**



575

576 **Figure 1. Maps associated with COVID-19 cases by province from February 19th to March 20th, 2020.** A) Quartiles were

577 corresponding to rates smoothed through an empirical Bayes procedure. B) Excess or relative risk.



578

579 **Figure 2. Spatial clustering associated with rates of COVID-19 cases by province from February 19th to March 20th,**
580 **considering queen contiguity weights.** A) Significant spatial clustering obtained through Local Indicators of Spatial Autocorrelation
581 (LISA) comparisons. Four types of a cluster are possible: High-High, Low-Low, High-Low, and Low-High. For instance, the High-High
582 cluster (red) indicates provinces with high values of a variable that are significantly surrounded by regions with similarly high values.
583 B) P-values associated with the spatial clustering in A), C) Scatter plot associated with the smoothed rates vs. their corresponding
584 spatially lagged values, including the associated linear regression fitting, whose slope corresponds to the Moran's I statistic, a global
585 spatial autocorrelation measure.

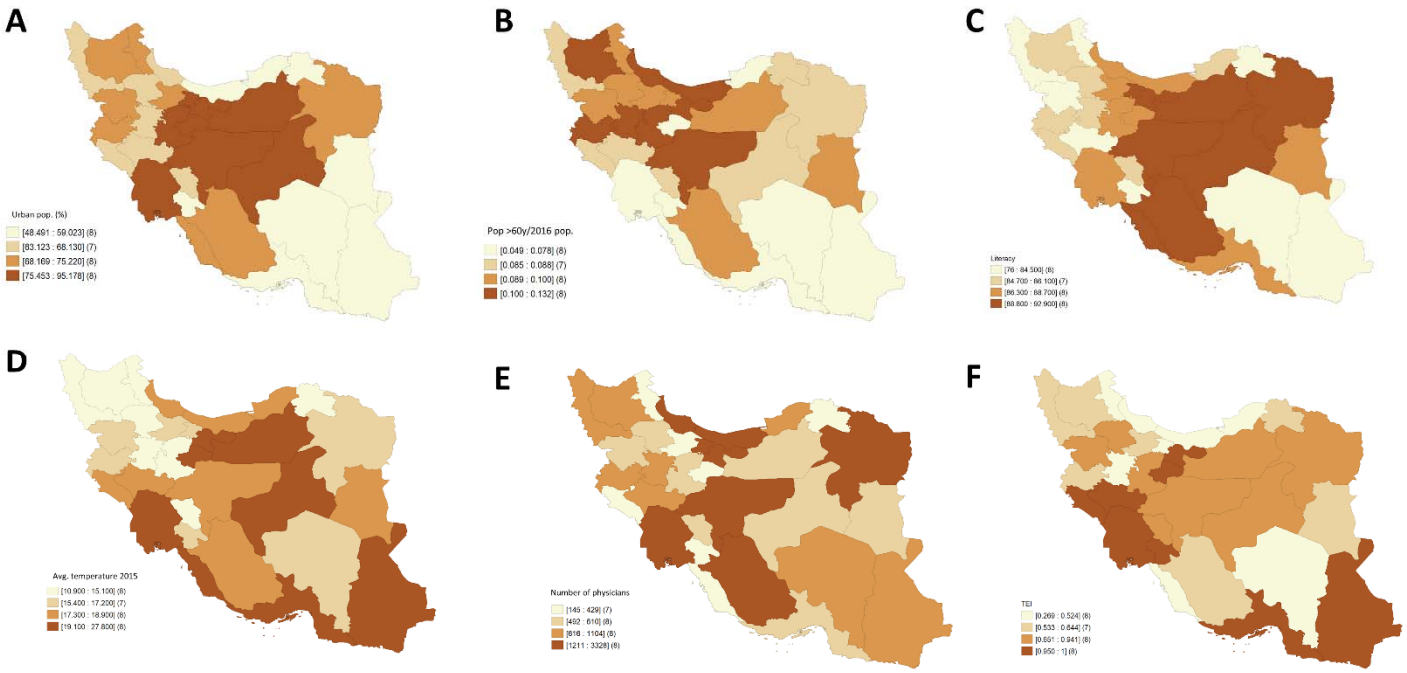


Figure 3. Quartiles associated with all the explanatory variables significant in the spatial lag model with response variable the logarithm of the number of COVID-19 cases. A) People settled in urban areas in 2016 (%). B) People aged ≥60 years, rates obtained through empirical Bayes smoothing. C) Literacy of population aged ≥6 years in 2016 (%). D) Average temperature (°C) of provincial capitals in 2015. E) Number of physicians employed by the ministry of health and medical education in 2006. F) Transportation Efficiency Index (TEI).

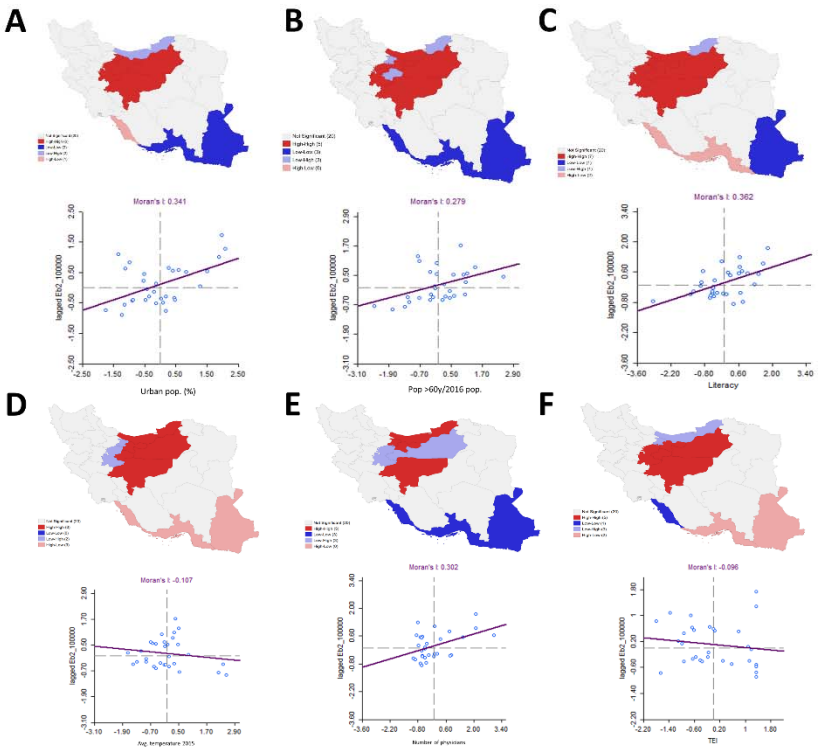
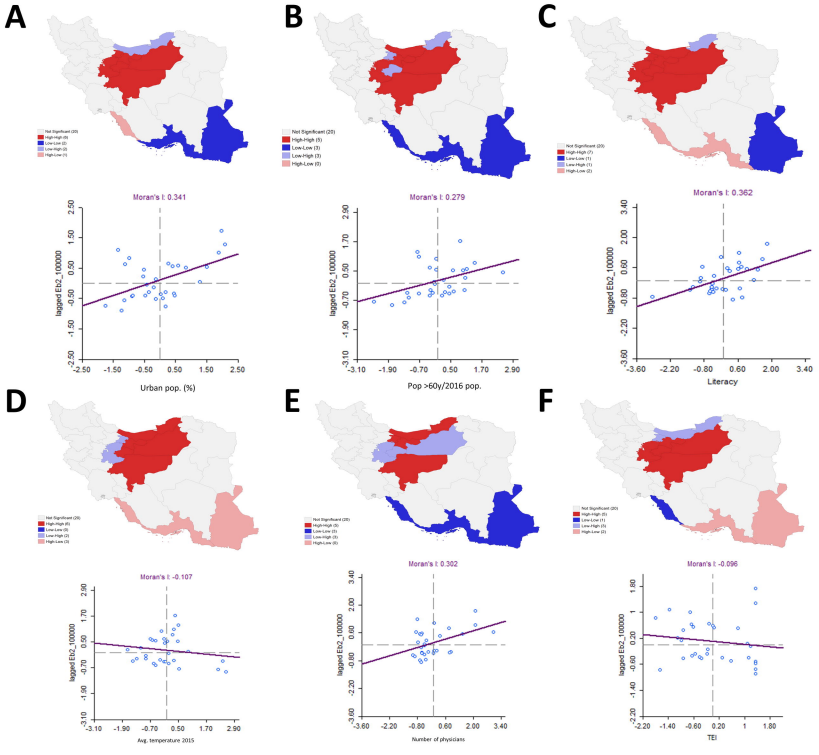
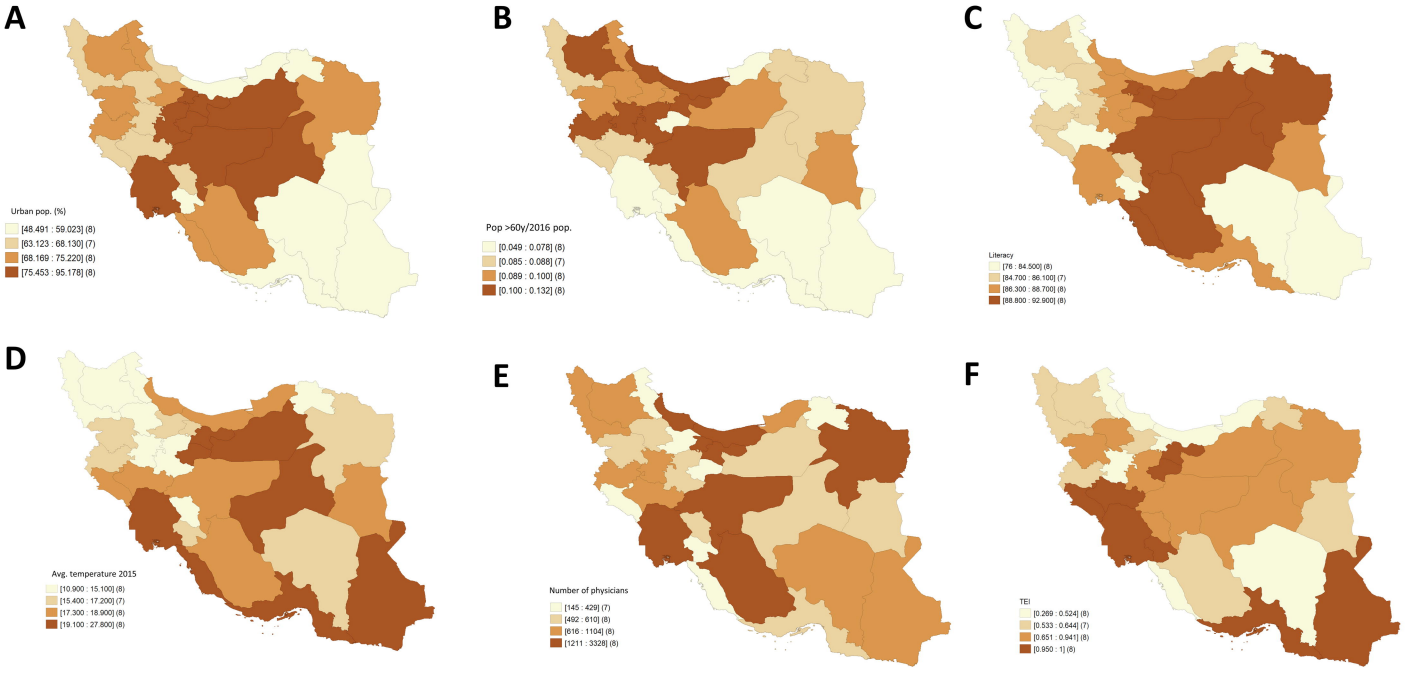


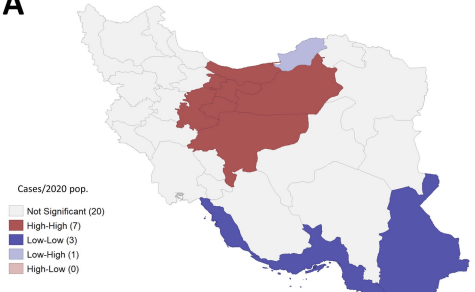
Figure 4. Bivariate LISA's significant spatial clustering between each of the significant variables in the spatial lag model and the rate of cases with COVID-19 smoothed through the empirical Bayes approach. The scatter plots associated with the variables vs. the spatially lagged smoothed rate of COVID-19 cases are presented as well, including the associated linear regression fitting, whose slope corresponds to the bivariate Moran's I statistic, a global spatial bivariate autocorrelation measure. A) People settled in urban areas in 2016 (%). B) People aged ≥60 years, rates obtained through empirical Bayes smoothing. C) Literacy of

599 population aged ≥6 years in 2016 (%). D) Average temperature (°C) of provincial capitals in 2015. E) Number of physicians employed
600 by the ministry of health and medical education in 2006. F) Transportation Efficiency Index (TEI).
601
602
603

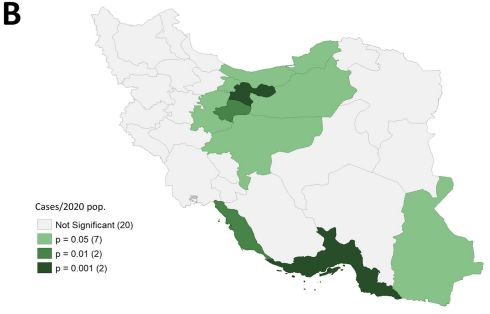

A
B
C
D
E
F
Moran's I: 0.341
Moran's I: 0.279
Moran's I: 0.362
Moran's I: -0.107
Moran's I: 0.302
Moran's I: -0.096
lagged Eb2_100000
Urban pop. (%)
Pop >60y/2016 pop.
Literacy
Avg. temperature 2015
Number of physicians
TEI


A
Urban pop. (%)
[48.491 : 59.023] (8)
[63.123 : 68.130] (7)
[68.169 : 75.220] (8)
[75.453 : 95.178] (8)
B
Pop >60y/2016 pop.
[0.049 : 0.078] (8)
[0.085 : 0.088] (7)
[0.089 : 0.100] (8)
[0.100 : 0.132] (8)
C
Literacy
[76 : 84.500] (8)
[84.700 : 86.100] (7)
[86.300 : 88.700] (8)
[88.800 : 92.900] (8)
D
Avg. temperature 2015
[10.900 : 15.100] (8)
[15.400 : 17.200] (7)
[17.300 : 18.900] (8)
[19.100 : 27.800] (8)
E
Number of physicians
[145 : 429] (7)
[492 : 610] (8)
[616 : 1104] (8)
[1211 : 3328] (8)
F
TEI
[0.289 : 0.524] (8)
[0.533 : 0.644] (7)
[0.651 : 0.941] (8)
[0.950 : 1] (8)

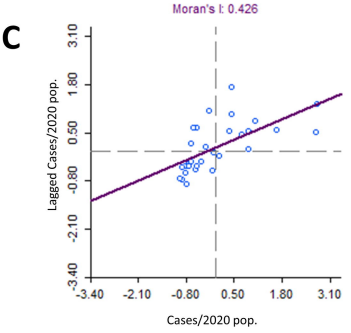
A

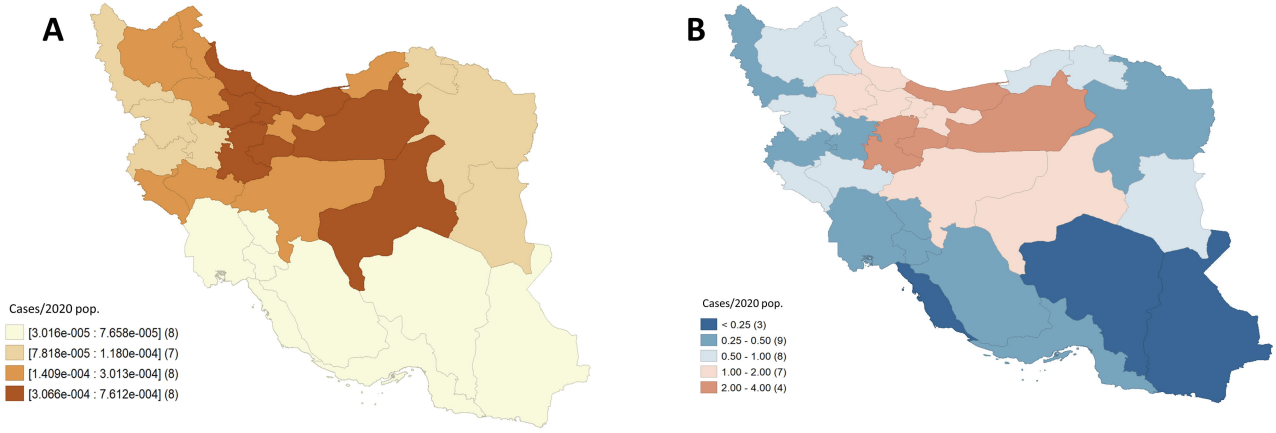

Cases/2020 pop.
Not Significant (20)
High-High (7)
Low-Low (3)
Low-High (1)
High-Low (0)

B


Cases/2020 pop.
Not Significant (20)
p = 0.05 (7)
p = 0.01 (2)
p = 0.001 (2)

C


Moran's I: 0.426
Lagged Cases/2020 pop.
3.10
1.80
0.50
-0.80
-2.10
-3.40
-3.40
-2.10
-0.80
0.50
1.80
3.10
Cases/2020 pop.


A
Cases/2020 pop.
[3.016e-005 : 7.658e-005] (8)
[7.818e-005 : 1.180e-004] (7)
[1.409e-004 : 3.013e-004] (8)
[3.066e-004 : 7.612e-004] (8)
B
Cases/2020 pop.
< 0.25 (3)
0.25 - 0.50 (9)
0.50 - 1.00 (8)
1.00 - 2.00 (7)
2.00 - 4.00 (4)