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 at the 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. 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 the socioeconomic features of Iranian provinces might predict the number of cases. We identified a cluster of provinces with significantly higher rates of COVID-19 cases around Tehran, which indicated that the spread of COVID-19 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. Interestingly, literacy is a protective factor that might be directly related to health literacy and compliance with public health measures. These features indicate that policies related to social distancing, protecting older adults, and vulnerable populations, as well as promoting health literacy, might be targeted to reduce SARS-CoV2 spread in Iran. 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.

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 has one of the highest COVID-19 mortality rates fluctuating between 8 to 18 percent daily, 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 or 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 economic situation and inflation rates are among the highest in the region, which 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 novel multidisciplinary tool for the analysis of spatial epidemiology, concerning mapping and statistical analyses of spatial and spatial-temporal incidences of different pathogens. Spatial analyses could contribute to public health measures by providing insight to inform the implementation of interventions or to understand socio-demographic factors that contribute to SARS-CoV2 spread and COVID-19 heterogeneity as it has been applied to previous infectious diseases (9). Here, we performed spatial analyses in order to understand better the COVID-19 outbreak situation on Iran, which, by considering the socioeconomic and structural factors which facilitate disease spread within Iranian provinces could aid to understand the burden of COVID-19 and its implication on public health within Iran and similar countries(10).

MATERIAL AND METHODS

Data sources

To obtain province-specific data, we extracted information mainly from the Statistical Center of Iran and the Iran data portal (11). We used and calculated variables at a province-level considering 31 provinces or polygons in Iran **(Table 1)** including: people settled in urban areas in 2016 (%), people aged ≥60 years in 2016, population density (people per km^2) in 2016, literacy rate of population aged ≥6 years in 2016, average temperature (°C) of provincial capitals in 2015, annual precipitation levels (mm) in 2015, number of physicians employed by the ministry of health and medical education in 2006, province contribution to gross domestic product (GDP) in 2004, number of beds in operating medical establishments in 2006, 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 a Transportation Efficiency Index (TEI)(12), with values between zero and one, where values near to one indicate provinces better

communicated, and considering such scale; we decided to standardize the TEI. The total number of cases with confirmed COVID-19 by Province from February 19th to March 20th, 2020, was also obtained. 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 linearly extrapolated using the corresponding information in the population and housing censuses from 2011 and 2016.

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 (13). Since raw and smoothed rates were surprisingly similar, only results of smoothed rates are reported. We also obtained quantile 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 (14). 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, and its significance was assessed through a random permutation inference technique based on Monte Carlo simulations (15). 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 (16). 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 (17). 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 linear models

Spatial linear models were fitted to identify variables that significantly impact the number of log-transformed COVID-19 cases (18). 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. 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. For significant variables in the linear spatial model, interpretations in the original scale (i.e., as counts) were derived. All statistical analyses were conducted using GeoDa version 1.14.0. A two-tailed p-value $p<0.05$ 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 are located in the Northern region of Iran and correspond to the provinces of Qom, Marzaki, Mazandaran, and Semnan, with significant rates also for 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 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 purple, has significantly lower COVID-19 rates despite being surrounded by a cluster of provinces with higher rates (**Figure 2**).

Selection of linear spatial model for COVID-19 spread

We confirmed that a spatial model was necessary since the error term from the OLS fitting showed significant spatial autocorrelation (Moran's I= 0.106, p-value=0.042). 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=8.1953, p-value=0.004; Robust LM=10.568, p-value=0.0011), which did not occur with the spatial error model. Fitting a spatial lag model, we obtained a significant spatial parameter (ρ = 0.687, 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, $\sigma^2 = 0.1455$). Normality and homoscedasticity assumptions were reasonably satisfied.

Predictors of COVID-19 spatial spread in Iranian provinces

Variables that significantly impact over a log-transformed number of COVID-19 cases included the percentage of people settled in urban areas, smoothed rate of people aged ≥60 years, literacy rate, average temperature, number of physicians employed, and the TEI (**Table 2**). A 10% increase in urban population or a 1% increase in the population aged ≥60 years has a multiplicative effect of 1.370 (95%CI 1.282-1.733) and 1.490 (95%CI 1.337-1.405), 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 multiplicative effects of 1.112 (95%CI 1.046-1.182), 1.0011 (95%CI 1.0004-1.0018), and 1.160 (95%CI 1.003-1.342), respectively, over the number of COVID-19 cases. Finally, a 1% increase in the literacy rate showed a protector effect of 0.887 (95%CI 0.817-0.962) on the number of cases. An increase in the number of COVID-19 cases is strongly associated with the urban population, people aged ≥60 years, and how well the provinces are communicated. These interpretations do not consider the spatial effect, which is accumulated since the spatially lagged response is part of the explanatory variables and consider fixed values for all variables except the one being interpreted.

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, we performed bivariate LISA significant clustering between each of the identified significant variables of the spatial lag model and the rate of cases with COVID-19, smoothed through the empirical Bayes approach. 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 confirmed a positive spatial relationship between literacy and disease rates (Moran's I=0.362, p-value=0.005). This result and the High-High cluster for COVID-19 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), indicating that global areas with higher temperatures are spatially related to areas with lower disease rates. This result does not contradict the result obtained in the spatial lag model 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. 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.

As expected, 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. Overall, the variables spatially correlate with the COVID-19 province clustering indicating a consistent association with the observed variables.

DISCUSSION

Here, we demonstrate that the smoothed rates of COVID-19 cases within Iranian provinces are spatially correlated, probably due to the origin of the outbreak, since there is an important province cluster with the highest number of COVID-19 cases in the north 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 (19–21). 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-CoV2 transmission and provide additional evidence that human mobility or province interconnectedness might be significant in favoring disease spread (22).

Importantly, our approach demonstrates that urbanization, aging population, education, average temperatures, number of physicians, and inter-province communications are determinants of case numbers amongst Iranian provinces. The rate of older adults ≥60 years has one of the most important effects over the number of cases of COVID-19, an increase of 1% in the rate has a multiplicative effect of 1.490 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 (23,24). 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 (25). 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 (26).

Urbanization rates also show a multiplicative effect 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 (22). 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-CoV2 have been linked to major structural factors that have been reported in other studies, including population growth, climate change, urbanization, and pollution (27,28). 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 (29).

The only protective effect observed 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 (32,33). 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 (34).

Our study had some strengths and limitations. We approached COVID-19 using spatial analysis, which allowed us to identify province-level factors that facilitate 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 (35,36). 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. When fitting estimates using both projected and reported cases, we observed no significant changes in the model, which confirms the feasibility of our approach. 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 which determine 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-CoV2, must be understood in order to implement evidence-based regional public health policies aimed at improving mitigation policies and diminishing the likelihood of disease re-emergence.

ORCID

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

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

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

Conflict of Interest:

Nothing to disclose.

Funding:

Nothing to disclose

Author contributions

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

REFERENCES

1. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (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. 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.
10. 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;
11. Iran Data Portal [Internet]. [cited 2020 Apr 17]. Available from: https://irandataportal.syr.edu/
12. 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

13. 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
14. Cressie, Noel. Statistics For Spatial Data (Wiley Classics Library). Revised. Hoboken, NJ: Wiley-InterScience; 1993.
15. Moran PAP. Notes on continuous stochastic phenomena. Biometrika. 1950 Jun;37(1/2):17.
16. Anselin L. Local Indicators of Spatial Association-LISA. Geogr Anal. 2010 Sep 3;27(2):93–115.
17. 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.
18. Ward M, Gleditsch K. Spatial Regression Models. 2455 Teller Road, Thousand Oaks California 91320 United States of America : SAGE Publications, Inc.; 2008.
19. Sahafizadeh E, Sartoli S. Estimating the reproduction number of COVID-19 in Iran using epidemic modeling. medRxiv. 2020 Mar 23;
20. 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;
21. 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;
22. 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;

23. 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.
24. 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.
25. 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;
26. Kashnitsky I, Aburto JM. COVID-19 in unequally ageing European regions. 2020 Mar 18;
27. 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.
28. 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.
29. Kost GJ. Geospatial Hotspots Need Point-of-Care Strategies to Stop Highly Infectious Outbreaks: Ebola and Coronavirus. Arch Pathol Lab Med. 2020 Apr 16;
30. Kassem AZE. Do Weather Temperature and Median-age affect COVID-19 Transmission? medRxiv. 2020 Apr 17;
31. Pawar Shrikant, Stanam A, Chaudhari M, Rayudu D. Effects of temperature on COVID-19 transmission. medRxiv. 2020 Mar 30;
32. 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;

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

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

35. 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;

36. 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;

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

Table 2. A spatial lag model estimated via maximum likelihood to predict log-transformed COVID-19 case distribution between Iranian provinces.

Variable	Coefficient	SE	z-value	p-value
Spatial parameter (ρ)	0.687	0.117	5.851	<0.001
Model constant	2.748	2.564	1.072	0.284
Urban population (%)	0.032	0.013	2.485	0.013
Population aged ≥60	0.399	0.077	5.183	<0.001
Population density	0.0002	0.0008	0.280	0.779
Literacy	-0.122	0.041	-2.896	0.004
Average temperature (°C)	0.106	0.031	3.404	0.001
Precipitation levels (mm)	-.0004	0.0003	-1.186	0.236
Physician distribution	0.0011	0.0003	3.237	0.001
GDP	-0.020	0.052	-0.374	0.708
Hospital beds	-0.0001	0.0001	-0.867	0.386
Consumer Price Index	0.048	0.043	1.121	0.262
TEI	0.145	0.074	1.999	0.046

Likelihood Ratio Test=12.8143, p= 0.00034 (no spatial vs spatial model); R^2=0.877; AIC=58.421; σ^2= 0.1455

FIGURE LEGENDS

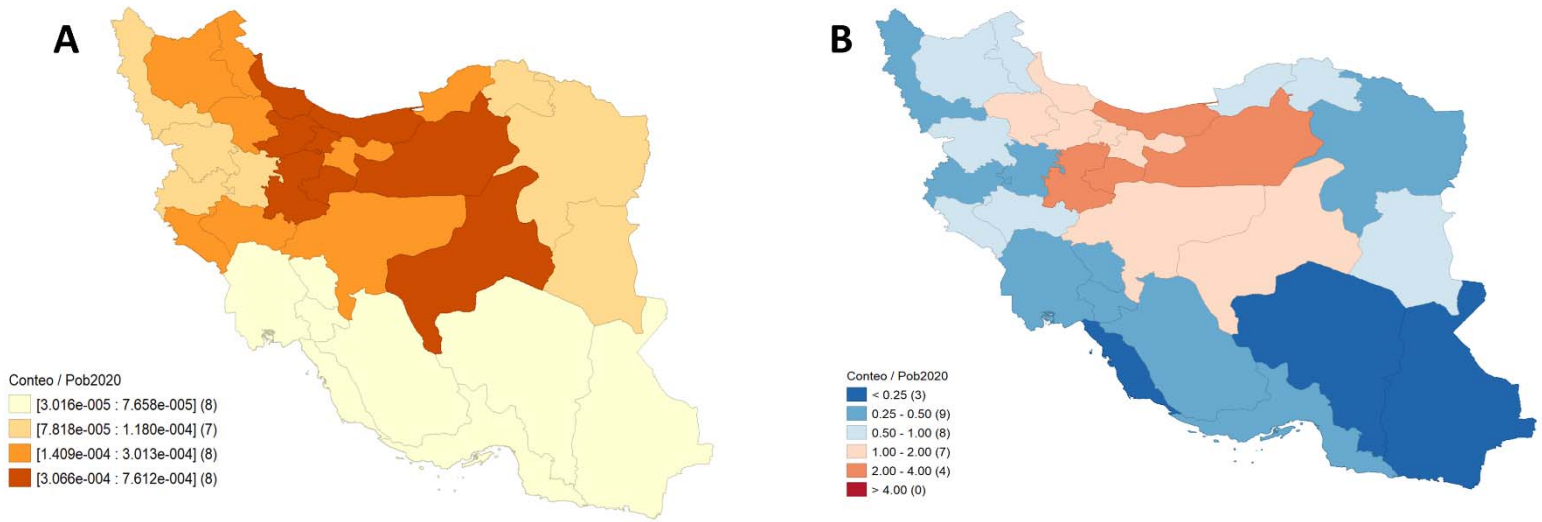


Figure 1. Maps associated with COVID-19 cases by province from February 19th to March 20th, 2020. A) Quartiles were corresponding to rates smoothed through an empirical Bayes procedure. B) Excess or relative risk.

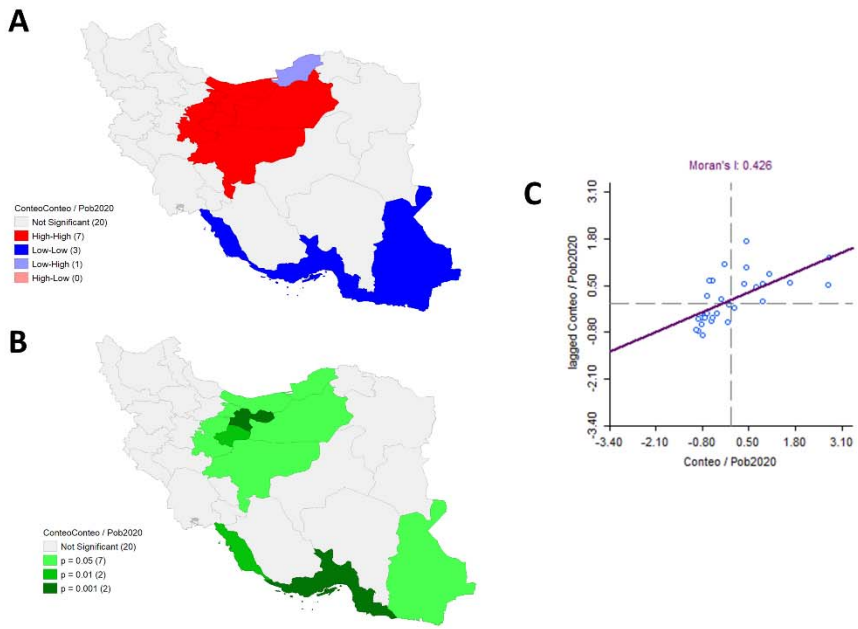


Figure 2. Spatial clustering associated with rates of COVID-19 cases by province from February 19th to March 20th, considering queen contiguity weights. A) Significant spatial clustering obtained through Local Indicators of Spatial Autocorrelation (LISA) comparisons. Four types of a cluster are possible: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

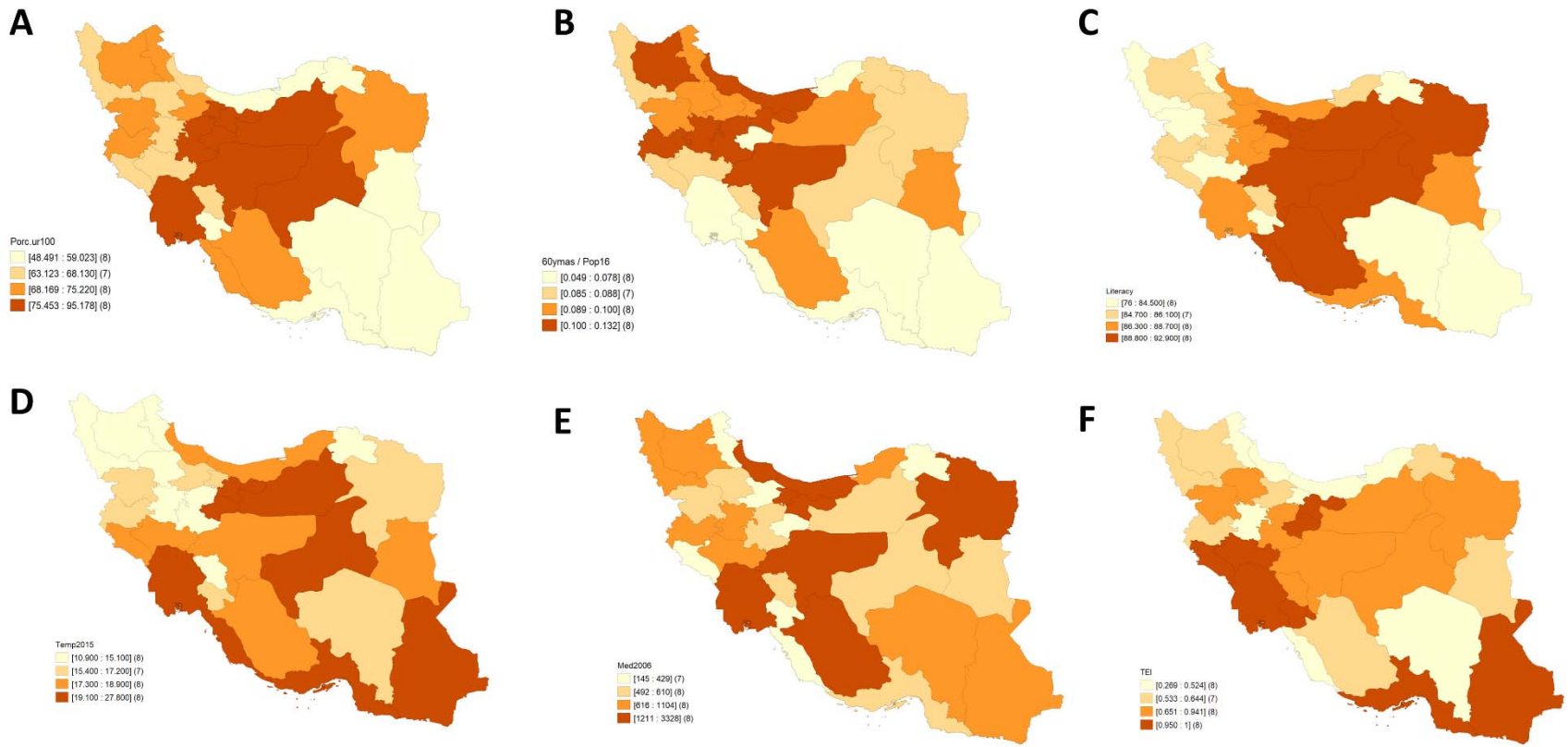


Figure 3. Quartiles associated with all the explanatory variables significant in the spatial lag model with response variable the logarithm of the number of COVID-19 cases. A) People settled in urban areas in 2016 (%). B) People aged ≥60 years, rates obtained through empirical Bayes smoothing. C) Literacy 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).

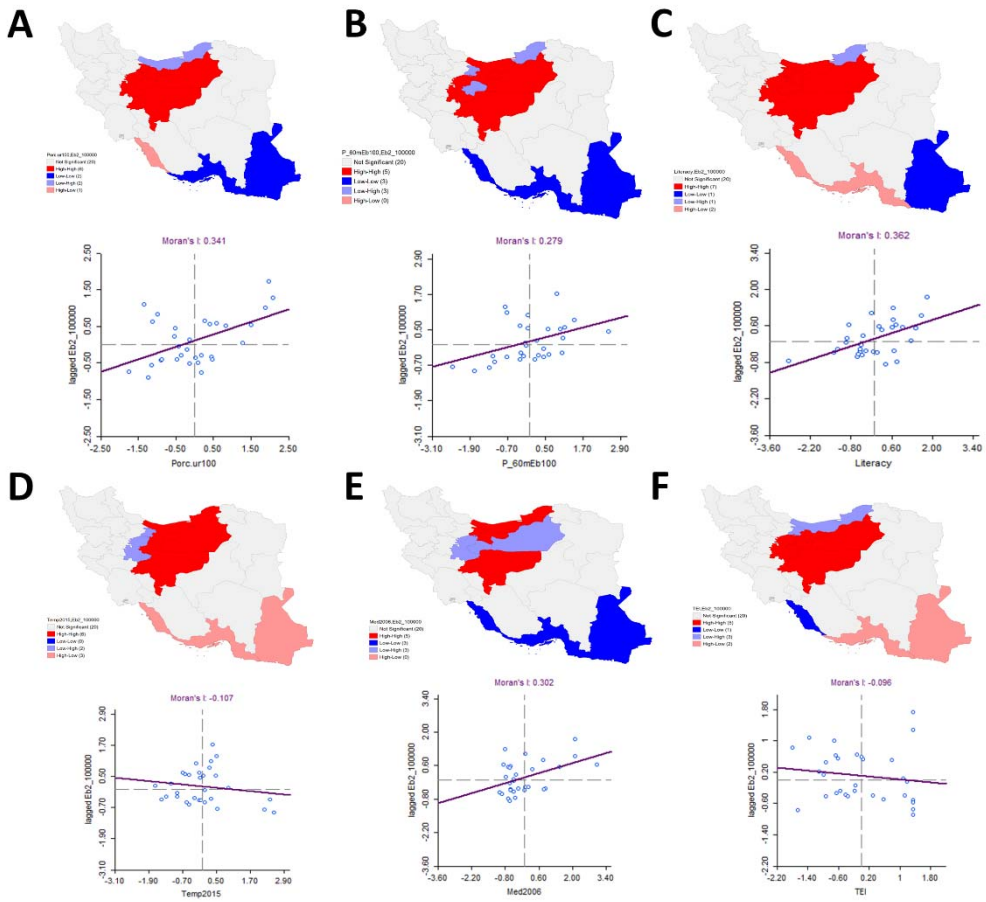
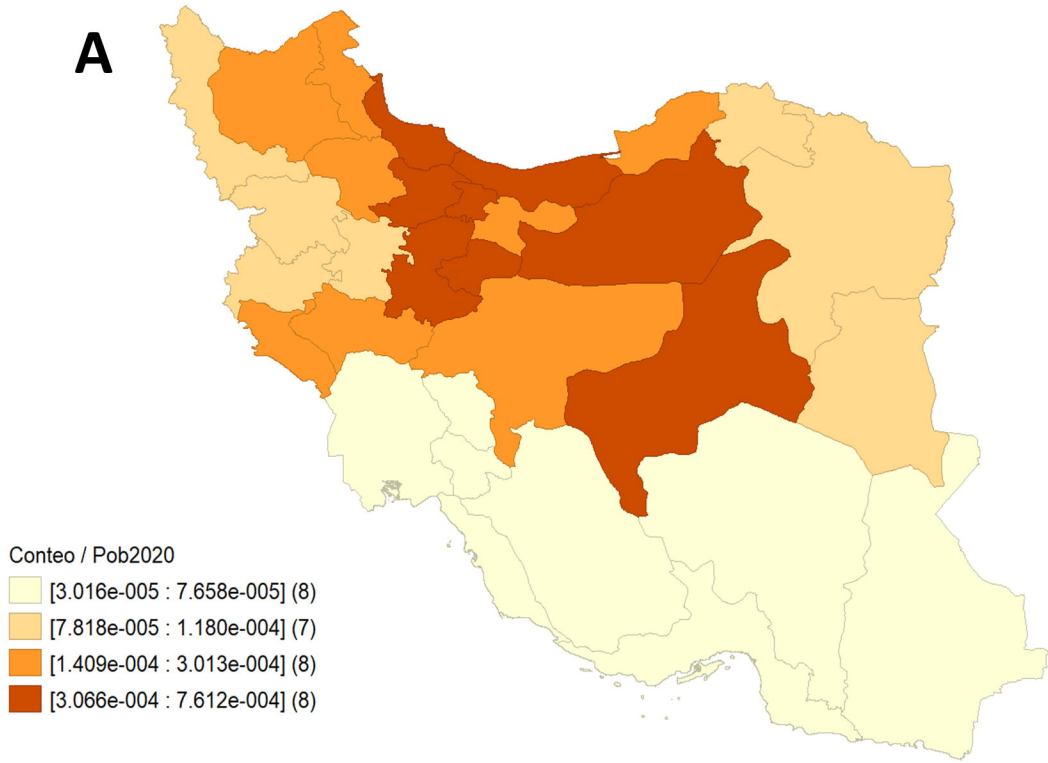
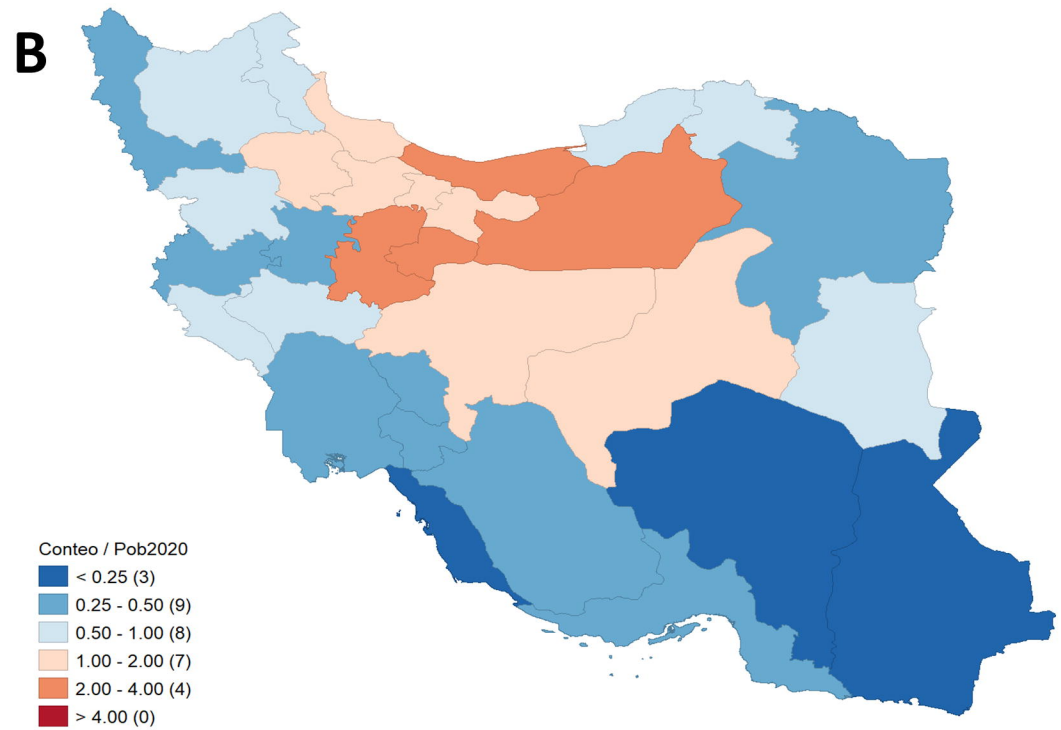


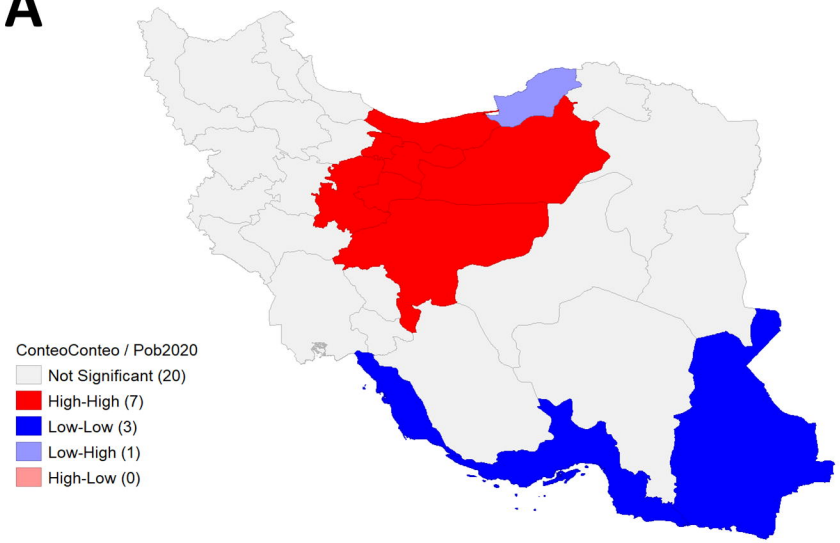
Figure 4. Bivariate LISA's significant spatial clustering between each of the significant variables in the spatial lag model and the rate of cases with COVID-19 smoothed through the empirical Bayes approach. The scatter plots associated with the variables vs. the spatially lagged smoothed rate of COVID-19 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 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).

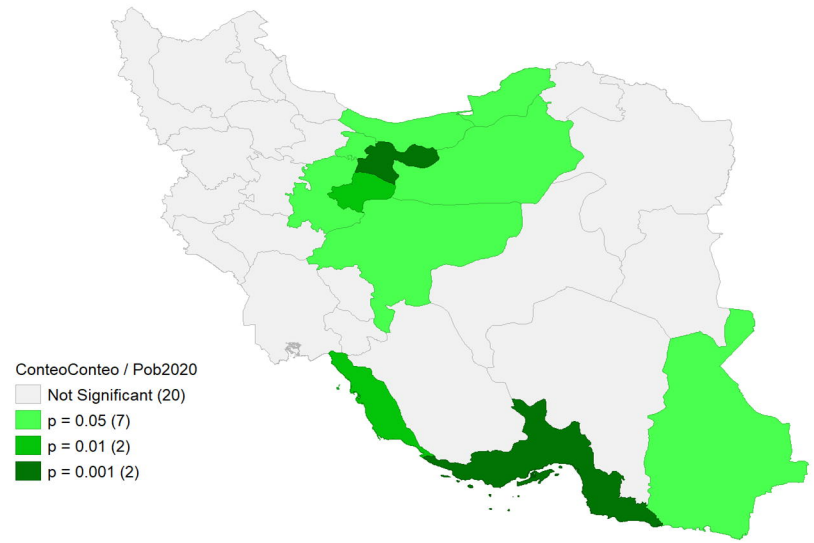

A
Conteo / Pob2020
[3.016e-005 : 7.658e-005] (8)
[7.818e-005 : 1.180e-004] (7)
[1.409e-004 : 3.013e-004] (8)
[3.066e-004 : 7.612e-004] (8)


B
Conteo / Pob2020
< 0.25 (3)
0.25 - 0.50 (9)
0.50 - 1.00 (8)
1.00 - 2.00 (7)
2.00 - 4.00 (4)
> 4.00 (0)

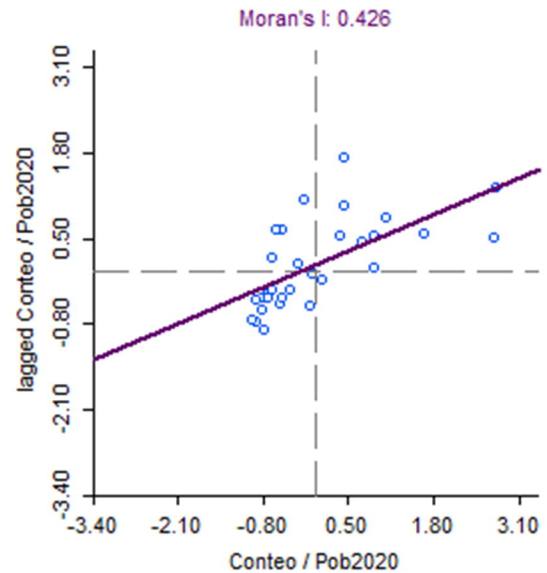
A

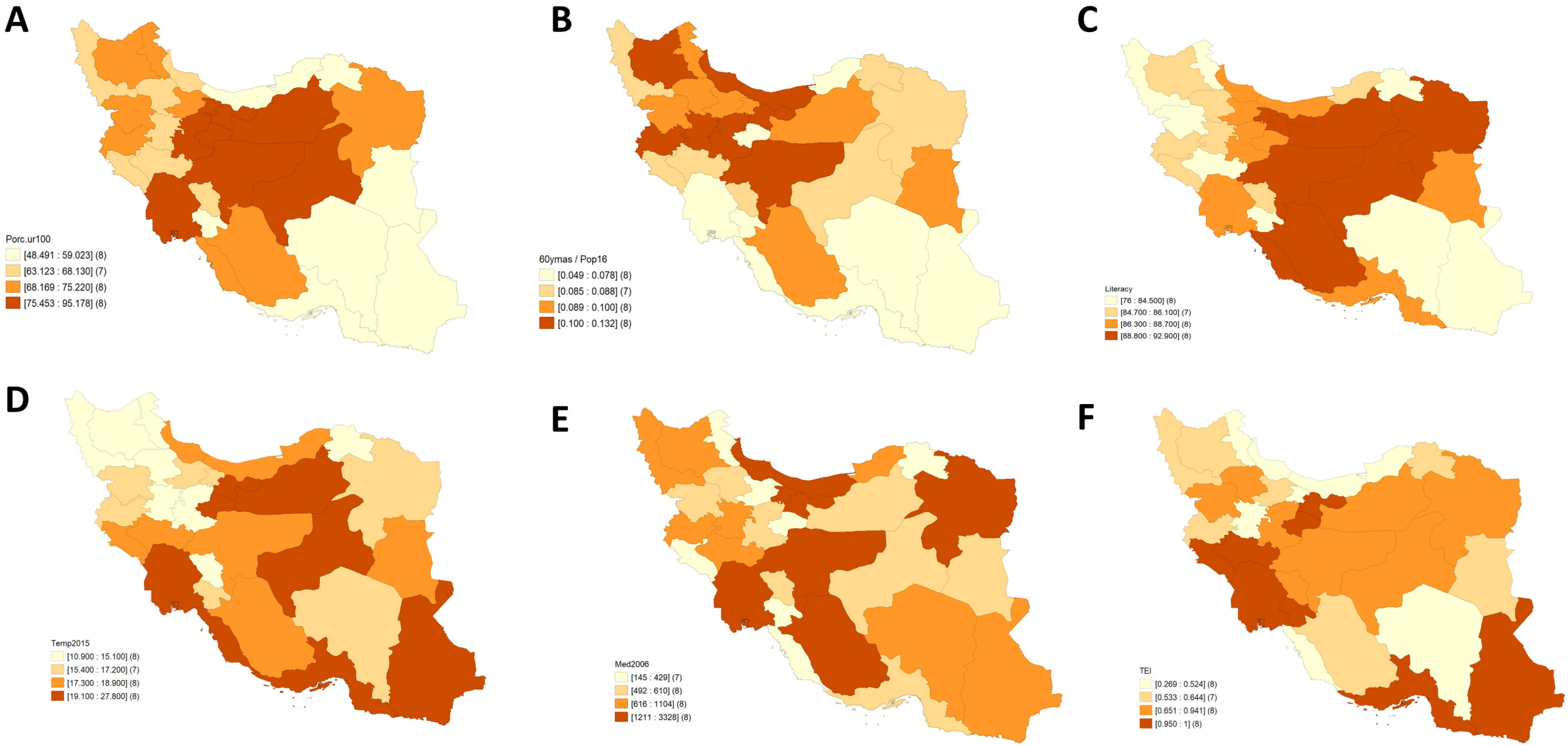

ConteoConteo / Pob2020
Not Significant (20)
High-High (7)
Low-Low (3)
Low-High (1)
High-Low (0)

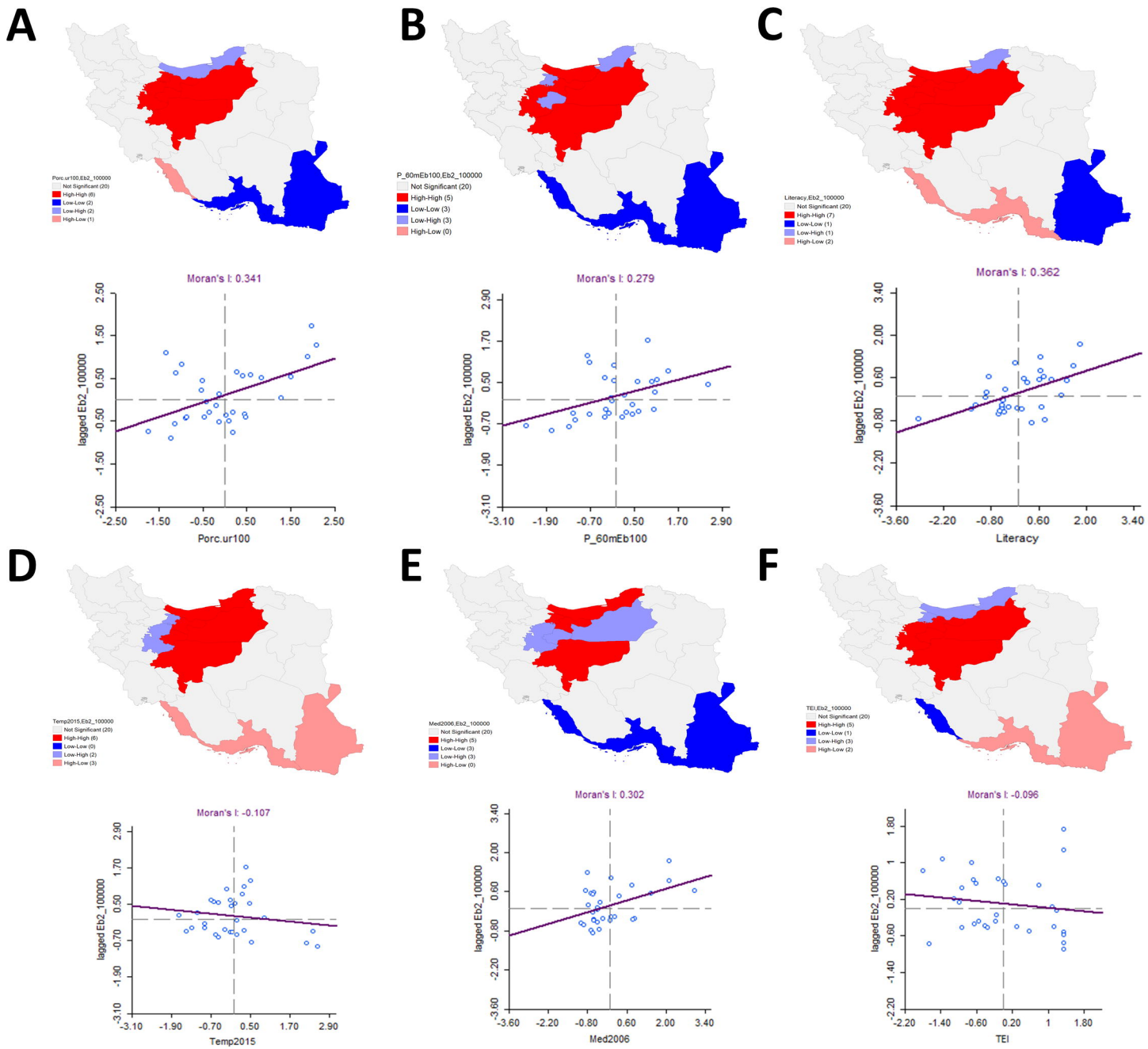
B


ConteoConteo / Pob2020
Not Significant (20)
p = 0.05 (7)
p = 0.01 (2)
p = 0.001 (2)

C


Moran's I: 0.426
lagged Conteo / Pob2020
Conteo / Pob2020
-3.40
-2.10
-0.80
0.50
1.80
3.10


A
Porc.ur100
[48.491 : 59.023] (8)
[63.123 : 68.130] (7)
[68.169 : 75.220] (8)
[75.453 : 95.178] (8)
B
60ymas / Pop16
[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
Temp2015
[10.900 : 15.100] (8)
[15.400 : 17.200] (7)
[17.300 : 18.900] (8)
[19.100 : 27.800] (8)
E
Med2006
[145 : 429] (7)
[492 : 610] (8)
[616 : 1104] (8)
[1211 : 3328] (8)
F
TEI
[0.269 : 0.524] (8)
[0.533 : 0.644] (7)
[0.651 : 0.941] (8)
[0.950 : 1] (8)


A
Porc.ur100,Eb2_100000
Not Significant (20)
High-High (6)
Low-Low (2)
Low-High (2)
High-Low (1)
Moran's I: 0.341
lagged Eb2_100000
Porc.ur100
B
P_60mEb100,Eb2_100000
Not Significant (20)
High-High (5)
Low-Low (3)
Low-High (3)
High-Low (0)
Moran's I: 0.279
lagged Eb2_100000
P_60mEb100
C
Literacy,Eb2_100000
Not Significant (20)
High-High (7)
Low-Low (1)
Low-High (1)
High-Low (2)
Moran's I: 0.362
lagged Eb2_100000
Literacy
D
Temp2015,Eb2_100000
Not Significant (20)
High-High (6)
Low-Low (0)
Low-High (2)
High-Low (3)
Moran's I: -0.107
lagged Eb2_100000
Temp2015
E
Med2006,Eb2_100000
Not Significant (20)
High-High (5)
Low-Low (3)
Low-High (3)
High-Low (0)
Moran's I: 0.302
lagged Eb2_100000
Med2006
F
TEI,Eb2_100000
Not Significant (20)
High-High (5)
Low-Low (1)
Low-High (3)
High-Low (2)
Moran's I: -0.096
lagged Eb2_100000
TEI