

A bootstrap method based on linear regression to estimate COVID-19 Ecological Risk in Catalonia

Nicolas Ayala-Aldana ^{1,*}, Antonio Monleon-Getino ¹, Jaume Canela-Soler ², Petia Radeva ³ and Javier Rodenas ³

1 Department of Genetics, Microbiology, and Statistics, Section of Statistics, Faculty of Biology, University of Barcelona, Barcelona, Spain.

2 Department of Clinical Foundations, School of Medicine and Health Sciences, University of Barcelona, Barcelona, Spain

3 Department of Mathematics and Computer Science, Faculty of Mathematics and Computer Science, University of Barcelona, Spain.

World Journal of Advanced Research and Reviews, 2023, 17(01), 324–332

Publication history: Received on 02 December 2022; revised on 09 January 2023; accepted on 12 January 2023

Article DOI: <https://doi.org/10.30574/wjarr.2023.17.1.0047>

Abstract

Background: SARS-CoV-2 is a new type of coronavirus that causes COVID-19. It is affecting the entire planet. Despite the widespread use of ecologic analysis in epidemiologic research and health planning, health scientists and practitioners have given little attention to the methodological aspects of this approach. The study of risk factors linked to the COVID-19 pandemic is one of the most current and exciting topics for epidemiologists. These risks in many cases are unknown. This research covers the study of risk factors in the case of COVID-19 and proposes the use of an ecologic method known to epidemiologists in the case of aggregated data. The present study aims to compute a model that allows to easily calculate the risk of infection in different types of populations, using aggregated data to approximate the individual risk of COVID-19 transmission by a person.

Methods: The case of Catalonia, in Spain, is presented as an example, as it is one of the areas where the incidence of the virus among the population is being higher. The proposed method is known as an ecological study and is based on the statistical regression model between the incidence (or variable that represents it) and the risk factors but using aggregated data and obtaining a risk ratio (RR).

Results: The results obtained have made it possible to find the risk of contracting COVID-19 concerning risk factors for high family income (RR=1.157491), more mobility (RR=1.065475), and high density of population (1.00002).

Conclusions: This method could be used to design an app that predicts how the risk will evolve and calculate the risk of contagion in one area or another to take the proper action. The calculated RR can help us to understand how the variables become risks or protective factors at an ecological level (understanding aggregate data).

Keywords: COVID-19; Outbreak; Epidemic dynamics; Modelling; Relative risk

1. Introduction

SARS-CoV-2 is a new type of coronavirus (a broad family of viruses that normally affect only animals) that can affect people and causes COVID-19. It was detected for the first time in 2019/12 in the city of Wuhan (China). Coronaviruses produce clinical conditions ranging from the common cold to more serious diseases. For example, the coronavirus that

* Corresponding author: Nicolas Ayala-Aldana 0000-0002-3530-6734

caused severe acute respiratory syndrome (SARS-CoV) a few years ago and the coronavirus that causes the Middle East respiratory syndrome (MERS-CoV) (1,2).

The mathematical models in epidemiology are useful to compute incidence, prevalence and estimate the consequences in the population affected by COVID-19. These models use simple and multivariate linear regression, Bayesian statistics, deterministic and stochastic models. These types of predictions are difficult to manage, but are useful for individual cases, like for example, to know if outbreaks will appear in one place or another in the territory studied, or what is the consequence of applying one measure or another.

That is why it is necessary to delve into other models that approximate the risk of individual contagion, not only knowing the prevalence and incidence of those affected, but also through knowing the associated risk factors, such as the mobility of the person, characteristics of the studied area (wealth, population density, etc.), and many others. The calculation of a probability of contagion is not easy and multiple attempts have been made to obtain and simplify it since multiple factors make one or another person contagious (incidence, prevalence, social contact of the person, associated pathologies, age, etc).

The calculation of the risk of contagion of COVID-19 is proposed through aggregated data using an approach of the ecological study type, as proposed by Morgenstern (4), Beral et al (5) and Silcocks et al (6). Ecologic studies are empirical investigations involving the group as the unit of analysis. In ecological studies, disease rates and exposures are measured in each of a series of populations and their relation is examined and has been extensively studied by works such as Morgenstern in "Uses of Ecologic Analysis in Epidemiologic Research" (1982) (4). Frequently, the information about disease and exposure is abstracted from published statistics and therefore does not require expensive or time-consuming data collection.

In ecological studies, usually, the group is carefully chosen so that it belongs to a geographically defined area (e.g., city, region, country). This eases the data collection because statistics can be obtained by combining existing data files on large populations. Ecologic studies are generally less expensive and take less time than studies involving the individual as the unit of analysis. Ecologic studies allow establishing a relationship between variables or risk factors such as geographical comparisons, time trends, migrant population and social class with the disease or dependent variable of the study. On the other hand, data on many variables (e.g., clinical histories, analysis, questionnaires, etc.) may not be available at the ecologic level, and the results of ecologic analyses are subject to certain limitations not applicable to many other study designs (e.g., clinical trials, cohort studies, etc). One of the most known limitations is the ecological fallacy where conclusions are inappropriately inferred about individuals from aggregate data results. It is erroneously concluded that the statistical correlation between two variables is equal to the correlation between the corresponding variables at the individual level. (7,8).

Objective

The present study aims to compute a model that allows to easily calculate the risk of infection in Catalonia (Spain). It is intended to use a model of the incidence as a function of the mobility, density of population, etc. to determine the relative ecological risk of the population and of the person contracting COVID-19, using ecological epidemiology methods.

2. Material and methods

2.1. Statistical data

The data used to feed the model was obtained from different public sources such as:

IDESCAT (Institute of Statistics of Catalonia): <https://www.idescat.cat/pub/?id=aec&n=250&lang=es>;
<https://www.idescat.cat/emex/?id=080193#h40000000>, Generalitat of Catalonia (Catalonia Government):
<http://aquas.gencat.cat/ca/actualitat/ultimes-dades-coronavirus/mapa-per-abs/>, Municipal data from environmental pollution stations: <https://analisi.transparenciacatalunya.cat/Medi-Ambient/Dades-d-immissi-dels-punts-de-mesurament-de-la-Xar/uy6k-2s8r/data, etc>.

The data collection and the epidemiological models can be seen alongside the study code at <https://github.com/nicolas-ayala-aldana/Model-Risk>. The aggregated data consists of 233 health areas in the rows (sanitary area) compiled in a spreadsheet by these different databases where the following variables are the columns

- Health_area: Part of the municipality where the incidence of COVID19 has been noted. It is related to a sanitary area where a certain number of inhabitants is controlled.
- Municipality: Municipality or village of Catalonia, Spain.
- Positive_cases: Total positive cases.
- Suspected_cases: Total suspected cases.
- Raw_rate_10k: Raw rate per 10,000 inhabitants.
- Standard_rate_10k: Standardized crude rate per 10,000 inhabitants.
- Insured_people: Insured people in the area.
- Surface_km2: Area occupied by the health area in km².
- Density: Population density by km².
- NOX_reduction_january2020: Reduction of the percentage of NOx in January 2020 compared to the previous 3 years.
- NOX_reduction_february2020: Reduction of the percentage of NOx in February 2020 compared to the previous 3 years.
- NOX_reduction_march2020: Reduction of the percentage of NOx in March 2020 compared to the previous 3 years.
- NOX_reduction_april2020: Reduction of the percentage of NOx in April 2020 compared to the previous 3 years.
- NOX_reduction_may2020: Reduction of the percentage of NOx in May 2020 compared to the previous 3 years.
- Income_euros: Family income (thousands of euros). Catalan acronym called “RFBD”
- Income_euros_inhab: Family income per inhabitant (thousands of euros). Catalan acronym called “RFBD per inhabitant”.
- Income_index_100: Family income per inhabitant (index between 0-100%) referred to the whole of Catalonia.

At the time of composing this study, we used the data available until the middle of May 2020, at which point the daily case curve reached the inflection point and was no longer growing. In many cases, the complete information on the variables was not available, since the values were missing, or not all the municipalities had statistical data or environmental information. The pollution information (NOx Reduction) was used to estimate mobility indirectly, so we assumed that the more pollution reduction the less mobility since a smaller number of vehicles emit less NOx. Our main variable (Y) to predict is standard rate, which is the standardized crude COVID19 rate per 10,000 inhabitants in the different health areas.

2.2. Relative risk (RR) estimation using aggregated data of current epidemiological ecological study

As previously mentioned, it is very difficult to calculate the individual risk of COVID-19 infection in a heterogeneous population (e.g., countryside, city, high and low density of inhabitants, rich areas, poor areas, etc.) in a territory as extensive as Catalonia (32,108 km²). That is why an indirect system has been devised for its study from the ideas raised by the ecological studies used in the epidemiological field. One way to approximate this probability is by using the Relative Risk (RR).

In mathematical terms, the RR is the ratio of the probability of an outcome in an exposed group to the probability of an outcome in an unexposed group (9). It is computed as:

$$RR = \frac{\text{Risk in exposed}}{\text{Risk in nonexposed}}$$

RR measures the association between exposure and the outcome. Relative risk can be estimated from a 2x2 contingency table in a clinical or epidemiological study (Table 1)

Table 1 Calculation of relative risk in epidemiology

	Group	
	Disease develops	The disease does not develop
Exposed	a	b
Non-exposed	c	D

Where *a*, *b*, *c* and *d* are frequencies of the different cells for group-event. The point estimate of the relative risk (RR) is:

$$RR = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

Thus, different factors of interest that may be related to the risk of COVID-19 infection could be studied, such as mobility of people, population density, family income, etc. These factors are reasonably easy to know but unfortunately, the epidemiology of COVID-19 is complex and data for individual risk at habitant levels is not possible to obtain. As these are not available, we use the crude rate per 10,000 inhabitants, which would correspond to the independent variable of the model (Y). Finally, the variables that can be used as regressors (independent terms of the model, X) are aggregate and quantitative variables (e.g.: population density per city, family income per area, mobility per area).

Different authors suggest that it is possible to use the *RR* (relative risk) or risk ratio as a tool calculated by the linear regression models [$RR=1+(\text{slope}/\text{intercept})$] using ecological studies (4,5). This is an indirect and inexpensive method for estimating relative risk between disease in exposed people versus disease in non-exposed people. Even though problems of confusion factors could be inconvenient, it seems feasible in exploratory designs (6).

The calculation of the relative ecological risk approximation as a method to approximate *RR* related to COVID-19 from the aggregated data is presented below, using the method proposed by Morgenstern (4) using a linear regression model, simple or multiple, to obtain an estimation of *RR*. The level of aggregation, in this case, is the municipality/portion present in the Health System of the Generalitat de Catalonia (see data explanation).

The linear regression model was used to estimate *RR* using the method proposed by Morgenstern (4) where the distribution of a single response variable (Y = standardized crude rate per 10,000 inhabitants) is related to several explanatory variables, X_1, X_2, \dots , by a regression model Monleon, 2017 (10):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} + \varepsilon_i$$

Where β_0 is the intercept, β_1, \dots, β_p the model coefficients, and ε_i the error term, where $\varepsilon_i \sim N(0, \sigma)$ is the distribution of the errors, has mean zero and captures the residual variability.

Independent variables used in the linear models (X_i) are:

- Density of population: Inhabitants / km²
- Reduction % NOx for March 2020 vs 2017-2019
- Reduction % NOx for April 2020 vs 2017-2019
- Reduction % NOx for May 2020 vs 2017-2019
- Family income per inhabitant (thousands of euros)

First, a univariate regression model (see equation 2) with a single X_i was calculated. *RR* was computed using:

$$RR = 1 + \frac{\beta_1}{\beta_0}$$

In a second step, a regression model with different X_i and their interactions $X_i X_j$ was tested. In this case, the significance level used was 0.001. We proposed a highly demanding alpha because it is more difficult to accept interactions.

3. Results

Figure 1 shows the number of daily cases of COVID-19 from February to July 2020. The peak of new cases observed was 1895 on 20th March, then daily cases had a significant decrease until reaching 51 cases on 6th May. After that, the number of cases fluctuated until July 2020. The statistical package R (version 3.6) was used to perform the analyses using simple linear regression calculations between the incidence of COVID-19 as the dependent variable (Y) and the commented risk factors (X_i) as independent variables. Once the regression coefficients were obtained, the relative ecological risk for each factor was calculated and after that, a model with multiple significant factors for the individual model was presented (see Table 2 and Table 3).

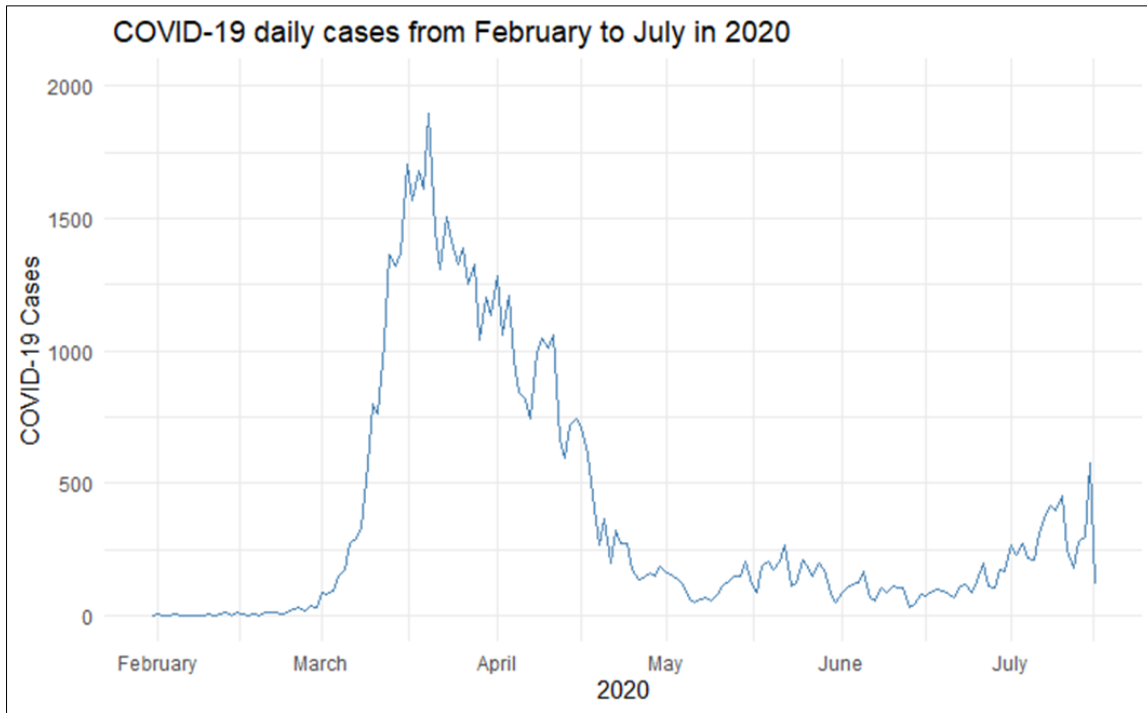


Figure 1 Representation in time of the number of cases in Catalonia (north-east of Spain) from February to July 2020.

Table 2 Estimations of β_0 and β_1 (regression coefficients in a simple linear model), RR: relative ecological risk approximation, and R^2 = coefficient of determination, also the p-values are presented

Factor	R^2	β_0	β_1	RR
Density (inhabitants /km ²)	0.1025 (p=2.182e-05)	6.494e+01 (p=<2e-16)	1.318e-03 (p=2.18e-05)	1.00002 (1.000021,1.000021,1.00002)*
Mobility (Reduction of the percentage of NOx in March 2020 compared to the previous 3 years)	0.07924 (p=0.0005279)	20.669 (p=0.223378)	1.353 (p=0.000528)	1.065475 (1.112431,0.9752694,0.8381082)*
Mobility (Reduction of the percentage of NOx in April 2020 compared to the previous 3 years)	0.9987749 (p=0.8204)	86.9407 (p=0.00283)	-0.1065 (p=0.82039)	0.9987749 (1.007246,1.003234,0.9992213)*
Mobility (Reduction of the percentage of NOx in May 2020 compared to the previous 3 years)	0.09093 (p=0.0002017)	56.94779 (p=< 2e-16)	0.35091 (p=0.000202)	1.006162 (1.006911,1.006368,1.005825)*
Family income per inhabitant	0.08054 (p=3.755e-05)	19.3861 (p=0.13)	3.0531 (p=3.76e-05)	1.157491 (1.180263,1.114656,1.04905)*

*Estimation of RRE (ULCI95%, mean, LLCI95%) using resampling with 1000 iterations and n=200 samples with replacement.

Table 3 Estimations of $\beta_0, \beta_1, \dots, \beta_p$ (regression coefficients in a multiple linear model), RR: relative ecological risk approximation, and R^2 = coefficient of determination, also the p-values are presented

Regression model Coefficient, name	β_0	β_1	β_2	β_3	β_4
		Density (inhabitants/km ²)	Mobility (Reduction of the percentage of NOX in May 2020 compared to the previous 3 years)	Family income per inhabitant (thousands of euros)	Interaction between NOx reduction of May 2020 and Family income per inhabitant
Regression model Coefficient, value	-1.147e+02	7.303e-04	3.955e+00	8.384e+00	-2.059e-01
p-value	0.08506	0.07547	0.00318	0.02052	0.00204
R ²	0.2082				
	p-value = 6.45e-06				

Assuming the causal effect between the exposure (Xi) and the outcome (Y), values of RR can be interpreted as follows:

- RR = 1 means that exposure does not affect the outcome.
- RR < 1 means that the risk of the outcome is decreased by exposure.
- RR > 1 means that the risk of the outcome is increased by exposure.

In Table 1 the estimations of β_0 and β_1 (regression coefficients in a simple linear model) used to compute RR are shown: relative ecological risk approximation. The compute of β_1 variables are: Density ($\beta_1 = 1.318e-03$; $p = 2.18e-05$), March 2020 mobility ($\beta_1 = 1.353$; $p = 0.000528$), April 2020 mobility ($\beta_1 = -0.1065$; $p = 0.000202$), May 2020 Mobility ($\beta_1 = 0.35091$; $p = 3.76e-05$) and Family Income ($\beta_1 = 3.0531$; $p = 3.76e-05$). They are significant to reject the null hypothesis of $H_0: \beta_1 = 0$ ($p < .001$) and in consequence accepting the alternative hypothesis.

When analysing the RR, we can state that people who live in high-density areas have 1.00002 times more risk of infection than people who live in low-density areas (RR=1.00002). Regarding mobility, the results obtained for each month were: March 2020 (RR=1.065475), April 2020 (RR=0.9987749) and May 2020 (RR=1.006162). Therefore, the reduction of mobility can be considered as a protective factor for COVID-19 risk infection. In April 2020, the risk of infection for people who live in areas with less mobility is 1.0012251 times lower than people who live in high-mobility areas. Finally, the RR obtained for family income per inhabitant is 1.157491, meaning that the higher the family income per inhabitant, the more risk of infection.

Density, mobility, family income and interaction of multivariate analysis are presented in Table 3. Density and Family Income per inhabitant have a p-value >0.01 and therefore it cannot rule out that RR=0. Because the p-values are greater than 0.01, no relationship can be established between variables such as density ($\beta_1 = 7.303e - 04$; $p = 0.07547$) and Family income ($\beta_3 = 8.384e + 00$; $p = 0.02052$). Mobility ($\beta_2 = 3.955e + 00$; $p = 0.00318$) has p-value less than 0.01 therefore a positive relationship can be established between the mobility and COVID-19 cases. Interaction between variables NOx reduction of May 2020 and family income per inhabitant is observed ($\beta_4 = -2.059e - 01$; $p = 0.00204$). Generally, it is suggested not to use multivariate models with joint factors for their use in epidemiology.

4. Discussions

In a study with 1031 suburban areas of 314 Latin American cities a 10% decrease in weekly mobility was associated with an 8.6% lower incidence of COVID-19 in the following week (12). In a study that used population mobility from Google services in 34 OECD countries plus Singapore and Taiwan, in two-thirds of the countries examined, reductions of up to 40% in mobility were associated with a decline of COVID-19 cases, especially at the beginning of the pandemic (13). In a study where 52 countries were analysed according to WHO and CDC databases, the reduction in mobility in 73% of the countries was associated with a decrease in new cases of COVID-19 (14). It is striking in our study that in April there was an inverse relationship between the decrease in NO_x and the cases of COVID-19. For March and May, the relationship was weak. The April measure was $\beta_1 = -0.1065$; $p = 0.82039$, with an $RR = 0.9987749$. Reduced mobility should be even more conclusive, but it is interesting to calculate the relative risk directly from this simple method used in epidemiology. It can be explained by public policies such as the restriction of mobility. The lockdown in Spain began on March 14th and lasted until May. After that, the population was gradually allowed to carry out outdoor activities. On June 21, the Spanish government authorized the extension of the state of alarm, allowing the population to move freely from 6am until midnight. This situation was changing in the following months according to the prevalence and incidence by the autonomous community (11).

According to population density, studies show a linear relationship between the number of cases, incidence and mortality of COVID-19 in areas with higher population density such as in the case of Italy (15), England (16), Malaysia (17), Germany and Japan (18). The areas with higher density could explain the difficulties in applying adequate physical distancing and problems in isolation. In our case, no relevant association was observed, and the computation of the relative risk would assume a higher density as a risk for the disease.

Regarding family income as a social determinant in health, other studies have shown that groups with high socioeconomic vulnerability have a higher risk of having COVID-19 determinate by difficulty in carrying out physical distancing in their homes and difficulty in accessing health services (19–21). In Catalonia, a higher incidence of COVID-19 has been founded in the poorest areas of the town of Barcelona (22). In our case, a weak positive relationship is observed compared to the effect of population density. It may be that the per capita family income is not an optimal measurement variable in a regression model, or the variable is representing a fallacy at this ecological level.

Limitations

Results of this study should be interpreted cautiously, bearing in mind limitations attributable to the ecological design of the study. The methodological problem fundamental in assuming individual associations based on group data (ecological fallacy) is well known (7,8). Some limitations of this study could be the absence of measurement of individual socioeconomic characteristics, pollutions levels and range of density by subareas. These variables were measured on an ecological level making it impossible to adequately control the confounding factors, effect modifiers and mediators at an individual level. To overcome the limitations and bias, it is recommended to compute a multilevel analysis considering an individual and contextual level. It is impossible to distinguish the individual and contextual effects of a variable using an ecological design.

5. Conclusion

In conclusion, the proposed method is known as an ecological study and is based on the statistical regression model between the incidence (or variable that represents it) and the risk factors using aggregated data and obtaining a risk ratio (RR). The present study shows the risk factors linked to the COVID-19 in the population of Catalonia (Spain) in the first wave in 2020. Low density of population and reduction of mobility are related with lower risk of contagions of COVID-19. High family income outcomes were related as a risk factor. The ecologic study method could be an effective way to design an APP that predicts how the population is infected with COVID-19 according to risk factors and identify areas most susceptible to contagion. Its development could help in decision-making in public health.

Compliance with ethical standards

Acknowledgments

This work was supported by the University of Barcelona, Catalonia, Spain.

Disclosure of conflict of interest

All authors have no conflict of interest to declare.

References

- [1] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med*. 2020 Mar, 382(13):1199–207.
- [2] Center for Coordination of Health Alerts and Emergencies. Ministry of Health S. Document about COVID-19. 2020 [Internet]. 2020. Available from: https://www.mschs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov-China/documentos/20200224.Preguntas_respuestas_COVID-19.pdf
- [3] López MG, Chiner-Oms Á, García de Viedma D, Ruiz-Rodríguez P, Bracho MA, Cancino-Muñoz I, et al. The first wave of the COVID-19 epidemic in Spain was associated with early introductions and fast spread of a dominating genetic variant. *Nat Genet* [Internet]. 2021, 53(10):1405–14. Available from: <https://doi.org/10.1038/s41588-021-00936-6>
- [4] Morgenstern H. Uses of ecologic analysis in epidemiologic research. *Am J Public Health*. 1982 Dec, 72(12):1336–44.
- [5] Beral V, Chilvers C, Fraser P. On the estimation of relative risk from vital statistical data. *J Epidemiol Community Health* [Internet]. 1979 Jun, 33(2):159–62. Available from: <https://pubmed.ncbi.nlm.nih.gov/490098>
- [6] Silcocks PB, Murphy M. Relative risk estimation from vital statistical data: validation, a pitfall and an alternative method. *J Epidemiol Community Health*. 1987 Mar, 41(1):59–62.
- [7] Schwartz S. The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. *Am J Public Health* [Internet]. 1994 May, 84(5):819–24. Available from: <https://pubmed.ncbi.nlm.nih.gov/8179055>
- [8] Loney T, Nagelkerke NJ. The individualistic fallacy, ecological studies and instrumental variables: a causal interpretation. *Emerg Themes Epidemiol* [Internet]. 2014 Nov 19, 11:18. Available from: <https://pubmed.ncbi.nlm.nih.gov/25745504>
- [9] Celentano , Szklo, M., Gordis, Leon,, DD. *Gordis epidemiology*. 2019.
- [10] Monleón A, Gomez C. *Probability and Statistics I*. 1st ed. Barcelona U of, editor. Barcelona, 2017.
- [11] La Moncloa. Estado de Alarma [Internet]. [cited 2022 Mar 22]. Available from: <https://www.lamoncloa.gob.es/covid-19/Paginas/estado-de-alarma.aspx>
- [12] Kephart JL, Delclòs-Alió X, Rodríguez DA, Sarmiento OL, Barrientos-Gutiérrez T, Ramirez-Zea M, et al. The effect of population mobility on COVID-19 incidence in 314 Latin American cities: a longitudinal ecological study with mobile phone location data. *Lancet Digit Heal* [Internet]. 2021/08/26. 2021 Nov, 3(11):e716–22. Available from: <https://pubmed.ncbi.nlm.nih.gov/34456179>
- [13] Oh J, Lee H-Y, Khuong QL, Markuns JF, Bullen C, Barrios OEA, et al. Mobility restrictions were associated with reductions in COVID-19 incidence early in the pandemic: evidence from a real-time evaluation in 34 countries. *Sci Rep* [Internet]. 2021 Jul 2, 11(1):13717. Available from: <https://pubmed.ncbi.nlm.nih.gov/34215764>
- [14] Nouvellet P, Bhatia S, Cori A, Ainslie KEC, Baguelin M, Bhatt S, et al. Reduction in mobility and COVID-19 transmission. *Nat Commun* [Internet]. 2021, 12(1):1090. Available from: <https://doi.org/10.1038/s41467-021-21358-2>
- [15] Pluchino A, Biondo AE, Giuffrida N, Inturri G, Latora V, Le Moli R, et al. A novel methodology for epidemic risk assessment of COVID-19 outbreak. *Sci Rep* [Internet]. 2021 Mar 5, 11(1):5304. Available from: <https://pubmed.ncbi.nlm.nih.gov/33674627>
- [16] Bray I, Gibson A, White J. Coronavirus disease 2019 mortality: a multivariate ecological analysis in relation to ethnicity, population density, obesity, deprivation and pollution. *Public Health* [Internet]. 2020/07/07. 2020 Aug, 185:261–3. Available from: <https://pubmed.ncbi.nlm.nih.gov/32693249>
- [17] Ganasegeran K, Jamil MF, Ch'ng AS, Looi I, Peariasamy KM. Influence of Population Density for COVID-19 Spread in Malaysia: An Ecological Study. Vol. 18, *International Journal of Environmental Research and Public Health* . 2021.

- [18] Diao Y, Kodera S, Anzai D, Gomez-Tames J, Rashed EA, Hirata A. Influence of population density, temperature, and absolute humidity on spread and decay durations of COVID-19: A comparative study of scenarios in China, England, Germany, and Japan. *One Heal* [Internet]. 2021, 12:100203. Available from: <https://www.sciencedirect.com/science/article/pii/S2352771420303049>
- [19] Niedzwiedz CL, O'Donnell CA, Jani BD, Demou E, Ho FK, Celis-Morales C, et al. Ethnic and socioeconomic differences in SARS-CoV-2 infection: prospective cohort study using UK Biobank. *BMC Med* [Internet]. 2020, 18(1):160. Available from: <https://doi.org/10.1186/s12916-020-01640-8>
- [20] Whittle RS, Diaz-Artiles A. An ecological study of socioeconomic predictors in detection of COVID-19 cases across neighborhoods in New York City. *BMC Med* [Internet]. 2020 Sep 4, 18(1):271. Available from: <https://pubmed.ncbi.nlm.nih.gov/32883276>
- [21] Yoshikawa Y, Kawachi I. Association of Socioeconomic Characteristics With Disparities in COVID-19 Outcomes in Japan. *JAMA Netw Open* [Internet]. 2021 Jul 14, 4(7):e2117060–e2117060. Available from: <https://doi.org/10.1001/jamanetworkopen.2021.17060>
- [22] Baena-Díez JM, Barroso M, Cordeiro-Coelho SI, Díaz JL, Grau M. Impact of COVID-19 outbreak by income: hitting hardest the most deprived. *J Public Health (Bangkok)* [Internet]. 2020 Nov 23, 42(4):698–703. Available from: <https://doi.org/10.1093/pubmed/fdaa136>