

# Spatial analysis of risk of morbidity and mortality by COVID-19 in Europe and the Mediterranean in the year 2020

JESÚS ENRIQUE ANDRADES-GRASSI<sup>1</sup> | LEDYZ CUESTA-HERRERA<sup>2</sup> |  
GUILLERMO BIANCHI-PÉREZ<sup>1</sup> | HILDA CRISTINA GRASSI<sup>1</sup> |  
JUAN YGNACIO LÓPEZ-HERNÁNDEZ<sup>1</sup> | HUGO TORRES-MANTILLA<sup>3</sup>

Recibido: 14/06/2020 | Aceptado: 19/10/2020 | Publicado online: 27/10/2020

## Abstract

Disease mapping seeks to represent the risk of a disease. This paper focuses on the spatial analysis of risk for pandemic COVID-19 in Europe and the Mediterranean. Morbidity and mortality data for 54 countries in ratio format were used. Two hypotheses were considered, the first one is that the data are homogeneous and the second one is that the ratios are defined in a heterogeneous manner requiring the stratification on the basis of covariables and the methodology of Jenks' intervals. Spatial risk models were applied as well as methods for the representation of clusters. The results show that the best representation is obtained with the Poisson-Gamma Model under stratification. The variations in the ratios are due to the individual policies of each country for the management of the pandemic. The cluster analysis shows that there is a high mortality process in Eastern Europe. The behavior of the pandemic should be evaluated in the space-time process as well as in other heterogeneous and highly unequal regions.

---

Keywords: COVID-19; disease spatial risk; Europe; Mediterranean; disease clusters; morbidity; mortality

---

## Resumen

*Análisis espacial de riesgo de morbilidad y mortalidad por COVID-19 en Europa y el Mediterráneo en el año 2020*

El mapeo de enfermedades busca representar el riesgo de una enfermedad. El objetivo de este trabajo es hacer un análisis del riesgo para la pandemia de COVID-19 en Europa y el Mediterráneo. Se utilizaron los datos de morbilidad y mortalidad en formato de tasas de 54 países. Se aplicaron dos hipótesis, la primera es que los datos son homogéneos y la segunda es que las tasas son definidas de forma heterogénea por lo que se estratificó en base a covariables y la metodología de los intervalos de Jenks. Se aplicaron modelos espaciales de riesgos así como métodos de representación de clústers. Los resultados muestran que el modelo Poisson-Gamma bajo estratificación es el que mejor representa el proceso. Las variaciones de las tasas se deben a la heterogeneidad en las políticas individuales de cada país para el manejo de la pandemia. Los análisis clusters muestran que existe un fuerte proceso de mortalidad ubicado en Europa del Este. Debe evaluarse

---

1. Universidad de Los Andes, Venezuela. Correspondencia: [cristinagrassi@gmail.com](mailto:cristinagrassi@gmail.com)

2. Universidad Católica del Maule, Chile.

3. Universidad de Santander, Colombia.

el comportamiento del proceso de la pandemia en el espacio-tiempo así como en otras regiones heterogéneas y altamente desiguales.

---

Palabras clave: COVID-19; riesgo espacial para enfermedades; Europa; Mediterráneo; clústers de enfermedades; morbilidad y mortalidad

---

## 1. Introduction

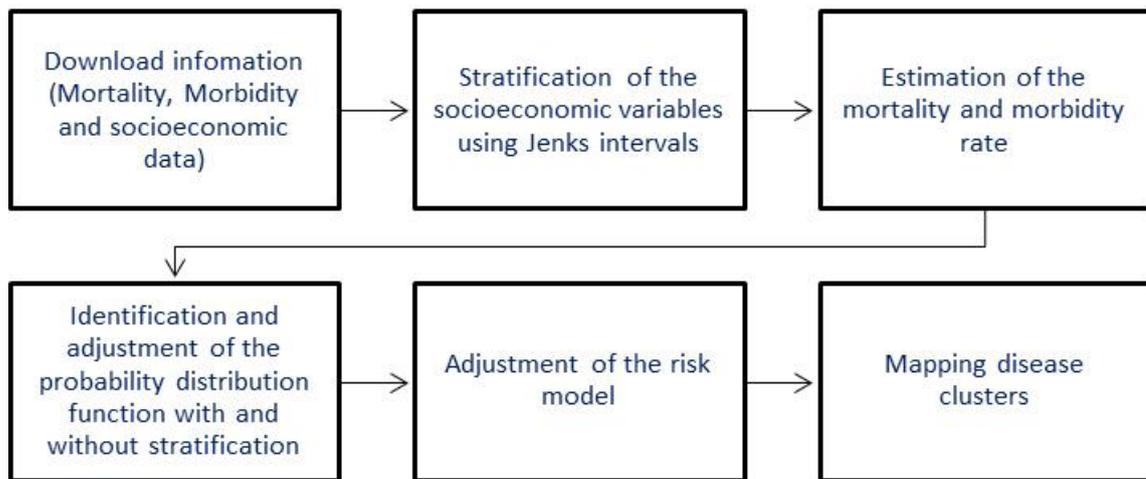
The mapping of disease incidence and prevalence has long been a part of public health, epidemiology and the study of disease in human populations (Koch, 2005). The aim of disease mapping is to provide a representation of the spatial distribution of the risk of a disease in the study area, which we will assume is divided into several non-overlapping smaller regions, the risk may reflect actual deaths due to the disease (mortality) or, if it is not fatal, it may reflect the number of people who suffer from the disease (morbidity) in a certain period of time for the population at risk (Bivand, Pebesma, & Gomez-Rubio, 2013). COVID-19 is a pandemic infection caused by the virus SARS-CoV-2.9 which was first detected as an epidemic disease of Wuhan 2019 causing pneumonia by coronavirus with high morbidity and mortality (Hui, Azhar, Madani, & et al, 2020). In this article we use the concept of the first law of geography which expresses that everything is related to everything else, but near things are more related than distant things (Tobler, 1970), and we focus on the challenge of obtaining reliable statistical estimates of COVID-19 disease risk based on counts of observed cases in countries in Europe and in the Mediterranean.

The spatial issue of COVID-19 was approached using Moran's Index as a statistics tool (Huang, Liu, & Ding, 2020), applying it to the spatial panel and showing that the infection by COVID-19 is space dependent and propagates mainly from the Province of Hubei in central China to the neighboring regions. Additionally, other authors (Chen et al., 2020) found that there is a high impact of the mobility of people on the propagation of the disease. In Europe it has been observed that differences in mortality by COVID-19 are due to unequal population structures in reference to age and age and sex (Kashnitsky & Aburto, 2020). The analysis of the decomposition of the general wage inequality in Europe shows that the quarantine measures and social distancing promote a double process of divergence: they increase the inequalities within the countries as well as among them (Palomino, Rodríguez, & Sebastian, 2020).

## 2. Methods

Figure 1 shows the general diagram of the methods as they were applied. The classical design-based solutions are often not viable for this type of problem because the sample sizes within each region required for desired levels of statistical precision, are often unavailable or inaccessible, in contrast, model-based approaches offer a mechanism to "borrow strength" across small areas in order to improve local estimates, resulting in the smoothing of extreme rates based on small local sample sizes. Such approaches are often expressed as mixed effects models (Gelfand et al., 2010). This explains why the observed number of cases alone gives no information on the risk of the disease, given that the cases are mainly distributed according to the underlying population. In order to obtain an estimate of the risk, the observed number of cases must be compared to an expected number of cases, thus we will denote by  $P_{ij}$  and  $O_{ij}$  the population and observed number of cases in region  $i$  and stratum  $j$  (Bivand, Pebesma, & Gomez-Rubio, 2013).

Figure 1. Diagram of the methods used in this work.



Summing over all strata we can get the total population and number of cases per area, which we will denote by  $P_i$  and  $O_i$ . Summing again over all the regions will give the totals which will be denoted by  $P_+$  and  $O_+$ . If  $P_i$  and  $O_i$  are already available, which is the simplest case, the expected number of cases in region can be calculated as  $E_i = P_{ir+}$  where  $r_+$  is the overall incidence ratio equal to  $O_+/P_+$ . Because the data are grouped in strata, we computed the ratio for each population stratum as  $r_j = \sum_i O_{ij} / \sum_i P_{ij}$ , this is called Standardized Mortality Ratio (SMR) and Standardized Incidence Ratio (SIR) in the case of morbidity, both considered as SMR/SIR. Thus, the expected number of cases in region is given by  $E_i = \sum_j P_{ij} r_j$ , this standardisation is also called *internal standardization* (Waller & Gotway, 2004).

54 countries were selected from Europe, North of Africa and the Near East. European micro-states were excluded because the SMR/SIR may cause distortions in the analysis. The mortality and morbidity data was obtained from the plugin HCMGIS from Qgis (QGIS.org, 2020) from the John Hopkins Institute of April 29, 2020 (John Hopkins Institute, 2020) and the R libraries *ctv* (Zeileis & Hornik, 2018), *maptools* (Bivand & Lewin-Koh, 2020), *rgdal* (Bivand, Keitt, & Rowlingson, 2020), *spdep* (Bivand R., 2020), *ncf* (Bjornstad, 2020), *pgirmess* (Giraudoux, 2018), *ape* (Paradis et al, 2020), *spatialeco* (Hothorn, Zeileis, Farebrother, & Cummins, 2020), *lmtree* (Zeileis & Hothorn, 2002), *regeos* (Bivand & Rundel, 2020), *splm* (Millo & Piras, 2018), *spgwr* (Bivand & Yu, Roger Bivand, Danlin Yu. 2020), *epitools* (Aragon, 2020), *DCluster* (Gómez-Rubio, Ferrándiz-Ferragud, & López-Quílez, 2015), *raster* (Hijmans, 2020) y *fitdistrplus* (Delignette-Muller, Dutang, & Siberchicot, 2020) and *classInt* (Bivand, Ono, Dunlap, Stigler, Denney, & Gómez, 2020). However, it is important to point out that the available COVID-19 data may be underestimated or overestimated due to the high number of asymptomatic cases and also due to the number of tests performed per day in each country; all this causes differences in the quality of the reports of the data by country (Modig, Ahlbom, & Matthews, 2020). For the modeling of the COVID-19 the basic data must include the population at risk and number of cases of mortality and morbidity in each area. Two hypotheses were considered. The first one is that the country's risk factor is homogeneous, which implies that there is no available variable that controls SMR/SIR and this index is the same for each country. In this case the estimation of SMR/SIR was  $r = \sum O / \sum P$ .

The second hypothesis is that SMR/SIR have spatial heterogeneity. In this case, in addition to the statistical phenomenon of spatial autocorrelation, another equally important statistical concept

must be included, spatial heterogeneity (HE), which refers to the variation in the relationships between variables in space (LeSage, 1999), that is, it is a phenomenon that is due to a real and substantive variation that evidences the existence and validity of the geographical context in the definition of social behavior (O’Loughlin & Anselin, 1992). The variables that may affect the risk of COVID-19 which were used in this work were obtained from data of the World Bank (The World Bank, 2020), American Cancer Society (American Cancer Society) and John Hopkins Institute (John Hopkins Institute, 2020), and are the following: 1) Population over 65 years (%) by country; 2) The Global Health Security Index by country; 3) Daily passenger dispatch by air and water; 4) Poverty (%) and 5) Cigarette consumption *per capita* per year and by country (Zhou et al., 2020); (Aitken et al., 2020); (Patel et al., 2020); (Correa-Martínez et al., 2020); (Zhao et al., 2020), see Table 1. This process was performed as a method to incorporate the effect of socio-economic variables on morbidity and mortality of COVID-19.

Table 1. Variables that may affect risk of COVID-19

Variables	Risk aspects	Protective aspects
Population over 65 years	Higher mortality reported from the first series of cases. Health conditions such as COPD*, diabetes, and coronary heart disease are frequent in this population (Zhou et al., 2020)	Older adults without these diseases may not be at increased risk.
Global Health Security Index**	Index summarizes existing policies and procedures in case of pandemic	Poor prediction of preparedness. Global Health Security Index (GHSI)** is predictive of COVID-19 burden, but in the opposite direction (Aitken et al., 2020)
Air and marine traffic	Easy spread of contagion by tourism and travelling (Correa-Martínez et al, 2020)	Better surveillance of imported cases
Poverty	Overcrowded housing, occupations without opportunities to work from house, public transport use and heightened stress by the economic situation could be risk factors. Health services, personal protection items and test availability differs (Patel et al., 2020)	Reduced travelling and tourism before the pandemic
Cigarette consumption per capita	Ongoing smoking history is attributed to the worse progression and outcome of COVID-19 (Zhao et al., 2020)	The cellular receptor for the virus has been reported to be decreased in smokers.

\*COPD: Chronic Obstructive Pulmonary Disease, \*\*GHSI, The Global Health Security Index is an evaluation across six items about the responsiveness of countries in the event of a pandemic, developed by the Johns Hopkins Center for Health Security, the Nuclear Threat Initiative (NTI) and the Economist Intelligence Unit (EIU) (Nuclear Threat Initiative, Johns Hopkins Center for Health Security). Besides the criticisms pointed out in the Table, we must mention the ecological bias or fallacy that can be considered a case within the problem of small area estimation in which the measurement scale of a phenomenon does not correspond to the scale where the real interactions occur between the variables. We assumed within-area variability is sufficiently small in order to ignore it until we have data at an appropriate scale (Wakefield, 2010).

Jenks’ interval technique (Jenks, 1967) was used as a method of univariant stratification which seeks to minimize each class’s average deviation from the class mean, while maximizing each class’s deviation from the means of the other groups. The optimization method is also known as the goodness-of-variance-fit (GVF) method and can be used as acceptance criteria. The other criterion that is frequently used is the Tabular Accuracy Index (TAI). The acceptance criteria for the methodology included natural intervals, minimum number of classes obtained with a GVF and a TAI which is higher than 0.90. Then for the analysis of the natural intervals, the classes were assigned for each variable/country according to the established criteria of the natural intervals. Finally, the strata for each variable was added so that  $S = E_{v_1} + E_{v_2} + \dots + E_{v_n}$ , then SMR/SIR were estimated by strata.

It is mandatory to use a probability distribution model to analyze risk. Therefore, in order to select the most suitable model, the first step was the identification and adjustment of the frequency distribution for mortality and morbidity, SMR/SIR, in both hypothesis (homogeneity and heterogeneity). Cullen and Frey's diagrams, (Cullen & Frey, 1999) were used for this purpose as a method for the identification of the closest distribution function and the method suggested by Venables and Ripley (Venables & Ripley, 2002) for the adjustment of the probability distribution model. Then the following Bayesian risk models were adjusted: Poisson-Gamma Model, Log-Normal Model, (Clayton & Kaldor, 1987) Marshall's Global EB Estimator (Marshall, 1991) and Marshall's Local EB Estimator (Bivand, Pebesma, & Gomez-Rubio, 2013). The graphic method suggested by Bivand et al (Bivand, Pebesma, & Gomez-Rubio, 2013) was used as a selection method, generating box-and-whisker plots for each model and for the real data as well as for an analysis of the performance of the models in terms of Root-Mean-Square Deviation (RMSE).

After performing the analysis of the presence of groups, the heterogeneity of the relative risks must be evaluated. Three global risk tests were performed on the expected values. The first one is a Chi squared because for each area we have the observed number of cases and the expected number has been calculated, the idea is to assess significant (global) differences among these two quantities, the statistical test is defined by (Bivand, Pebesma, & Gomez-Rubio, 2013):

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - \theta E_i)^2}{\theta E_i}$$

where  $\theta$  is global SMR/SIR, and, asymptotically, it follows a chi-square distribution with degrees of freedom. The second test is the one suggested by Potthoff and Whittinghill. (Potthoff & Whittinghill, 1966). This author proposed another test of homogeneity of the means of different Poisson distributed variables which can be used to test the homogeneity of the relative risks. The test statistic is given by:

$$PW = E + \sum \frac{O_i(O_i - 1)}{E_i}$$

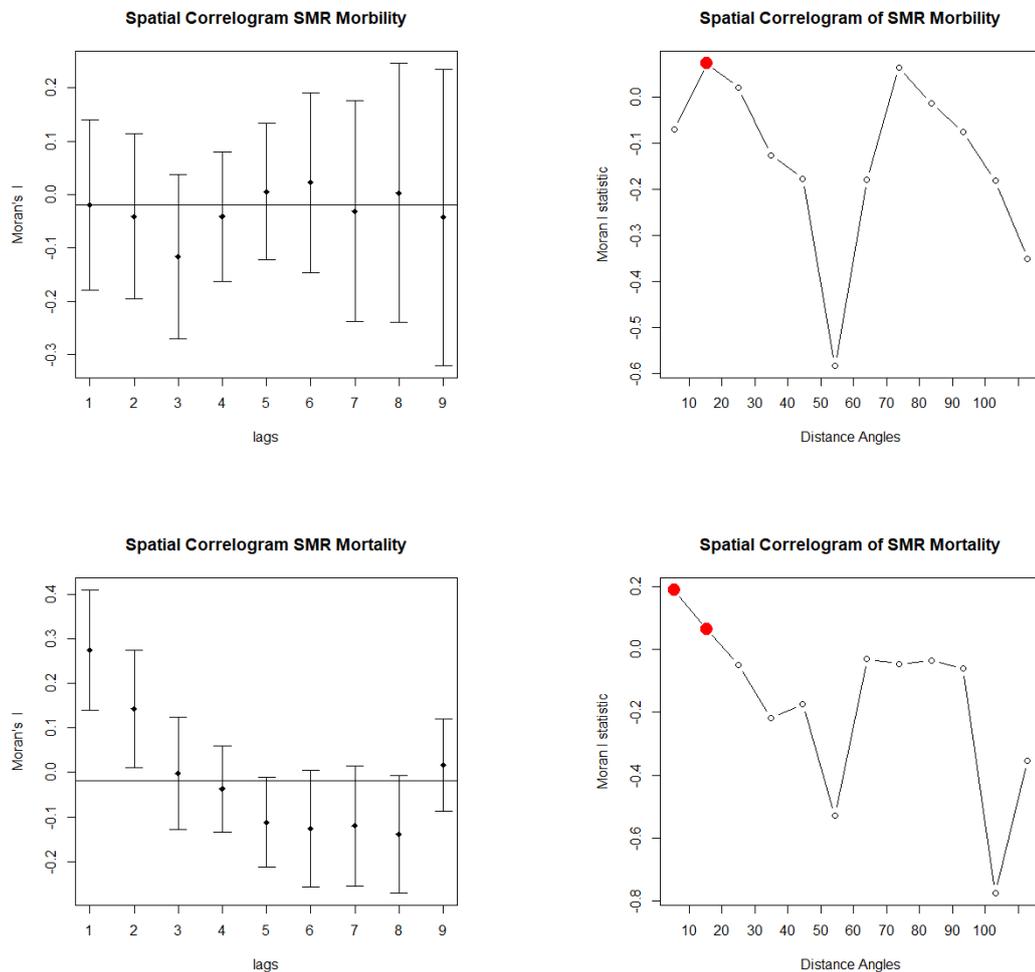
The alternative hypothesis of this test is that the  $O_i$  are distributed following a Negative Binomial distribution. Finally, the Moran test (Moran, 1950) which corresponds with a linear test was applied and also the Pearson coefficient correlation was applied to quantify the existence of spatial autocorrelation:

$$I = \frac{N}{S_0} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (x_i - \mu)(x_j - \mu)}{\sum_{i=1}^N (x_i - \mu)^2}$$

Here we apply Moran's I statistic to the SMR/SIR to account for the spatial distribution of the population. If we computed Moran's statistic for the  $O_i$  we could find spatial autocorrelation only due to the spatial distribution of the underlying population, because it is well known that the higher the population, the higher the number of cases (Bivand, Pebesma, & Gomez-Rubio, 2013).

However, the execution of this test requires the exploration of the spatial autocorrelation for the distance. Autocorrelation was explored with this method because the nature of global exchange as a result of long distance terrestrial, maritime and air travelling is very efficient especially in Europe. Furthermore, by allowing free circulation (especially in the countries of the European Union) the terrestrial distance is the main determinant in the mobility among countries rather than the number of terrestrial borders. For that purpose the spatial correlogram was estimated and the point where spatial autocorrelation is maximized was evaluated. In these terms, the spatial weighing matrix was selected. The exploration is presented in Figure 2 which shows that the autocorrelation is maximized at a distance of 20 degrees and at 3 spatial lags in the case of Morbidity SMR and at 20 degrees and 2 spatial lags in the case of Mortality SMR. With this, a Queen type second order matrix was used for the estimation of Moran's I.

Figure 2. Spatial Correlograms of SMR Mortality and SIR Morbidity for COVID-19.



So far we have considered methods that only assess the presence of heterogeneity of risks in the study area and give a general evaluation of the presence of clusters. In order to detect the actual location of the clusters present in the area, a different approach must be followed (Bivand, Pebesma, & Gomez-Rubio, 2013). A useful family of methods that can help in this purpose are scan statistics, Kulldorff and Nagarwalla (Kulldorff & Nagarwalla, 1995) developed a new test for the detection of clusters based on a window of variable size that only considers the most likely

cluster around a given region. Kulldorf's statistic works with the regions within a given circular window and the overall relative risk in the regions inside the window is compared to that of the regions outside the window. The null hypothesis, of no clustering, is that the two relative risks are equal while the alternative hypothesis (clustering) is that the relative risk inside the window is higher. This is resolved by means of a likelihood ratio test, which has two main advantages. First, the most likely cluster can be detected as the window with the highest value of the likelihood ratio and, second, there is no need to correct the p-value because the simulations for different centers are independent (Waller and Gotway, 2004, p. 220). Two variants of the test were applied, first an approximation where the strata do not have much weight (Binomial) and second, an approximation of a Poisson distribution was used where the strata possess a significant weight (Kulldorff & Nagarwalla, 1995).

### 3. Results and Discussion

The results of the strata obtained with Jenks covariable intervals show that Poverty has 7 strata, Population over 65 years (%) by country has 8 strata, Mortality has 11 strata, Daily passenger dispatch by air and water has 7 strata, Cigarette consumption per capita per year has 10 strata and The Global Health Security Index has 11 strata. It should be noted that in order to meet the previously established criteria for GVF and TAI, the number of classes is relatively high, indicating that the data is highly variable.

Table 2 shows the descriptive statistics for the estimated SMR/SIR and the Goodness of Fit Statistics of the distribution models. The goodness-of-fit statistics aims to measure the distance between the fitted parametric distribution and the empirical distribution (Delignette-Muller & Dutang, 2015), with the purpose of orienting the selection of either the homogeneity or heterogeneity model according with the least statistical value. The statistics favor the selection of the homogeneity model for SIR and a model with strata for SMR.

Table 2. Descriptive statistics and Goodness of Fit Statistics for SMR/SIR

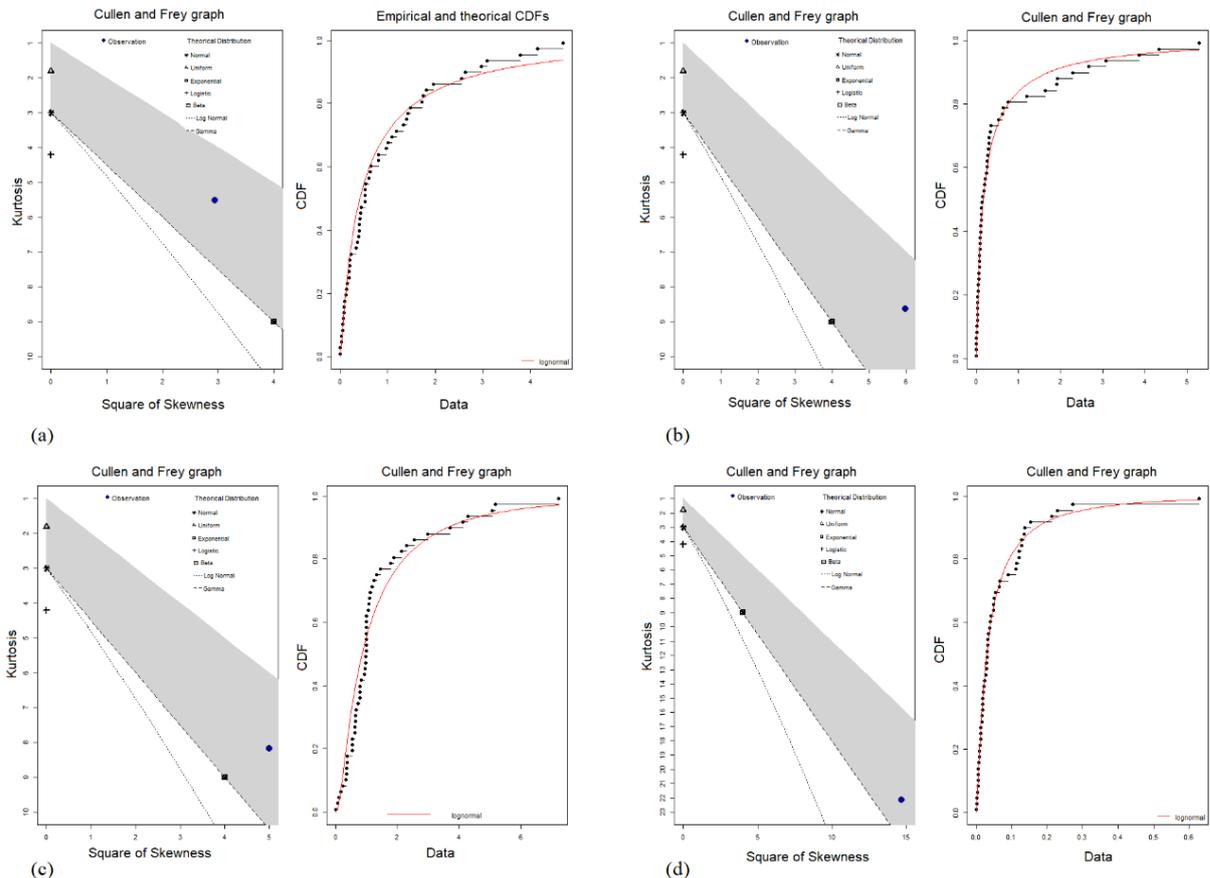
Method	SIR		SMR	
	Homogeneity	Heterogeneity	Homogeneity	Heterogeneity
Mean	0.95	1.38	0.66	0.065
St. Dev.	1.12	1.44	1.18	0.09
Skw.	1.71	2.23	2.44	3.83
Kurt.	5.51	8.18	8.62	22.16
K-S	0.11	0.15	0.07	0.08
C-VM	0.09	0.29	0.04	0.04
A-D	0.65	1.57	0.35	0.28
AIC	111.9	153.75	25.7	-189.67
BIC	115.88	157.73	29.68	-185.69

Kolmogorov-Smirnov (K-S); Cramer-von Mises; Anderson-Darling (A-D); Akaike Information Criterion (AIC); Bayesian Information Criterion (BIC).

Figure 3 shows the results of the identification of the probability distribution and the adjusted model, with a type Log-Normal adjusted distribution for all cases. It should be noted that in the

process of adjustment of the probability of the Log-Normal distribution, the statistics suggested by Delignette-Muller y Dutang (Delignette-Muller & Dutang, 2015) for the analysis of the goodness of fit of the model indicate a better adjustment when the heterogeneity is defined by the strata because in general, the adjusted models have less entropy (AIC and BIC).

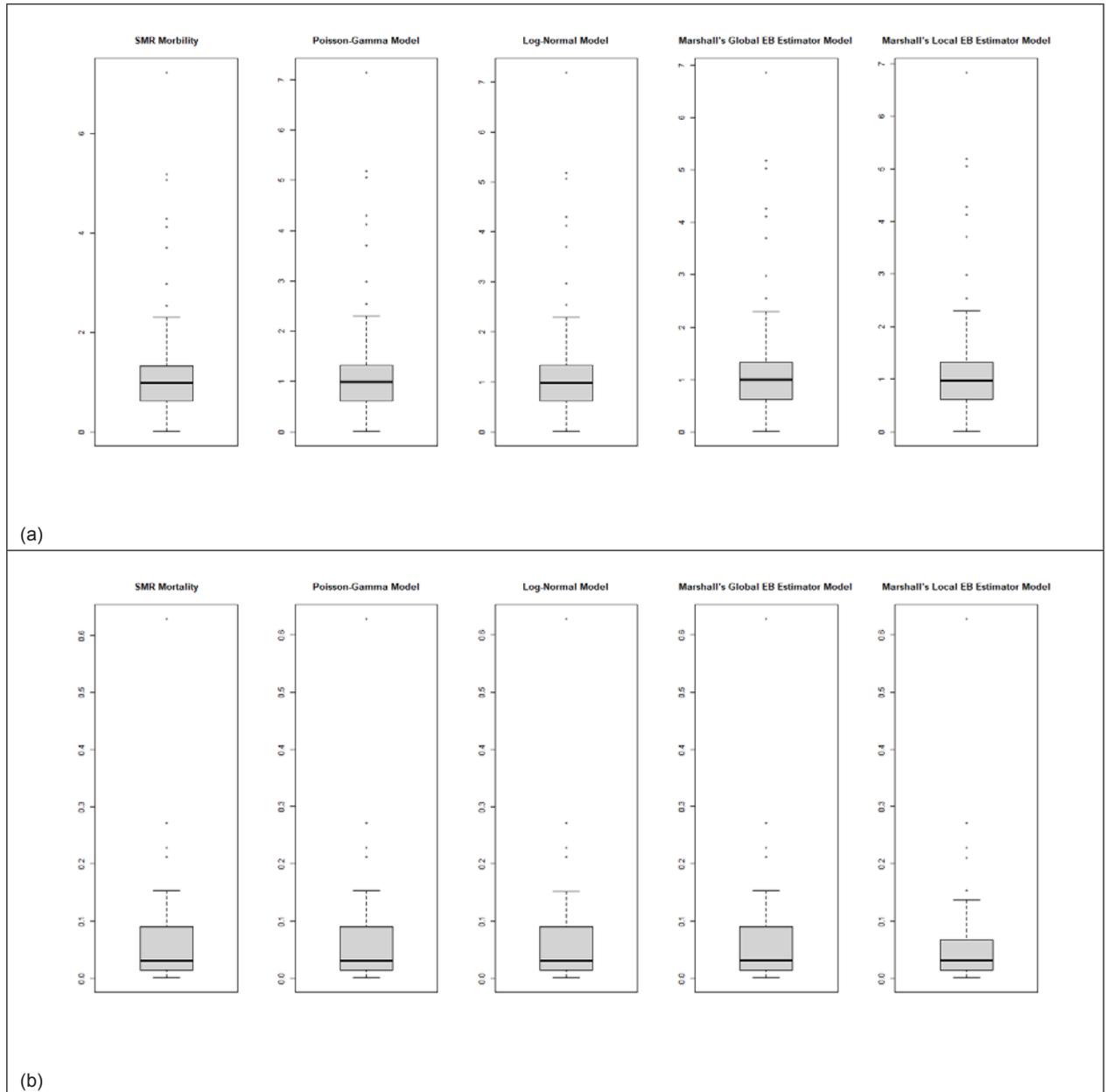
Figure 3. Cullen and Frey Graph, Histogram and theoretical densities, Empirical and theoretical CDFs



a) Morbidity with Homogeneity; b) Mortality with Homogeneity; c) Morbidity with Heterogeneity; d) Mortality with Heterogeneity.

Figure 4 shows the box diagrams of the adjusted and observed models. The models Poisson-Gamma, Marshall's Global EB Estimator and Marshall's Local EB, have a similar performance and the smaller RMSE was obtained with the Poisson-Gamma Model both in Morbidity (0.01180) and in Mortality (0.00134).

Figure 4. Risk Box Plot Models adjusted for CoViD-19



a) Morbidity Risk; b) Mortality Risk.

This model implies that one key issue is that for this distribution the mean and the variance of  $O_i$  are supposed to be the same. It is often the case that data are “over-dispersed”, meaning that the variance of the data is higher than their mean and the statistical model needs to be expanded. This formulation is known as the Poisson-Gamma (PG) model and is structured following a two-level model (Bivand, Pebesma, & Gomez-Rubio, 2013):

$$O_i | \theta_i, E_i \sim Po(\theta_i E_i)$$

$$\theta_i \sim Ga(\nu, \alpha)$$

In this model, we also consider the relative risk  $\theta_i$  as a random variable which is drawn from a Gamma distribution with mean  $\nu/\alpha$  and variance  $\nu/\alpha^2$ . Note that now the distribution of  $O_i$  is conditioned on the value of  $\theta_i$ . The posterior expectation of  $\theta_i$  is:  $E[O_i|\theta_i, E_i] = \nu + O_i/\alpha + E_i$ . Which can also be expressed as a compromise between the prior mean of the relative risks and  $SMR_i$  or  $SIR_p$ , so that this is a shrinkage estimator (Bivand, Pebesma, & Gomez-Rubio, 2013):

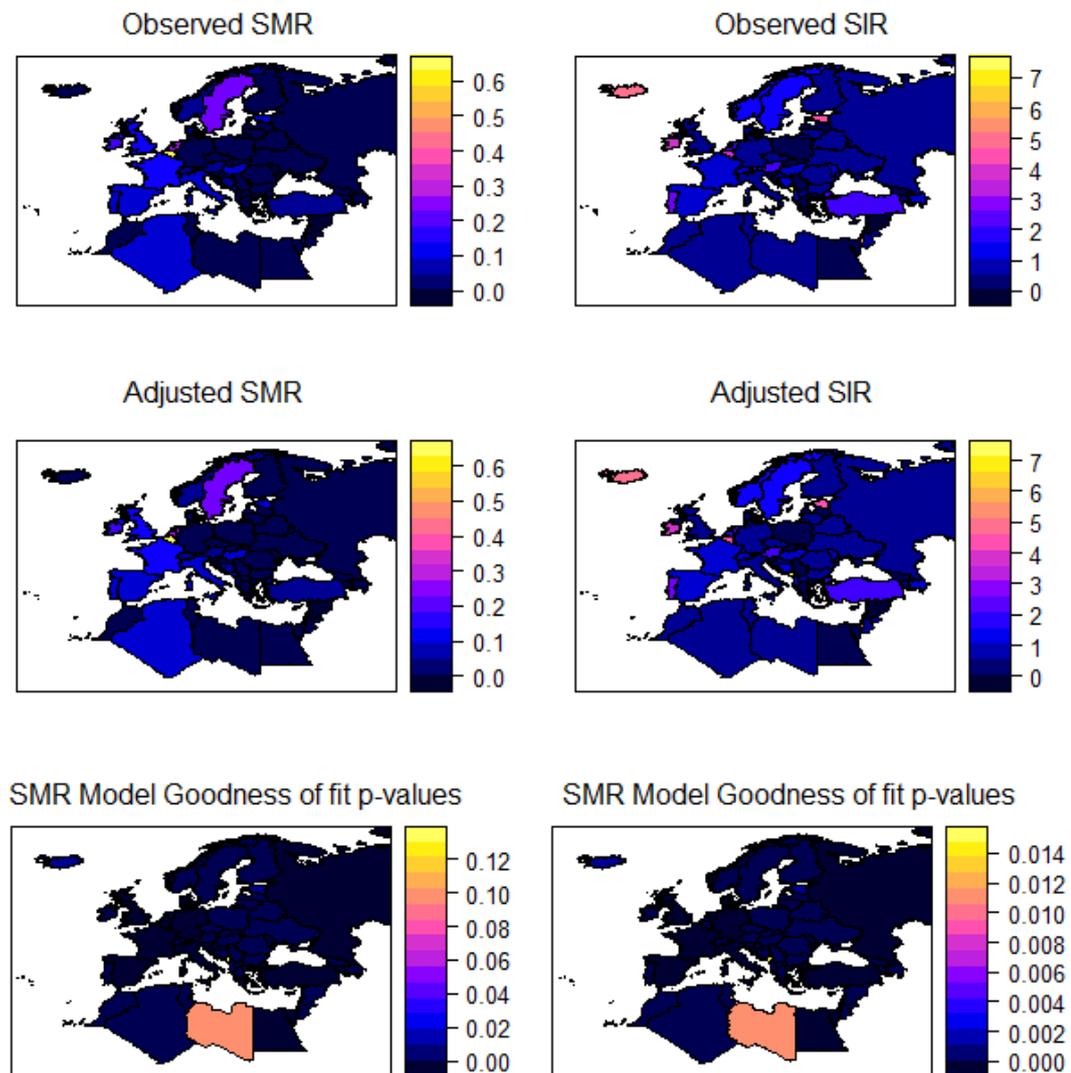
$$E[O_i|\theta_i, E_i] = \frac{E_i}{\alpha + E_i} SMR_i + \left(1 - \frac{E_i}{\alpha + E_i}\right) \frac{\nu}{\alpha}$$

Two issues should be noted from this estimator. First of all, when  $E_i$  is small, as often happens in low population areas, a small variation in  $O_i$  can produce dramatic changes in the value of  $SMR_i$ . For this reason, according to the previous expectation, the  $SMR_i$  will have a low weight, as compared to that of the prior mean. Secondly, information is borrowed from all the areas in order to construct the posterior estimates given that  $\nu$  and  $\alpha$  are the same for every region. This concept of borrowing strength can be modified and extended to take into account a different set of areas or neighbours (Bivand, Pebesma, & Gomez-Rubio, 2013).

Probability maps are a convenient way of representing the significance of the observed values (Figure 5), these maps show the probability of a value being higher than the observed data according to the assumption we have made about the model. We noted that the Poisson-Gamma model was more appropriate in this case due to over-dispersion, and we should try to make inferences based on this model. As expected, the p-values for the Poisson-Gamma model are higher because more variability is permitted. Nevertheless, there are still countries of high risk in the case of mortality and morbidity such as Libya and Montenegro. It should be considered that this model represents the study area except these two countries that had a decreased response of their health care system due to war conflicts or due to a lack of effectivity in the stratification of these spatial units.

It is important to point out that it seems that the problem pointed by Bivand et al. (Bivand, Pebesma, & Gomez-Rubio, 2013) is not present in this case because the data of the population under study does not include areas of low population such as micro-states, indicating that a small variation in  $O_i$  may produce dramatic changes in the value of  $SMR_i$ . The risk predictions estimated from this model are shown in Figure 6 with significantly high rates of morbidity and mortality in countries such as Iceland, Ireland, Portugal, Belgium, Israel and Estonia. Due to this variability it is important to consider the state policies that each country has for pandemic control, especially in countries of the European Union which have relatively homogeneous standards in their health care systems. Thus each country's individual state policies during the pandemic must generate subtle changes in  $SMR/SIR$ . The rapid increase in the cases of COVID-19 has threatened the health care systems in many countries. As a result, the affected countries had to consider public health strategies achieved through non-pharmaceutical strategies. In general, these strategies have been classified in two categories: i) "mitigation" with the purpose of obtaining herd immunity allowing the propagation of virus SARS-CoV-2 among the population and mitigating the load of the disease, and ii) "suppression" with the purpose of reducing drastically the rate of endogenous transmission within the population (Brett & Rohani, 2020).

Figure 5. Poisson-Gamma model response for Morbidity and Mortality of Covid-19.



United Kingdom and countries such as Sweden are promoting “herd immunity” as opposed to other European countries. Sweden has avoided “suppression” measures. For Sweden the first concern was if the population or the economy would be able to bear a long shut down or cycles of opening and shutting down. The second concern was the impact on the no-COVID cases in a scenario in which the population was locked down while the health care systems had to consider COVID-19 cases as the priority. This stands out in places such as United Kingdom where health specialties such as cancer, heart and organ transplant have reached a difficult situation (Orlowski & Goldsmith, 2020).

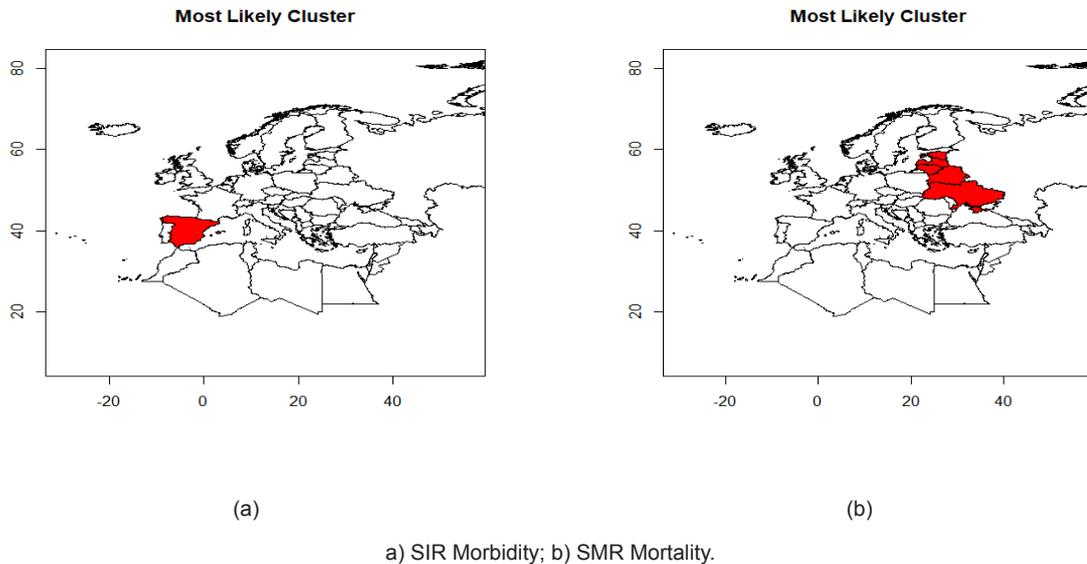
These examples show how the country’s individual state policies play a role in the variations of risk assessment for morbidity and mortality for COVID-19 in spite of the fact that the health care systems of the European countries are relatively homogeneous, possess high standards and have an important ability to respond to situations such as this pandemic. Based on these ideas, it is important to analyze the state policies in the short, medium and long term in regions where

the ability to respond to a pandemic is extremely heterogeneous. An example would be the Latin American countries which comprise the most unequal region of the world (CEPAL, 2016) and where each country's ability to respond is extremely heterogeneous. It could be expected that a small number of COVID-19 cases could cause a collapse of the health care system which may not be appropriate enough for the management of the disease.

However, these results have to be analyzed with caution because it is possible that not all the cases are reported as it is explained in the following: 1) There is a great deal of asymptomatic cases which are not diagnosed; 2) The number of tests carried out by country by day are not the same; 3) The interpretation of the data by country may be highly heterogeneous which would explain the results obtained for Libya and Montenegro; 4) The pandemic has not reached an end up to the time of the presentation of this paper. Therefore, in this research we must assume that the data presented by country is only approximately real and that the Standardized Ratio for Mortality as well as for Morbidity may be underestimated and invariant in time. Therefore it is not possible to distinguish if the statistical model is effectively estimating the data for SIR and SMR or the fluctuations of the quality of the data from each country. There are differences on how the cause of death is determined, on the strategies for the tests, what the tests are measuring or detecting and the time for the presentation of reports in each country (Modig, Ahlbom, & Matthews, 2020) (Okell *et al.*, 2020) .

The results of the tests of Homogeneity of the Relative Risks reject the null hypothesis of the homogeneity of the relative risk of COVID-19 (NS 5%), with the exception of Moran's I for the relative risk for morbidity. It can be concluded that there does not exist a cluster defined by morbidity (p-value 0.925), this is in agreement with the characteristics of a pandemic which affects all levels and strata of the society (Porta M, 2008). The mortality ratio is also in agreement with the expected behavior because the process for this variable must be highly influenced by each country's ability to respond, especially in the health care system which was quantified by the Global Health Index (John Hopkins Institute, 2020). Kulldorff's test (Figure 6), shows that in the case of SIR, only Spain is a significant cluster this is related with the results obtained with Moran's I for Morbidity SIR in which the process was characterized as random. It is surprising that countries like Ukraine, Belarus, Latvia, Lithuania and Estonia comprise a cluster of SMR according to Kulldorff's test, maybe due to state policies, seasonal climate changes or any other factor not considered. It is important to point out that in the analysis of Mortality and Morbidity SMR/SIR, Italy nor the United Kingdom appear as countries with high numbers of deaths by COVID-19. The results of Italy may possibly be due to the fact that the pandemic was centered in the region of Lombardy and its epidemiologic focus was relatively regional, another reason is that possibly the Mortality and Morbidity SMR/SIR are diluted in the country's spatial unit.

Figure 6. Results of the Kulldorff test



#### 4. Conclusions and Recommendations

This analysis presents a preliminary approximation of the behavior of mortality and morbidity risk of COVID-19 in Europe and the Mediterranean in a ratio format. The conclusion is that the risk model is an approximately global process type Poisson-Gamma, expected in a pandemic of this nature, which responds in a relatively homogeneous manner in Europe because the health care systems of those countries are relatively homogeneous, have high standards and an elevated response capacity for epidemics. The variations that were obtained are due to the individual policies of each country's government in the short term. It is also important to point out that up to the time of the presentation of these results the pandemic has not ended, so the Standardized Ratio for Morbidity and Mortality may change. Furthermore, it must be assumed that temporarily, the statistic risk process is invariant in time, so there are no temporal changes in the probability distribution. Thus it is mandatory to perform a spatio-temporal monitoring using risk models for the dynamics of behavior of the pandemic in the future. This will give indications on the ability of response of each country in time.

Can we predict the behavior of the risk models for COVID-19 in regions with great inequality? It is expected that more complex estimations will be required such as the hierarchical Bayesian disease risk models because there are countries with heterogeneous long term policies and although in the short term the response may be similar, the health care system may collapse with even a small number of COVID-19 cases. It is recommended that the process should be evaluated in time and that a script be made in a statistical language so that it automizes the processes as the morbidity and mortality data become available.

## 5. References

- John Hopkins Institute (2020). Retrieved from [https://www.google.com/url?sa=t&source=web&rct=j&url=https://coronavirus.jhu.edu/map.html&ved=2ahUKEwjRrPmmjufpAhWumuAKHaptB9sQFjACegQIAxAH&usg=AOvVaw18d-sqMa4OtFiUgMW\\_Xtnb](https://www.google.com/url?sa=t&source=web&rct=j&url=https://coronavirus.jhu.edu/map.html&ved=2ahUKEwjRrPmmjufpAhWumuAKHaptB9sQFjACegQIAxAH&usg=AOvVaw18d-sqMa4OtFiUgMW_Xtnb): <https://coronavirus.jhu.edu/map.html>
- The World Bank (2020). Retrieved from <https://www.google.com/url?sa=t&source=web&rct=j&url=https://data.worldbank.org/&ved=2ahUKEwiU4dzqjefpAhVRnOAKHVeYAcMQFjAAegQIBxAD&usg=AOvVaw1NoaWr-8xIJmHBEorHiFSZ>
- Aitken, T., Chin, K. L., Liew, D., & Ofori-Asenso, R. (2020). Rethinking pandemic preparation: Global Health Security Index (GHSI) is predictive of COVID-19 burden, but in the opposite direction. *Journal of Infection*, 81(2): 318–356. doi: 10.1016/j.jinf.2020.05.001
- American Cancer Society. In *The Tobacco Atlas*. Retrieved from: <https://tobaccoatlas.org/topic/consumption/>
- Aragon, T. J. (2020). *Package 'epitools'*. Retrieved from <https://cran.r-project.org/web/packages/epitools/index.html>
- Bivand, R. (2020). *Package 'spdep'*. Retrieved from <https://cran.r-project.org/web/packages/spdep/index.html>
- Bivand, R. S., Pebesma, E. J., & Gomez-Rubio, V. (2013). *Applied spatial data analysis with R*. New York: Springer.
- Bivand, R., & Lewin-Koh, N. (2020). *Package 'maptools'*. Retrieved from <https://cran.r-project.org/web/packages/maptools/index.html>
- Bivand, R., & Rundel, C. (2020). *Package 'rgeos'*. Retrieved from <https://cran.r-project.org/web/packages/rgeos/index.html>
- Bivand, R., & Yu, D. (2020). *Package 'spgwr'*. Retrieved from <https://cran.r-project.org/web/packages/spgwr/index.html>
- Bivand, R., Keitt, T., & Rowlingson, B. (2020). *Package 'rgdal'*. Retrieved from <http://cran.r-project.org/web/packages/rgdal/index.html>
- Bivand, R., Ono, H., Dunlap, R., Stigler, M., Denney, B., & Gómez, D. H. (2020). *Package 'classInt'*. Retrieved from <https://cran.r-project.org/web/packages/classInt/index.html>
- Bjornstad, O. N. (2020). *Package 'ncf'*. Retrieved from <https://cran.r-project.org/web/packages/ncf/index.html>
- Brett, T., & Rohani, P. (2020). COVID-19 herd immunity strategies: walking an elusive and dangerous tightrope. *medRxiv*. doi: 10.1101/2020.04.29.20082065
- CEPAL. (2016). *La Matriz de la Desigualdad Social en América Latina*. Retrieved from <https://www.cepal.org/es/publicaciones/40668-la-matriz-la-desigualdad-social-america-latina>
- Chen, H., Chen, Y., Lian, Z., Wen, L., Sun, B., Wang, P., et al. (2020). Correlation between the migration scale index and the number of new confirmed coronavirus disease 2019 cases in China. *Epidemiology & Infection*, 1-11. doi: 10.1017/S0950268820001119
- Clayton, D., & Kaldor, J. (1987). Empirical Bayes Estimates of Age-standardized Relative Risks for Use in Disease Mapping. *Biometrics*, 43(3), 671-681.
- Correa-Martínez, C. L., Kampmeier, S., Kumpers, P., Schwierzeck, V., Marc Hennies 4, W. H., Hennies, M., et al. (2020). A pandemic in times of global tourism: superspreading and exportation of COVID-19 cases from a ski area in Austria. *Journal of Clinical Microbiology*, 58(6). doi: 10.1128/JCM.00588-20
- Cullen, A. C., & Frey, H. C. (1999). *Probabilistic techniques in exposure assessment: a handbook for dealing with variability and uncertainty in models and inputs*. New York, U.S.A.: Springer Science & Business Media.
- Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, 64(4), 1-34.
- Delignette-Muller, M.-L., Dutang, C., & Siberchicot, A. (2020). *Package 'fitdistrplus'*. Retrieved from <https://cran.r-project.org/web/packages/fitdistrplus/index.html>
- Gelfand, A. E., Diggle, P., Guttorp, P., & Fuentes, M. (2010). *Handbook of spatial statistics*. Boca Raton, Florida, U.S.A.: CRC Press.
- Giraudoux, P. (2018). *Package 'pgirmess'*. Retrieved from <https://cran.r-project.org/web/packages/pgirmess/index.html>
- Gómez-Rubio, V., Ferrándiz-Ferragud, J., & López-Quílez, A. (2015). *Package 'DCluster'*. Retrieved from <https://cran.r-project.org/web/packages/DCluster/index.html>

- Hijmans, R. J. (2020). *Package 'raster'*. Retrieved from <https://cran.r-project.org/web/packages/raster/index.html>
- Hothorn, T., Zeileis, A., Farebrother, R. W., & Cummins, C. (2020). *Package 'spatialEco'.x*. Retrieved from <https://cran.r-project.org/web/packages/spatialEco/index.htm>
- Huang, R., Liu, M., & Ding, Y. (2020). Spatial-temporal distribution of COVID-19 in China and its prediction: A data-driven modeling analysis. *The Journal of Infection in Developing Countries*, 14(03), 246-253.
- Hui, D. S., I Azhar, E., Madani, T. A., Ntoumi, F., Kock, R., Dar, O., Ippolito, G., Mchugh, T. D., Memish, Z. A., Drosten, C., Zumla, A., & Petersen, E. (2020). The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health — The latest 2019 novel coronavirus outbreak in Wuhan, China. *International Journal of Infectious Diseases*, 91, 264-266. <https://doi.org/10.1016/j.ijid.2020.01.009>
- Jenks, G. F. (1967). *The Data Model Concept in Statistical Mapping*, *International Yearbook of Cartography*. International Cartographic Association, 7:186-190.
- Kashnitsky, I., & Aburto, J. (2020). COVID-19 in unequally ageing European regions. *World development*, 136, 105170. <https://doi.org/10.1016/j.worlddev.2020.105170>
- Koch, T. (2005). *Cartographies of Disease: Maps, Mapping, and Medicine*. Redlands, California, U.S.A.: ESRI Press.
- Kulldorff, M., & Nagarwalla, N. (1995). Spatial disease clusters: Detection and Inference. *Statistics in Medicine* 14(8), 799-810. doi: 10.1002/sim.4780140809
- LeSage, J. (1999). The theory and practice of spatial econometrics. *University of Toledo. Toledo, Ohio*, 28(11).
- Marshall, R. M. (1991). Mapping disease and mortality rates using Empirical Bayes Estimators. *Royal Statistical Society: Series C Applied Statistics*, 40(2):283-94.
- Millo, G., & Piras, G. (2018). *Package 'splm'*. Retrieved from <https://cran.r-project.org/web/packages/splm/index.html>
- Modig, K., Ahlbom, A., & Matthews, A. (2020). Total mortalitet bättre vid jämförelser än död i covid-19 [Covid-19 - deaths and analysis]. *Lakartidningen*, 30;117:F3XL. Retrieved from: <https://pubmed.ncbi.nlm.nih.gov/32365212/>
- Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika* 37(1/2), pp. 17-23.
- Nuclear Threat Initiative, Johns Hopkins Center for Health Security, The Economist Intelligence Unit. In *Global Health Security Index*. Retrieved from: <https://www.ghsindex.org>
- Okell, L., Verity, R., Watson, O., Mishra, S., Walker, P., Whittaker, C., et al. (2020). Have deaths from COVID-19 in Europe plateaued due to herd immunity? *The Lancet*, 395(10241):e110-e111. doi: 10.1016/S0140-6736(20)31357-X
- O'Loughlin, J., & Anselin, R. (1992). Geography of International Conflict and Cooperation: Theory and Methods. In M. Ward (Ed.). *The New Geopolitics*. Philadelphia, U.S.A.: Gordon and Breach.
- Orłowski, E., & Goldsmith, D. (2020). Four months into the COVID-19 pandemic, Sweden's prized herd immunity is nowhere in sight. *Journal of the Royal Society of Medicine*, 113(8), 292-298.
- Palomino, J., Rodríguez, J., & Sebastian, R. (2020). Wage Inequality and Poverty Effects of Lockdown and Social Distancing in Europe. doi: <https://doi.org/10.1016/j.euroecorev.2020.103564>. Retrieved from: SSRN: <https://ssrn.com/abstract=3615615> or <http://dx.doi.org/10.2139/ssrn.3615615>
- Paradis, E., Blomberg, S., Bolker, B., Brown, J., Claramunt, S., Claude, J., et al. (2020). *Package 'ape'*. Retrieved from <https://cran.r-project.org/web/packages/ape/index.html>
- Patel, J., Nielsen, F., Badiani, A., Assi, S., Unadkat, V., Patel, B., et al. (2020). Poverty, Inequality & COVID-19: The Forgotten Vulnerable, Public Health. *Public Health*, 183, 110-111.
- Porta M, e. (2008). *Dictionary of Epidemiology*. New York, U.S.A.: Oxford University Press. doi: 10.1093/acref/9780195314496.001.0001
- Potthoff, R. F., & Whittinghill, M. (1966). Testing for homogeneity: II. The Poisson distribution. *Biometrika*, 53(1/2), 183-190.
- QGIS.org. (2020). *QGIS Sistema de Información Geográfica. Proyecto de Fundación Geoespacial de Código Abierto*. Retrieved from <http://www.qgis.org/>
- Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, 46(sup1), 234-240. doi: <https://doi.org/10.2307/143141>
- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S. Fourth edition*. New York, U.S.A.: Springer-Verlag.

- Wakefield, J. H., & Lyons, H. (2010). Spatial aggregation and the ecological fallacy. In Gelfand, A. E., Diggle, P. J., Fuentes, M., & Guttorp, P. (Ed.) *Handbook of spatial statistics*. Boca Ratón, Florida, U.S.A.: CRC Press.
- Waller, L. A., & Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data*. Hoboken, New Jersey, U.S.A.: John Wiley & Sons.
- Zeileis, A., & Hornik, K. (2018). *Package 'ctv'*. Retrieved from <https://cran.r-project.org/web/packages/ctv/index.html>
- Zeileis, A., & Hothorn, T. (2002). *Package 'lmtest'*. Retrieved from <https://CRAN.R-project.org/doc/Rnews/>
- Zhao, Q., Meng, M., Kumar, R., Wu, Y., Huang, J., Lian, N., et al. (2020). The impact of COPD and smoking history on the severity of COVID-19: a systemic review and meta-analysis. *Journal of Medical Virology*. doi: <https://doi.org/10.1002/jmv.25889>
- Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., et al. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, 395(10229):1054-1062. doi: [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3)