pagepress

Geospatial Health 2022; volume 17:1145

Spatiotemporal heterogeneity of SARS-CoV-2 diffusion at the city level using geographically weighted Poisson regression model: The case of Bologna, Italy

Addisu Jember Zeleke,[1] Rossella Miglio,[2] Pierpaolo Palumbo,[1] Paolo Tubertini,[3] Lorenzo Chiari,[1,4] Bologna MODELS4COVID Study Group of the University of Bologna and the National Institute for Nuclear Physics (INFN), Italy*

[1]Department of Electrical, Electronic, and Information Engineering Guglielmo Marconi, University of Bologna, Bologna; [2]Department of Statistical Sciences, University of Bologna, Bologna; [3]Enterprise information systems for integrated care and research data management (IRCCS), Azienda Ospedaliero-Universitaria di Bologna, Bologna; [4]Health Sciences and Technologies Interdepartmental Center for Industrial Research (CIRI SDV), University of Bologna, Bologna, Italy

***Members: Federico Baldo, Armando Bazzani, Valerio Carelli, Gastone Castellani, Luca Clissa, Ilaria D'Ascanio, Stefano Diciotti, Francesco Durazzi, Enrico Lunedei, Antonio Macaluso, Michela Milano, Roberto Morelli, Serena Moscato, Luca Palmerini, Daniel Remondini, Giulia Roli, Michele Scagliarini, Roberto Spighi, Vincenzo Vagnoni, Antonio Zoccoli*

Correspondence: Addisu Jember Zeleke, Department of Electrical, Electronic, and Information Engineering Guglielmo Marconi, University of Bologna, 40126 Bologna, Italy.
E-mail: addisu.zeleke2@unibo.it

Key words: COVID-19; local regression; mapping; spatial heterogeneity; Bologna; Italy.

Contributions: AJZ: conceptualization, methodology, formal analysis, data curation, and writing—original draft preparation; RM: conceptualization, methodology, supervision, and writing—review and editing; PP: data curation, and writing—review and editing; PT: data curation, and writing—review and editing; LC: conceptualization, methodology, supervision, and writing—review and editing.

Conflict of interest: The Authors declare no conflict of interest.

Funding: This research was partly supported by PoliclinicoSant'Orsola-Malpighi through the funding of the PhD scholarship of AJZ.

Institutional Review Board Statement: The AUSL (Local Health Unit) Bologna and the Municipality of Bologna are the data controllers. They anonymized the data and passed them to the Bologna MODELS4COVID Study Group members of the University of Bologna and INFN, Italy for statistical analyses. Data processing was done in agreement with the Italian emergency law (article 17 of the law decree 17th March 2020, n. 18, converted in to law n. 27, 24th April 2020).

Informed consent: Not applicable.

Availability of data and materials: The data are not publicly available due to ethical restrictions.

Acknowledgments: We are grateful to those who participated in the study. We also thank for their support the Director Generals of AUSL Bologna and PoliclinicoSant'Orsola-Malpighi, Dr. Paolo Bordon and Dr. Chiara Gibertoni, and Giuseppina Civitella from the Municipality of Bologna.

See the online version for the Appendix.

Received for publication: 16 August 2022.
Revision received: 10 November 2022.
Accepted for publication: 10 November 2022.

Geospatial Health 2022; 17:1145
doi:10.4081/gh.2022.1145

Abstract

This paper aimed to analyse the spatio-temporal patterns of the diffusion of SARS-CoV-2, the virus causing coronavirus 2019 (COVID-19, in the city of Bologna, the capital and largest city of the Emilia-Romagna Region in northern Italy. The study took place from February 1st, 2020 to November 20th, 2021 and accounted for space, sociodemographic characteristics and health conditions of the resident population. A second goal was to derive a model for the level of risk of being infected by SARS-CoV-2 and to identify and measure the place-specific factors associated with the disease and its determinants. Spatial heterogeneity was tested by comparing global Poisson regression (GPR) and local geographically weighted Poisson regression (GWPR) models. The key findings were that different city areas were impacted differently during the first three epidemic waves. The area-to-area influence was estimated to exert its effect over an area with 4.7 km radius. Spatio-temporal heterogeneity patterns were found to be independent of the sociodemographic and the clinical characteristics of the resident population. Significant single-individual risk factors for detected SARS-CoV-2 infection cases were old age, hypertension, diabetes and co-morbidities. More specifically, in the global model, the average SARS-CoV-2 infection rate decreased 0.93-fold in the 21–65 years age group compared to the >65 years age group, whereas hypertension, diabetes, and any other co-morbidities (present vs absent), increased 1.28-, 1.39- and 1.15-fold, respectively. The local GWPR model had a better fit better than GPR. Due to the global geographical distribution of the pandemic, local estimates are essential for mitigating or strengthening security measures.

Introduction

Importance of addressing COVID-19

As noted by the World Health Organization (WHO), the rapid spread of Coronavirus disease 2019 (COVID-19), caused by the SARS-CoV-2 virus, is a global health problem (WHO, 2020a).

From the first case detected in Wuhan in December 2019, COVID-19 has caused 495million incident cases and over 6.1 million deaths as of April 07th, 2022 (Worldometer, 2022). This posed significant challenges to health systems, particularly in Italy, the first European country to be infected (Specchia *et al.*, 2021), where the overwhelming increase of patients requiring hospitalization at the start of the pandemic threatened the national health care system with collaps (Armocida *et al.,* 2020). The COVID-19 pandemic underlines the need for a management plan to forecast hospital service demand so that healthcare facilities can efficiently make resources accessible when needed. Furthermore, older persons, including those with underlying health conditions, like cardiovascular disease, diabetes, chronic respiratory disease and cancer, are more likely to develop severe COVID-19 illness (WHO, 2020b).

Space-time risk models for COVID-19 are useful but rare/absent in the literature

Due to the pandemic's geographical spread, local estimates can play an essential role in mitigating or strengthening security measures. These estimates may regard the level of risk at each local area or the large-scale patterns of spreading of the contagion. However, estimates at the city level are rarely reported because official data on COVID-19 cases are only released, at least in Italy, at the provincial level (the number of people infected) or at the regional level (the number of deaths due to the virus). Furthermore, patient characteristics, such as health risk factors, pre-existing health conditions and from laboratory test results, which are helpful to estimate the level of risk in a population, are hardly ever available. Thanks to the joint effort of our Local Health Authority (AUSL Bologna), an Italian research hospital (IRCCS Policlinico Sant'Orsola-Malpighi) and the Municipality of Bologna, we could access these data in a common framework, after identification, linkage, and anonymization of the relevant data sources. This allowed us to estimate the level of risk in each city area and the space-and-time patterns of SARS-CoV h-2 diffusion across areas. It also permitted investigation how demographic inequalities, health risk factors and cases at different periods are associated with the infection dynamics. Spatial models are important tools for studying the geographic relationship between many explanatory variables and disease outbreaks (Mollalo *et al.,* 2015; Mollalo *et al.,* 2016). Spatial analysis of small geographical areas can provide useful information about at-risk areas, but it is not yet possible to investigate geographical variation or spatial analysis at the level of a single city.

Statistical methods to derive space-time models

Most studies in medical research are based on classical regression models like ordinary least square (OLS) regression and generalized linear models (GLM) (Choi *et al.,* 2017; Takele *et al.,* 2019; Zheng *et al.,* 2017) but these classical models produce bias by the resulting, average parameters over the whole study area without considering geographical variation. This kind of global regression models cannot detect non-stationary phenomena and may thus obscure differences in relationships between predictors and the outcome variable. Based on results from global models in which non-stationary is present but not detected, public policy inferences will be variable and may even be relatively poor in specific local/regional cases (Ali *et al.*, 2007). Local geographically weighted Poisson regression (GWPR) modelling, on the other hand, calculates local regression coefficients to help healthcare professionals better understand how the effects of independent factors vary depending on where they are located (Haque *et al.*, 2012; Matthews *et al.,* 2012). This allowed us to investigate possible geographical variations in disease infection rates and other health problems accounting for the level of risk at the local, resident population level (Nakaya *et al.,* 2005; Yang *et al.,* 2012; Zhou *et al.,* 2015). The unobserved spatial heterogeneity cannot be directly accounted for in standard models or global regression models since the estimated parameters are fixed and represent a mean relationship between dependent and explanatory variables. However, a GWPR model is best for relaxing the fixed association assumption between the response and the explanatory variables thereby capturing spatial heterogeneity. Recent studies have shown that the applications of GWPR in a wide variety of fields, including but not limited to, analysis of the effects of demographic and socio-economic factors on the spatial variation of COVID-19 (Chen *et al.,* 2022; Shawky *et al.,* 2021; Zhang *et al.,* 2021), health and other disease analysis (Bui *et al.,* 2018 ; Chen *et al.,* 2010; Goovaerts, 2005; Nakaya *et al.,* 2005; Poliart *et al.,* 2021; Yang *et al.,* 2009), traffic crash modelling (Li *et al.,* 2013; Zhao *et al.,* 2004), population density and housing (Mennis and Jordan, 2005; Chen *et al.,* 2017), poverty mapping (Benson *et al.,* 2005; Loubert *et al.,* 2018) and diseases mapping (Nakaya *et al.,* 2005). In most of these studies, however, one of the challenges is the presentation and synthesis of the large number of "mappable" results generated by local GWR models.

Risk factors for COVID-19 infection

Recent studies referred to by the U.S. Centers for Disease Control and Prevention (CDC) have shown that the elderly and individuals with underlying co-morbidities, such as cardiovascular disease, diabetes and hypertension, are more susceptible to complications and possibly death due to COVID-19 (CDC COVID-19 Response Team, 2020). These groups also stand an increased risk of becoming infected (CDC, 2020).

Infections are more likely to occur in those with diabetes who are older and have poor glycemic control (Klekotka *et al.,* 2015; Umpierrez *et al.,* 2002). In the absence of publicly available findings, it is clear that diabetes patients are more likely to contract COVID-19 infection (Choi *et al.,* 2016). Hypertensive patients had a higher rate of COVID-19 infection in a trial of 41 patients conducted in a Wuhan hospital (Huang *et al.,* 2020); a larger analysis of 138 hospitalized patients confirmed these results (Wang *et al.,* 2020). There is a heterogeneous nature to COVID-19 transmission processes over space (Wang *et al.,* 2020; Chen *et al.,* 2021). Variables such as local demographics, risk factors, epidemic waves and socio-economic characteristics, play a role in COVID-19 transmission (Dowd *et al.,* 2020; Gatto *et al.,* 2020; Pedro *et al.,* 2020; Qiu *et al.,* 2020). With the help of local modelling, researchers have been able to estimate the geographic variation in the relation between the outcome variable and explanatory factors more accurately. However, in some circumstances, predictors in the GWPR model might not all have spatial effects, so there is a need to check the spatial variability of the predictors in advance.

Aims

In this study, we aimed to analyze space and time patterns of SARS-CoV-2 diffusion across the 90 statistical area of Bologna, Italy, from February 1st, 2020 to November 20th, 2021, accounting for the space-referenced sociodemographic characteristics and health conditions of the resident population. Notably, we take an interest in how well global and local GPWR models explain the

variations in disease infection rates, which means that using the global model in the absence of variation and the local one if variations were found, comparing the results for optimal results. The secondary goal of this study was to derive a model for estimating the level of risk of infection with SARS-CoV-2, based on sociodemographic factors and health conditions and to identify and measure the place-specific factors associated with diseases and their determinants. That that end, COVID-19 infection levels were compared with respect to sociodemographic and clinical features at different time points over the entire study period.

Materials and Methods

Study area and population

Our study area refers to the city of Bologna, the capital and largest city of the Emilia-Romagna Region in northern Italy. It is the seventh most populous city in Italy, with about 400,000 inhabitants and 150 different nationalities, covering an area of 140.9 km^2 (Ufficio Statistica Regionale-Regione Emilia Romagna, 2019). Figure 1 shows the subdivision of the municipal territory into 90 statistical areas, pre-defined by the Statistical Bureau of the Bologna Municipality.

This study obtained data with permission from the Local Health Unit (AUSL) of the Municipality of Bologna. The reference population consisted of all residents in the city as of February 1st, 2020 (Figure 1).

Study variables

The infection rate of COVID-19 in each of the 90 statistical areas was taken as the response (outcome) variable); as usual, the population size was used as offset in the following models. The individual COVID-19 case status was classed as 1 = Yes and 0 = No. Some explanatory variables were referred to as population characteristics and others as area characteristics. In this study, the former comprised demographics, health-related characteristics, age (divided into <20, 20-65 and >65 years), gender, family (divided according to size (n=1), (n=2), (n=3) and (n=≥4), presence of co-morbidities (hypertension, diabetes or other). When tested for COVID-19 infection, people had documented their coexistent or preconditioned medical history according to the Index of Coexistent Disease (ICED).

The area variables included data about the 90 statistical areas of the city of Bologna. In particular, we used the total number of residents as well as the position (latitude and longitude) of the centroid of each statistical area. The minimum geo-point Y-coordinate (latitude) was 44.4592, with the maximum = 44.5463, while the minimum X-coordinate (longitude) was 11.2359, with the maximum = 11.3638. The COVID-19 case counts (all individual information) were aggregated by statistical area (into the 90 areas).

Study periods

The observation period extended from February 1st, 2020 to November 20th, 2021.

As in our previous study (Zeleke *et al.,* 2022), we partitioned our observation periods according to epidemic waves assuming that a wave starts with more than 500 total hospitalizations and ends when the hospitalization rate drops below 500. Using this cut-off, we chose a day in the stationary phase between the second and the third wave located in the middle of this period. We chose a middle point to divide the second from the third wave since hospitalization in that period never dropped to less than 500. Using this criterion, four wave periods were identified (Figure 2).

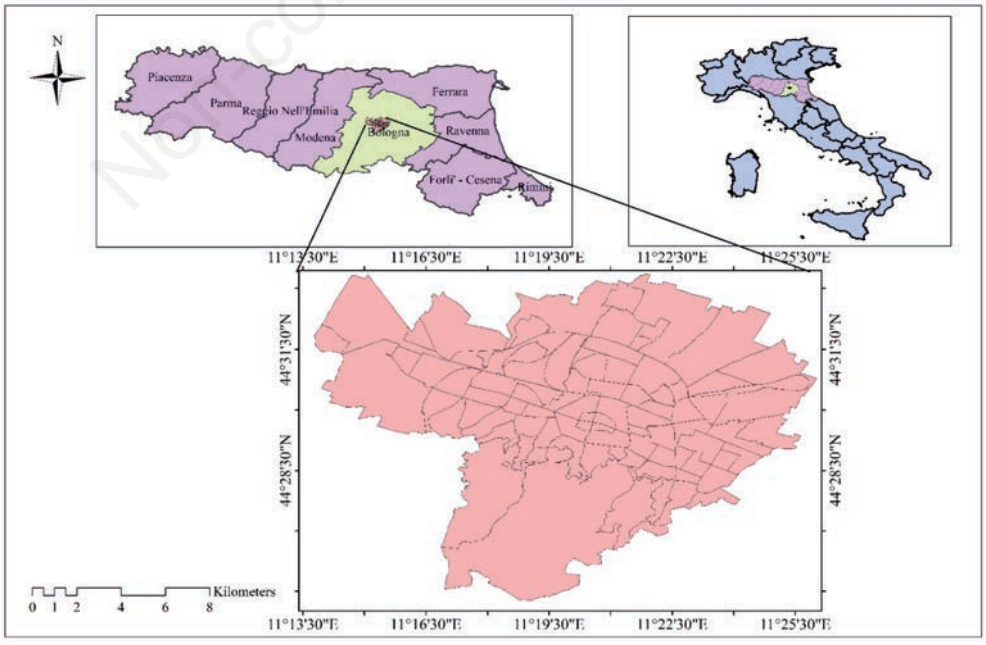

Figure 1. Location of the study area. The city of Bologna, here divided into 90 statistical areas, is the capital of the Emilia-Romagna Region (top-left panel).

Statistical modelling

Estimation of infection rates

COVID -19 infection rate each local statistical area was calculated using the following formula for:

$$IR_i = \frac{C_i}{P_i} \quad (1)$$

where IR^i is the estimated infection rate for statistical area unit i ; C_i the number of observed cases in the statistical area unit; and P_i is the total number of people living in the i^{th} statistical area.

Comparison of COVID-19 infection rate predictors during the different waves

COVID-19 infection rates were compared during different time points over the entire study period based on sociodemographic features, clinical features and case data. The characteristics and dates for each wave are listed in Figure 2.

Spatial distribution of predictors and the vaccination status by area

We looked at how each sociodemographic variable was distributed and classified in each statistical area in the city. Additionally, we tried to determine the vaccination status of the population in the areas. However, because individual vaccination varies over time and can be partial or complete, it would not be appropriate to use the COVID-19 infection rate as a measure of vaccination status. Instead, the time-to-event model, *i.e.* the survival analysis according to the Cox proportional-hazards (COXPH) model (Cox, 1972), was preferable since it also helped control the time-varying confounding factors and examine the relationship between a predictor variable and the survival time.

Testing for spatial variability or spatial non-stationarity

The variability of the local coefficient over space can be used to examine the plausibility of the stationarity assumption held in global regression. As suggested by Nakaya (2014) for the GWR4 Windows application, we performed a model comparison for each variable coefficient to determine its geographic variability based on the difference in model comparison indicator between the GWR model and the switched GWPR model (Diff of Criterion), which indicates whether the variable is a local variable (locally fixed) or a global variable (locally variable). For example, to compare the variability of the k^{th} varying coefficient, we compared the fitted GWPR - the original model, with the switched model, which keeps the remaining coefficients unchanged from the fitted GWPR. When the corrected Akaike Information Criterion (AICc) comparison suggests that the original GWPR model outperforms a switched GWPR model, it could be inferred that the k^{th} coefficient changes significantly with space (the DIFF of Criterion becomes negative). In a switched GWR model that is well fitted, there is no spatial variability in the local coefficients (the DIFF of Criterion becomes positive). Thus, when this indicator gives positive numbers, there is no spatial variability, which means that the global estimate can be considered; when negative, however, the DIFF of Criterion is a sign of spatial variability or non-stationarity.

Similarly, many other authors (Brunsdon *et al.*, 1996; Fotheringham *et al.*, 2003; Nakaya *et al.,* 2005) used the Monte Carlo test. Multiscale GWR (MGWR) 2.2 (Oshan *et al.,* 2019)allows running a Monte Carlo test to determine whether the spatial variability of the local estimates is due to sampling variation or other intrinsic processes. The Monte Carlo test calculates local parameter estimates once and then calculates new local parameter estimates after randomly rearranging the data points to see if the variability of each parameter surface is due to chance (*i.e.* using the 0.05 significance level, non-significant values ($p>0.05$) are considered as absence of spatial heterogeneity, while other significant values indicate spatial heterogeneity. Therefore local coefficient estimation for a variable is the next step. In our analysis, we considered both criteria to test for spatial non-stationarity.

Global Poisson Regression model (GPR)

A local regression model can evaluate the presence of changes in the importance of different variables over space, while global regression models express the relationship between the dependent

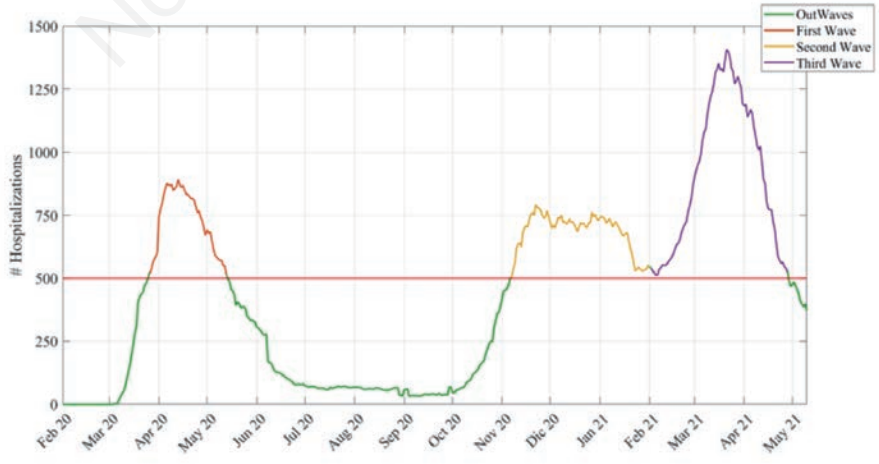

Figure 2. Number of hospitalizations over the study period and the empirical definition of infection waves. First Wave → 26 March 2020 - 13 May 2020 (48 days); Second Wave →07 November 2020- 01February 2021 (86 days); Third Wave →02February 2021 – 28 April 2021 (85 days); Out Waves → other periods.

and independent variables by a constant mean value across the study area. The relationship between COVID-19 infection rate and available covariates were considered using global and local regression models. The classical GLM Poisson regression model (Fotheringham *et al.,* 2013) is formulated as follows:

$$ln(E(IR_i|X)) = \beta_0 + \sum_k^p \beta_k x_{k,i} \quad (2)$$

where IR_i is the value of the outcome variable (here the infection rate of COVID-19) in each statistical area of i; β_0 the global Intercept; x_k is the k^{th} explanatory variable and β_k are the parameters corresponding to the explanatory variables.

Local GWPR

Fotheringham *et al.* (2003) introduced GWPR to describe a family of regression models in which the coefficients are allowed to vary spatially and focus attention on local variations. Unlike the global model, the linear GWR is a non-stationary regression model with an additional function of spatial location, where the estimated coefficient parameters vary over space. The equation of this fitted model is as follows:

$$ln(E(IR_i|X) = \beta_0(u_i, v_i) + \sum_k^p \beta_k(u_i, v_i) x_{ik} \quad (3)$$

where IR_i is the value of the outcome variable at the coordinate location i; (u_i, v_i); the two-dimensional geographical coordinates of the centroid of for each statistical area; β_0 and β_k represent the local estimated intercept and effect of variable j for location i, respectivelyThe iterative reweighted OLS method was used to calibrate the GWPR model using the GWR4.0 software (Nakaya *et al.,* 2014). The parameter estimation value of sample i is given by (Fotheringham *et al.,* 2003):

$$\hat{\beta}(u_i, v_i) = (X^T w(u_i, v_i) X)^{-1} X^T w(u_i, v_i) IR \quad (4)$$

where $\hat{\beta}$ (u_i, v_i) is the vector of the local parameters in each statistical area of i; X the matrix of the independent variables with a column of 1s for the intercept; and $w(u_i, v_i)$ is the spatial weight matrix that can be presented as:

$$w(u_i, v_i) = \begin{bmatrix} w_{i1} & 0 & \dots & 0 \\ 0 & w_{i2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & w_{in} \end{bmatrix} \quad (5)$$

where w_{ij} is the weight given to each statistical area of j during the calibrating procedure for the statistical area i.
The calculation of GWPR coefficients consists of two major steps, the first of which is choosing a proper kernel function to express the spatial relationship between the observed units. and the second the selection of the optimal bandwidth for which the spatial weight matrix calculation could contribute to a better fit.

Kernel function and bandwidth selection

Using a kernel regression method to calibrate the model to predict smoothed geographical parameter variations with a distance-based weighting scheme (Nakaya *et al.,* 2005) is a key step in the development of the GWPR model. There are five kernel functions used for GWR or GWPR modelling in different research areas: Box-car, bi-square, tri-cube, exponential and Gaussian weighted function (Li *et al.,* 2013). To obtain the most accurate possible estimation of spatial heterogeneity, bi-square was chosen after careful analysis and comparison of our dataset. It was expressed as follows:

$$w_{ij} = \begin{cases} [1 - (||u_i - v_j||/b)^2]^2 if\ ||u_i - v_j|| < b_i \\ 0 \qquad\qquad otherwise \end{cases} \quad (6)$$

where w_{ij} is the spatial/geographical weight; j a specific point in space where data are observe; i any point in space for which parameters are estimated; and $||u_i - v_j||$ is the Euclidean distance between observation i and regression location j. The kernel size is controlled by the parameter b_i (also known as bandwidth)

According to Fotheringham *et al.* (2003), bandwidth is defined as the optimum distance at which a location can still influence the observed location. Its size is important during model calibration since it determines the weights for the nearest spatial units. The optimal bandwidth was chosen using a bi-square adaptive kernel method. The GWR 4.0 software (Nakaya *et al.,* 2014) provides a setting for optimum bandwidth selection in the golden section search algorithm. The bandwidth of the kernel might be fixed and the distance determines the size of the kernel to the point of interest. The kernel is the same at any point in space, or adaptive, where the size of the kernel is determined by the number of neighbours at the point of interest. The adaptive kernel (bi-square function) was employed in this paper as it allows the weighting method to vary spatially according to the density of data. Several studies found that adaptive (optimal) bandwidth performs better than fixed bandwidth (Braun and Rousson, 2000; Davies and Lawson, 2019; Wang and Li, 2017).

Spatial heterogeneity maps

After implementing the GWPR model in MGWR2.2 and GWR4 software, the estimated local coefficients were calculated along with significance values for each one (local t-values) in each statistical area of Bologna. Finally, the spatial heterogeneity map of local coefficients and local t-values were plotted to show the non-stationarity and spatial variability of the parameters. The t-statistic used to examine the significance of the local β coefficients for the j^{th} parameter of the model was the following:

$$t_j(u_i) = \frac{\beta_j(u_i, v_i)}{se(\beta_j(u_i, v_i))} \quad (7)$$

where β_j is the model's j^{th} estimated parameter;the estimated parameter standard error; and $(u_i, v_i)$the geographical coordinates of reference point i.

Comparison of GWPR and GPR models

According to Nakaya *et al.* (2005, 2014), we used the following statistical approaches or bandwidth selection parameters to compare the performance and goodness of fit of the GWPR and GPR models: AICc, Bayesian information criterion (BIC), deviance and percent deviance explained. The model indicates the best model performance with the lowest AICc (Hadayeghi *et al.,* 2010; Fotheringham *et al.,* 2003; Nakaya *et al.,* 2005). GWPR was chosen as the final model form since it had the lowest AICc.

Results

Descriptive statistics

Figure 3 shows the aggregated COVID-19 infection rate for each of the 90 statistical areas from February 1st, 2020, to November 20th, 2021. The aggregation is based on the number of persons who lived in each of the 90 statistical areas at the time. Increasing infection infection is illustrated in the figure by the shift of shading from yellow to dark brown. The higher values were in the south-eastern districts and the lower in the north-western districts (Figure 3).

The number of people subjected to tests for COVID-19 in Bologna's 90 statistical areas between February 1st, 2020, and November 20th, 2021 were 154,544 and the cumulative positive rate amounted to 38,579 persons on the latter date, which represent 9.89 % of Bologna's population. The number of COVID-19 cases ranged from 1 to 55 in the different statistical areas (Table 1).

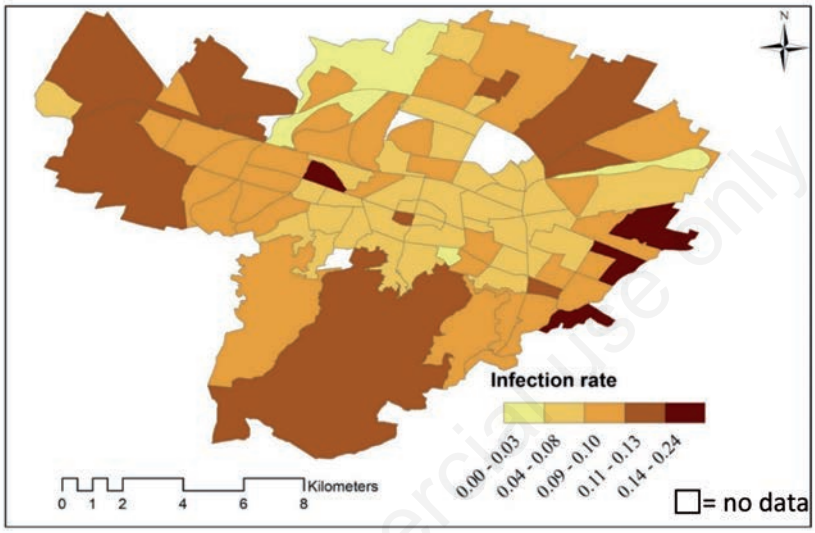

Figure 3. Distribution of the COVID-19 infection rate across Bologna' 90 statistical areas as of November 20th, 2021.

Table 1. Variables used in the analysis.

Variable	COVID infection (no.)	Percentage (%)	Infection rate (per 1,000 residents)
Age			
0-21	7,396	19.2	48
21-65	24,565	63.7	159
>65*	6,618	17.1	43
Gender			
Male	18,638	48.3	121
Female*	19,941	51.7	129
Family size			
1	8,462	21.9	55
2	8,364	21.7	54
3	8,661	22.5	56
≥4	13,092	33.9	85
Co-morbidity			
Any co-morbidity (yes/no*)	(7,388/31,191)	(19.2/80.8)	(48/202)
Hypertension (yes/no*)	(3,681/34,898)	(9.5/90.5)	(24/226)
Diabetes (yes/no)	(1,540/37,039)	(4.0/96.0)	(10/240)
Period			
First wave*	1,509	3.9	10
Second wave	15,696	40.7	102
Third wave	14,238	36.9	92
Out wave	7,136	18.5	46

*Indicates the reference categories of the covariates.

Comparison of COVID-19 infection predictors during the different waves

The descriptive comparison of COVID-19 cases during the different time points or epidemic waves over the entire study period, based on sociodemographic features, clinical features and case data, is shown in Table 2.

Spatial distribution of predictors in the area

To begin, we examined how the population was distributed or varied within each demographic covariate category for each statistical area. As shown in Figure 4, there is a variation of people living in the area for each sociodemographic category.

Vaccination status

As summed up in Table 3, people were said to be fully vaccinated when they had received their first and second dose of any vaccine by Pfizer, Moderna Biotech Spain, AstraZeneca S.P.A or one dose of Johnson and Johnson. Those who had received only one dose of Pfizer, Moderna Biotech Spain, or AstraZeneca S.P.A. were termed partially vaccinated individuals. Spatial distribution of Vaccination status in the 90 statistical areas of the population of Bologna are shown in Figure 5.

Testing for spatial variability

The results of the test for spatial variability using DIFF of Criterion and Monte Carlo Test are shown in Table S1 in Supplementary Materials. The Intercept, Second Wave, Third Wave, and Out Waves variables had negative values (based on DIFF of criterion result), indicating that those variables had significant local variation in the area. Similarly, the Intercept, Second Wave, Third Wave and Out Waves variables had a smaller p-value ($p<0.05$) in the Monte Carlo Test, indicating that the parameter estimates had significant local variation. As a result, a spatial variability local (varying) model was used for those parameters. In contrast, gender, age group, all family size categories, hypertension, diabetes, and co-morbidity variables showed positive values (see Table 4). Similarly, the p-values for variables >0.05, indicating that the parameter estimates do not exhibit significant local variation so that we used the global (fixed) model for the interpretation of the model result.

GPR estimates

Table 4 shows, based on the global GPR model result, that the intercept and age group (21–65 years) are statistically significant variables (at $p=0.05$) that were negatively related to COVID-19 infection rates. In contrast, hypertension, diabetes, and any co-morbidity, second wave, third wave, and out waves were statistically significant variables (at $p=0.05$) that positively affected the COVID-19 infection rate, while sex, the 0-20 years age group and family size did not significantly affect the COVID-19 infection rate. Comparisons were always made with reference categories.

Table 4 further shows that, globally, the estimated rate ratios for the intercept, the 21-65 age group, hypertension (present *vs* absent), diabetes (present *vs* absent), any comorbidity (present *vs* absent), the second wave, third wave and out waves were 0.074, 0.934, 1.275, 1.389, 1.149, 5.663 ,5.238 and 2.941, respectively, indicating that the average infection rates of COVID-19 decreased by 0.934-fold in the 21 – 65 years age group compared to the ≥65 years age group, while hypertension increased by 1.275-fold when compared to absence of hypertension and diabetes increased by 1.389-fold when compared to absence of diabetes, and the presence of any comorbidity increased by 1.149-fold when compared to a general absence of comorbidities. The infection rate of COVID-19 spiked (increased 5.663-fold) during the second wave compared to the first wave, 5.238-fold during third wave compared to the first wave and 2.941-fold during out waves compared to the first wave.

GWPR estimates

Based on the minimal AICc value, 4689 was calculated as the optimal bandwidth for the GWPR model. Table 4 shows the summary of parameter estimates in the local GWPR models. The effect

Table 2. Comparison of variables during the different COVID-19 waves in Bologna.

Variable	Total cases no. (%)	First wave no. (%)	Second wave no. (%)	Third wave no. (%)	Out waves no. (%)
Age					
0-21	7,396(19.2)	56(3.7)	2,524(16.1)	3,033(21.3)	1,783(25.0)
21-65	24,565(63.7)	1,053(69.8)	10,326(65.8)	8,856(62.2)	4,330(60.7)
≥65	656,618(17.1)	400(26.5)	2,846(18.1)	2,349(16.5)	1,023(14.3)
Gender					
Male	18,638(48.3)	594(39.4)	7,542(48.1)	6,944(48.8)	3,558(49.9)
Female	19,941(51.7)	915(60.6)	8,154(51.9)	7,294(51.2)	3,578(50.1)
Family size					
1	8,462(21.9)	513(34.0)	3,583(22.8)	2,887(20.3)	1,479(20.7)
2	8,364(21.7)	413(27.4)	3,538(22.5)	2,986(21.0)	1,427(20.0)
3	8,661(22.5)	273(18.1)	3,466(22.1)	3,341(23.5)	1,581(22.2)
≥4	1,3092(33.9)	310(20.5)	5,109(32.6)	5,024(35.2)	2,649(37.1)
Co-morbidity					
Any co-morbidity	7,388(19.2)	385(25.5)	3152(20.1)	2,678(18.8)	1,173(16.4)
None	31,191(80.8)	1,124(74.5)	12,544(79.9)	1,1560(81.2)	5,963(83.6)
Hypertension	3681(9.5)	219(14.5)	1,658(10.6)	1,297(9.1)	507(7.1)
Normal	34,898(90.5)	1,290(85.5)	14,038(89.4)	12,941(90.9)	6,629(92.9)
Diabetes	1,540(4.0)	86(5.7)	662(4.2)	582(4.1)	210(2.9)
Normal	37,039(96.0)	14,23(94.3)	15,034(95.8)	13,656(95.9)	6,926(97.1)
Total	38579(100)	1509(100)	15696(100)	14238(100)	7136(100)

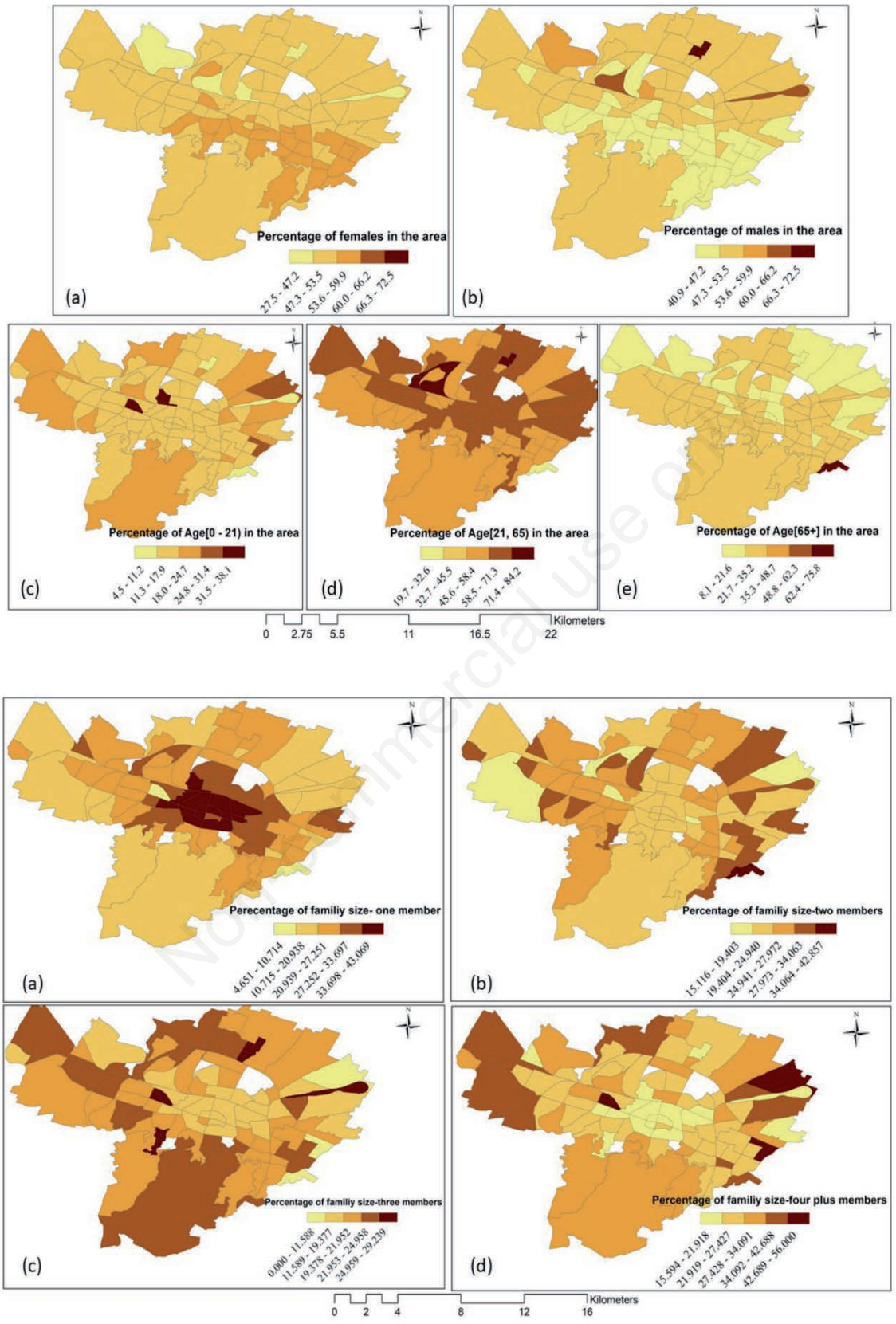

Figure 4. The spatial distribution of sociodemographic characteristics--percentage of residents living in the study area. White area = no data.

of some covariates varies across the study area. As a result, it is important to map the local parameter estimated to see where there is significant spatial heterogeneity between the independent and the dependent variables. The estimated coefficient of the local parameters are described by the summaries of central tendency and dispersion indices expressed by the mean, the standard the deviation, the minimum, the median and the maximum. Only four variables, including the intercept were finally accepted for GWPR calibration. The results of this model indicated a goodness-of-fit (R2) of 62.1% with an AICc = 6076.

Spatial heterogeneity maps

After estimating the GWPR parameters, the map of spatial variability for each GWPR parameter and a significant explanatory variables map was created by ArcMap10.8. Jenks' Natural Breaks algorithm (Slocum,1999) were used to classify each category. Class breaks are created in a manner that maximizes the differences between classes while grouping similar values. The features are classified into classes whose boundaries are determined by the differences in the data values. Local parameter estimate (beta-local coefficients) map plots for each variable that showed a significant variability are shown in Figure 6, and it is identified that the parameters have an obvious pattern of spatial non-stationarity. The values of local coefficients for the relationship between predictors and COVID-19 infection rate were not equal in all locations, with the direction and intensity of relationships changing depending on the geographical coordinates. The parameter of waves, the second wave ranges from 1.66 to 1.92, the third wave ranges from 1.61 to 1.93, and out waves ranges from 0.93 to 1.30. All the parameter signs are positive, indicating that all positively impacted the COVID-19 infection rates in the area. Whereas the baseline or intercept parameter value ranges from -2.74 to -2.54, the distribution implied that besides these three predictors, other factors were bound up with the COVID-19 infection rate. The colour code for the local parameter (the waves parameter) map plot indicates the estimated risk level for COVID-19 infection, where the high risk is represented by brown and decreases towards yellow (Figure 7).

Local estimate significance measure - the local t-value map plots are shown in Fig S1 in Supplementary Materials. A map showing a local parameter estimate significance measure - the spatial heterogeneity of local t-values estimated using the GWPR at 95% and 99% significance levels. For each of the 90 statistical area, local t-values greater than +1.96 and +2.58 indicate a positive and significant relationship between the predictors and COVID-19 infection rate; local t-values estimated between -1.96 and +1.96 indicate there is no significant relationship between the predictors and COVID-19 infection rate; and local t-values less than − 1.96 and − 2.58 indicate that there is a negative and statistically signif-

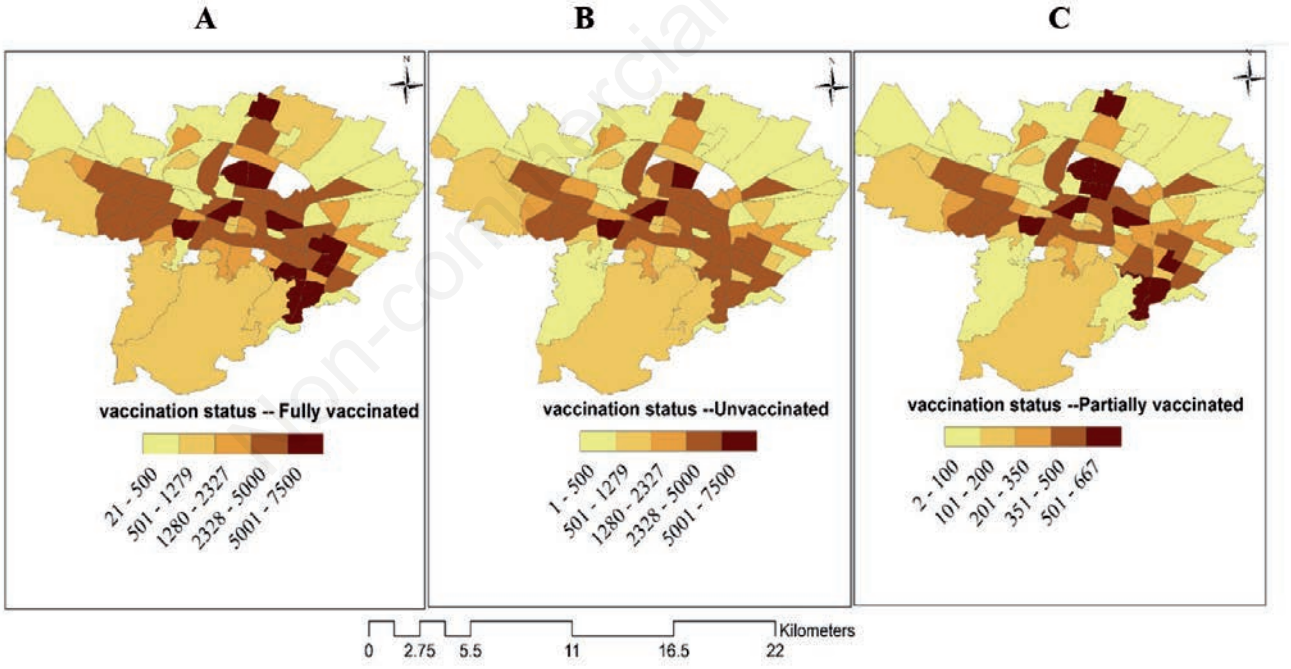

Figure 5. The spatial distribution of vaccinated people in the 90 statistical areas of Bologna. A) fully vaccinated; B) unvaccinated; C) partially vaccinated. White area = no data.

Table 3. Vaccination status of the Bologna population as of November 20th, 2021.

Status	People (no.)	Percentage (%)
Unvaccinated	161,561	41.22
Fully vaccinated (with or without booster)	210,180	53.63
Partially vaccinated	20,176	5.15

tiotemporal heterogeneity patterns were independent of the sociodemographic and clinical characteristics of the resident population. Presence of hypertension, diabetes and any co-morbidity were found to be significant, individual risk factors.

We used GIS and geographical data to explore the potential risk factors of COVID-19 infection rates in relation to sociodemographic factors, health risk factors, and epidemic waves. We compared the local GWPR model with a traditional GLM Poisson regression model to find the best fit for exploring the association between the available predictors and COVID-19 infection rates. The results showed that calibration of the GWPR model led to an improved fit compared to GLM Poisson regression.

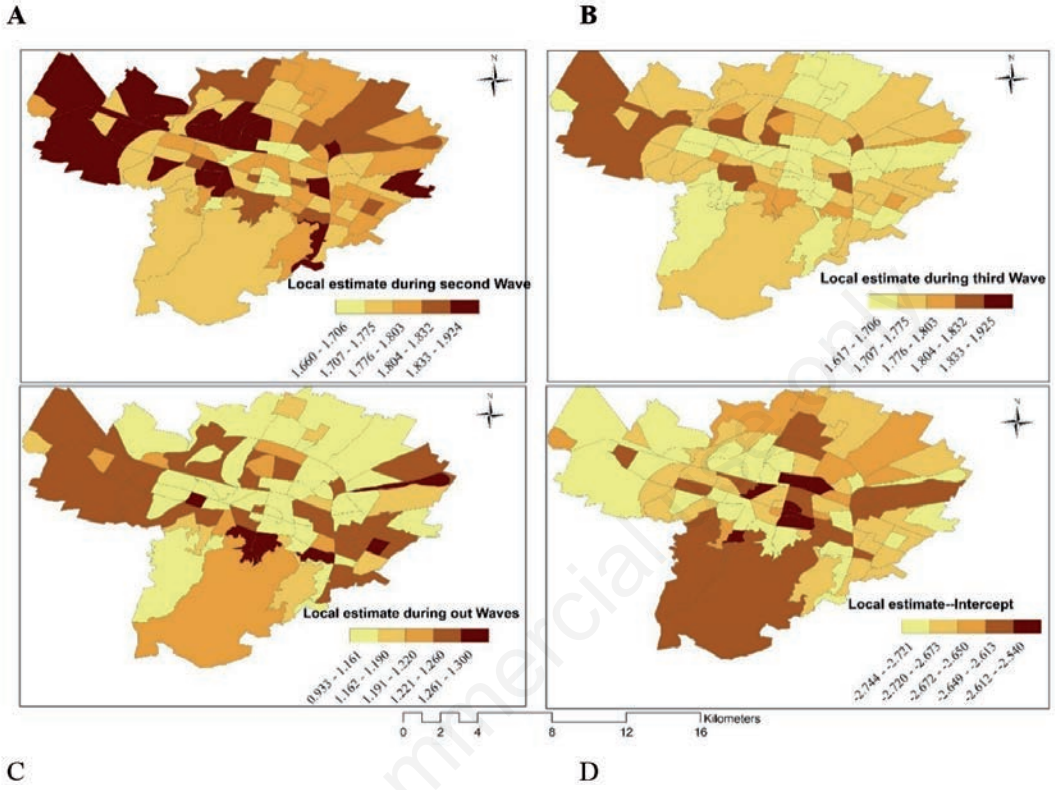

Figure 6. Spatial heterogeneity maps -the effects of predictors in describing COVID-19 infection rates at local level. Local parameter estimate /beta-local coefficients map plots; A) Second Wave; B) Third Wave; C) Out Waves; D) Intercept that showed aspatial variability.

Table 4. Summary statistics of the regression models.

Global model-GPR				Local model-GWPR				
Variable	Estimate	Rate ratio	P-value	Mean	STD	Min	Median	Max
Intercept	-2.601	0.074	<0.0001*	-2.653	0.050	-2.750	-2.650	-2.539
Male	0.014	1.014	0.166	-	-	-	-	-
Age [0–21)	-0.008	0.992	0.705	-	-	-	-	-
Age [21-65)	-0.068	0.934	<0.0001*	-	-	-	-	-
Family size-1	-0.007	0.993	0.655	-	-	-	-	-
Family size-2	-0.015	0.985	0.323	-	-	-	-	-
Family size-3	0.024	1.024	0.083	-	-	-	-	-
Hypertension	0.243	1.275	<0.0001*	-	-	-	-	-
Diabetes	0.329	1.389	<0.0001*	-	-	-	-	-
Any co-morbidity	0.139	1.149	<0.0001*	-	-	-	-	-
Second Wave	1.734	5.663	<0.0001*	1.786	0.064	1.656	1.778	1.815
Third Wave	1.656	5.238	<0.0001*	1.707	0.049	1.610	1.693	1.815
Out Waves	1.079	2.941	<0.0001*	1.138	0.087	0.894	1.153	1.260

GPR=global Poisson regression; GWPR=geographically weighted Poisson regression; STD=standard deviation; *Results statistically significant. Bandwidth size: 4689. The estimates are calculated for the reference categories (age 65+, family size - 4+, hypertension-no, diabetes-no, without any co-morbidities). As defined in the methodology, bandwidth is the optimum distance between a location and an observed location at which it is still possible to influence the location. The area-to-area influence was estimated to exert its effect over an area with 4.7 km radius.

icant relationship between the predictors and COVID-19 infection rate in each area. As shown in fig S1. in Supplementary Materials, all the Wave parameters are greater than +1.96 and +2.58, indicating that the t-value for this relationship is positive and statistically significant. However, we did not find a local t-value between -1.96 and +1.96, and the intercept parameter estimated value was lower than -1.96 and -2.58. This indicates that the predictors and COVID-19 infection rate have a negative and statistically significant relationship. According to the local R2 value in Fig S1 in Supplementary Materials top (right), the GWPR model fits better in the darker area (R^2>0.68).

Comparison of GWPR and GPR models

The performance of both GPR and local GWPR models is shown in Table 5. The GWPR model showed a smaller AICc, and larger percent deviance (it is equivalent to R^2, which explains how predictors affect outcome variables), so this was chosen as the best model. In our result, the best bandwidth size was 4689 with a minimum AICc of 6076. The results demonstrate that GWPR predicts COVID-19 infection rates in particular areas more accurately than GPR. In GWPR, the variability of COVID-19 infection rates across the statistical areas was better captured by the spatial heterogeneity in the data (Table 5).

Discussion

This is the first population-based study to look at how well GPR and local GWPR models can explain the variations of COVID-19 infection rates in a large city based on sociodemographic factors, health risk and different time periods-waves as explanatory variables. Different city areas were impacted differently during different epidemic waves. The area-to-area influence was estimated to be confined within a 4.7 km radius and these spa-

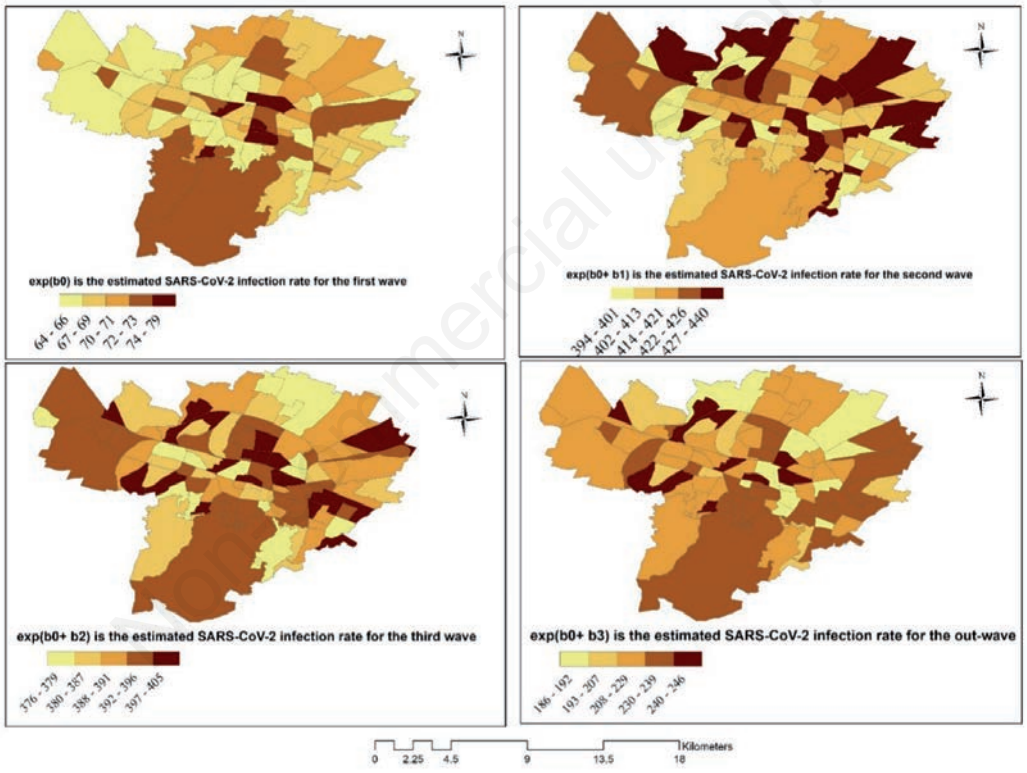

Figure 7. Estimated SARS-CoV-2 infection rates (per 1,000 inhabitants) for each statistical area and Wave. Estimated infection rates were calculated as exp(beta_i,0+beta_i,j), for j=2ndWave, 3rdWave, and OutWave.

Table 5. Comparison of model performance in the estimation of the effects of the explanatory variables with respect to the COVID-19 infection rate.

Model	AICc	Percent deviance explained
GPR model	6144	0.611
GWPR model	6076	0.621
Difference	68	0.010

Best bandwidth size: 4689 with minimum AICc : 6076.

The heterogeneous nature of COVID-19 transmission processes over space has been observed previously (Chen *et al.,* 2021; Wang *et al.,* 2020). Several factors contribute to COVID-19 transmission, including local demographic inequalities, health risk factors, different time periods or epidemic waves, and socioeconomic characteristics (Gatto *et al.,* 2020; Qiu *et al.,* 2019; Pedro *et al.,* 2020). Recent studies have found that the geo-epidemiological distribution of epidemic waves varies between countries, even between areas of limited size (Gaudart *et al.,* 2021). Geo-epidemiological analyses based on available data can therefore contribute greatly to the development of effective health policies (Hou *et al.,* 2021; Srivastava *et al.,* 2020) by focusing on the most impacted areas and developing tailored interventions.

Individuals, particularly the elderly, and if they have underlying co-morbidities, such as cardiovascular disease, diabetes and hypertension, are at higher risk for complications from COVID-19 (Klekotka *et al.,* 2015). In this research, key findings, based on GPR, were that sociodemographic, health risk factors, and epidemic waves were found to impact on COVID-19 infection rate, and based on the local model, COVID-19 infection rate varied geographically within the study area during epidemic Waves. More specifically, in the global model, the average COVID-19 infection rate decreased in age group (21 - 65) compared to age group (65 +), which is consistent with most previous findings (Dowd *et al.,* 2020; Shawky *et al.,* 2021; Sun *et al.,* 2020); whereas the average COVID-19 infection rate increased in hypertension – yes compared to hypertension – no, this is in line with other study (Landstra *et al.,* 2021), however, the opposite is observed in other studies (Fresán *et al.,* 2021; Jayaswal *et al.,* 2021); the average COVID-19 infection rate increased in diabetes – yes compared to diabetes - no, this is in line with most previous studies (Landstra *et al.,* 2021; Jayaswal *et al.,* 2021), also one review article (Nassar *et al.,* 2021) confirmed that diabetes mellitus is associated with a significant risk of complications, extended hospital stays, and mortality in COVID-19 infected patients ; the average COVID-19 infection rate increased in any comorbidities - yes compared to any comorbidities - no, this is in line with most previous studies (Landstra *et al.,* 2021); and the average COVID-19 infection rate increased significantly during the Second and Third Wave compared to the First Wave, also moderately increased during Out Waves compared to the First Wave, this is in line with other study (Gaudart *et al.,* 2020), in the Second and Third waves, it spiked significantly and showed spatial variability.

The period with the highest SARS-CoV-2 infection rate was the Second Wave (07 November 2020- 01 February 2021, 15696 cases). The rate was similar but lower during the 3rd Wave (02 February 2021 – 28 April 2021, 14238 cases) and lowest during the First Wave 26 March 2020 - 13 May 2020 ,1509 cases). The low rate during the First Wave is probably due to insufficient diagnostic capabilities at the beginning of the pandemic. Mostly, during the Second Wave, was the hardest hit and spiked significantly. This was due in part to the relaxation of the severe lockdown measures (Magnavita *et al.,* 2020). During these two waves (the second and the third), Italian cases reached their highest peak. However, after the Omicron variant, daily cases also spiked, even if the death rates were very low. The number of patients having health risk factors or proportion of the health risk factors (hypertension - yes, diabetes - yes, any comorbidities - yes) was very small compared to no categories like 9.5 %, 4 %, and 19.2 % respectively. As a result, we did not observe spatial variabilities in the area, whereas sex-male, age group [0, 20], and all three family size categories did not significantly affect the COVID-19 infection rate.

In contrast to global modelling, where the relationship between variables is constant across the study area, the results from GWPR overcame the disadvantages and drawbacks of global modelling. From the GWPR result, we only observed spatial variability in four parameters (the intercepts and the three Waves). For those parameters from spatial heterogeneity maps, COVID-19 infection rates were not equal in all locations, and it is identified that the parameters have obvious patterns of geographic variation. All the parameter signs of Waves were positive in all statistical areas indicating that all have positive impacts on the COVID-19 infection rates in the area, whereas the baseline or intercept signs were negative in all areas, implying that besides these three predictors, there were still other factors bound up with the COVID-19 infection rate. Showing epidemic trends in advance may provide information about geographic risks, social determinants, and health factors influencing COVID-19 transmission and how to respond to it.

Methodologically (Brunsdon *et al.,* 1996; Fotheringham *et al.,* 2003; Nakaya *et al.,* 2005), developed GWPR into a convenient yet powerful technique for exploring spatial non-stationarity and providing mappable statistics to visualize the spatial patterns of relationships between dependent and independent variables. Recent studies have suggested that GWPR is one of the geostatistical methods that should be promoted in health studies, given the locality of health outcomes (Young *et al.,* 2010). Furthermore, GWPR can potentially contribute significantly to health research, allowing researchers to understand etiology and spatial processes better, providing specific results beyond global models to facilitate place-specific health policy formation, and allowing scholars to investigate questions that cannot be answered with traditional (global) analytical models. Finally, as we mentioned above, GWPR is useful in a wide range of disciplines where spatial data are used and in applications where spatial non-stationarity is suspected (and should be confirmed).

Spatial modelling of disease epidemics can be useful for assessing where and why disease hotspots and clusters exist and providing explicit information about the spatial variation in disease incidence and transmission. However we cannot clearly identify outbreaks and clusters. Possible reasons: i) by visual inspection outbreaks/clusters are not evident because of the coarse spatiotemporal resolution (90 areas x 4 waves); ii) the range of human mobility (and, in turn, virus diffusion) is comparable to the size of the city of Bologna; iii) other analytical methods are needed for cluster identification. Similarly, little is known about how sociodemographic, epidemic, and risk factors affected COVID-19 infection rates within the Bologna area. As such, this study aimed to provide insights into the relationship between predictors and disease infection rates at the local level. Before our study , more detailed information about individual subjects(including pre-existing medical conditions, and detailed demographic information) with COVID-19 were not accessible in Bologna, thus no spatial or geographical analysis was undertaken so far. As a result, this research could be the first and most valuable.

As a research paper, we need to be aware of several limitations - the study involved a single centre at a city level, so the results cannot be generalized across all areas in Italy due to heterogeneity in practice, building infrastructure, population, and many other factors. Furthermore, obtaining all covariates other than COVID-19 case number was impossible. Furthermore, we could not obtain socio-economic information, detailed clinical characteristics, or detailed information about medical conditions and laboratory tests

for the patients in our study area in Bologna, which is likely to be one of the most influential factors affecting the rate of COVID-19 infection. Despite these limitations, our study can help in two ways: i) the results provide a clear understanding of the relationships between predictor characteristics and disease infection rates, both globally and at the local level in Bologna; ii) studies at the local level can also provide insight into national policies aimed at improving the health and welfare of citizens. Also the ideas presented here may provide a basis for evaluating public health policies using spatial model results.

Conclusions

Based on our results, the local GWPR model fits better than the traditional GLM – global. Different city areas were hit differently by SARS-CoV-2 infection during different epidemic waves. The area-to-area influence was estimated to be confined within a 4.7 km radius. These spatiotemporal heterogeneity patterns were independent of the sociodemographic and clinical characteristics of the resident population. Significant single-individual risk factors for detected SARS-CoV-2 infection were hypertension, diabetes, and comorbidities. On the basis of the global model results, sociodemographic factors, health risk factors, or pre-existing conditions, epidemic waves played a role in the COVID-19 infection rate variation, and the local model results confirmed that the COVID-19 infection rate did vary geographically during epidemic waves. When analysing the effectiveness of preventative programs and other health services, it is important to keep this spatial variability in mind. Moreover, we believe that this study can be used as a guide for other countries to understand the local spread of COVID-19, and the information we have obtained and the method we used could serve as a tool or as a reference in the event of similar epidemics in the future.

References

Ali K, Partridge MD, Olfert MR, 2007. Can geographically weighted regressions improve regional analysis and policy making? Int Reg Sci Rev 30:300-31.

Armocida B, Formenti B, Ussai S, Palestra F, Missoni E, 2020. The Italian health system and the COVID-19 challenge. Lancet Public Health 5:e253.

Benson T, Chamberlin J, Rhinehart I, 2005. An investigation of the spatial determinants of the local prevalence of poverty in rural Malawi. International Food Policy Research Institute. Washington DC, USA. Available from: http://gisweb.ciat.cgiar.org/povertymapping/download/Malawi.pdf

Braun JW, Rousson V, 2000. An autocorrelation criterion for bandwidth selection in nonparametric regression. J Statist Computat Simulat 68:89-101.

Brunsdon C, Fotheringham AS, Charlton ME 1996. Geographically weighted regression: a method for exploring spatial non-stationarity. Geogr Anal 28:281–98.

Bui LV, Mor Z, Chemtob D, Ha ST, Levine H, 2018. Use of Geographically Weighted Poisson Regression to examine the effect of distance on Tuberculosis incidence: A case study in Nam Dinh, Vietnam. PLoS One 13:e0207068.

CDC COVID-19 Response Team, 2020. Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) - United States, February 12-March 16, 2020. MMWR Morb Mortal Wkly Rep 69:343–6.

CDC, 2020. Coronavirus COVID-19: symptoms of coronavirus. Centers for Disease Control and Prevention. Accessed April 18, 2020. Available from: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html

Chen J, Liu L, Zhou S, Xiao L, Jiang C, 2017. Spatial variation relationship between floating population and residential burglary: a case study from ZG, China. ISPRS Int J Geo-Inf 6:246.

Chen M, Chen Y, Wilson JP, Tan H, Chu T, 2022. Using an eigenvector spatial filtering-based spatially varying coefficient model to analyze the spatial heterogeneity of COVID-19 and its influencing factors in mainland China. ISPRS Int J Geo-Inf 11:67.

Chen VY-J, Wu P-C, Yang T-C, Su H-J, 2010. Examining non-stationary effects of social determinants on cardiovascular mortality after cold surges in Taiwan. Sci Total Environ 408:2042–9.

Chen YY, Assefa Y, 2021. The heterogeneity of the COVID-19 pandemic and national responses: an explanatory mixed-methods study. BMC Public Health 21:835.

Choi M, Lee M, Lee MJ, Jung D, 2017. Physical activity, quality of life and successful aging among community-dwelling older adults. Int Nurs Rev 64:396–404.

Choi WS, Kang CI, Kim Y, Choi JP, Joh JS, Shin HS, Kim G, Peck KR, Chung DR, Kim HO, Song SH, Kim YR, Sohn KM, Jung Y, Bang JH, Kim NJ, Lee KS, Jeong HW, Rhee JY, Kim ES, Woo H, Oh WS, Huh K, Lee YH, Song JY, Lee J, Lee CS, Kim BN, Choi YH, Jeong SJ, Lee JS, Yoon JH, Wi YM, Joung MK, Park SY, Lee SH, Jung SI, Kim SW, Lee JH, Lee H, Ki HK, Kim YS; Korean Society of Infectious Diseases, 2016. Clinical presentation and outcomes of Middle East respiratory syndrome in the Republic of Korea. Infect Chemother 48:118–26.

Cox D, 1972. Regression models and life tables (with discussion). J Roy Stat Soc 34:187–202.

Davies TM, Lawson AB, 2019. An evaluation of likelihood-based bandwidth selectors for spatial and spatiotemporal kernel estimates. J Statist Computat Simulat 89:1131-52.

Dowd JB, Andriano L, Brazel DM, Rotondi V, Block P, Ding X, Liu Y, Mills MC, 2020. Demographic science aids in understanding the spread and fatality rates of COVID-19. Proc Natl Acad Sci U S A 117:9696-9698.

Fotheringham AS, Brunsdon C, Charlton M. 2003. Geographically weighted regression: The analysis of spatially varying relationships John Wiley & Sons.

Fotheringham AS, Kelly MH, Charlton M, 2013. The demographic impacts of the Irish famine: Towards a greater geographical understanding. Transact Inst Bri Geograph 38:221-237.

Fresán U, Guevara M, Trobajo-Sanmartín C, Burgui C, Ezpeleta C, Castilla J, 2021. Hypertension and related comorbidities as potential risk factors for COVID-19 hospitalization and severity: a prospective population-based cohort study. J Clin Med 10:1194.

Gatto M, Bertuzzo E, Mari L, Miccoli S, Carraro L, Casagrandi R, Rinaldo A, 2020. Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. Proc Natl Acad Sci U S A 117:10484-10491.

Gaudart J, Landier J, Huiart L, Legendre E, Lehot L, Bendiane MK, Chiche L, Petitjean A, Mosnier E, Kirakoya-Samadoulougou F, Demongeot J, Piarroux R, Rebaudet S,

2021. Factors associated with the spatial heterogeneity of the first wave of COVID19 in France: a nationwide geo-epidemiological study. Lancet Public Health 6:e222-e231.

Goovaerts, P. 2005. Analysis and detection of health disparities using geostatistics and a space-time information system: The case of prostate cancer mortality in the United States, 1970-1994. Proceedings of GIS Planet 2005. (May 30-June 2, 2005, Estoril, Portugal).

Hadayeghi A, Amer S. Shalaby, Bhagwant N. Persaud. 2009. Development of planning level transportation safety tools using Geographically Weighted Poisson Regression. Accid Anal Prev 42:676-88.

Haque U, Scott LM, Hashizume M, Fisher E, Haque R, Yamamoto T, Glass GE, 2012. Modelling malaria treatment practices in Bangladesh using spatial statistics. Malar J 201211:63.

Hou X, Gao S, Li Q, Kang Y, Chen N, Chen K, Rao J, Ellenberg JS, Patz JA, 2021. Intra county modelling of COVID-19 infection with human mobility: assessing spatial heterogeneity with business traffic, age, and race. Proc Natl Acad Sci U S A 118:e2020524118.

Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B, 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395:497-506.

Jayaswal SK, Singh S, Malik PS, Venigalla SK, Gupta P, Samaga SN, Hota RN, Bhatia SS, Gupta I, 2021. Detrimental effect of diabetes and hypertension on the severity and mortality of COVID-19 infection: A multi-center case-control study from India. Diabetes Metab Syndr 15:102248.

Klekotka RB, Mizgała E, Król W. 2015. The ethology of lower respiratory tract infections in people with diabetes. Pneumonol Alergol Pol 2015;83:401-8.

Landstra CP, de Koning EJP, 2021. COVID-19 and Diabetes: Understanding the Interrelationship and Risks for a Severe Course. Front Endocrinol (Lausanne) 12:649525.

Li Z, Wang W, Liu P, Bigham JM, Ragland DR, 2013. Using Geographically Weighted Poisson Regression for county-level crash modelling in California, Safety Sci 58:89-97.

Loubert L, 3.27 - Using GIS to Understand Schools and Neighborhoods. Comprehensive Geographic Information Systems. Elsevier, 2018, pp. 422-439.

Magnavita N, Sacco A, Chirico F, 2020. Covid-19 pandemic in Italy: pros and cons. Zdrowie Publicznei Zarządzanie 18:32-35.

Matthews SA, Yang TC. Mapping the results of local statistics: Using geographically weighted regression. Demogr Res 2012 26:151-166.

Mennis JL, Jordan LM, 2005. The distribution of environmental equity: exploring spatial non-stationarity in multivariate models of air toxic releases. Ann Ass Am Geographers doi:10.1111/j.1467-8306.2005.00459.x.

Mollalo A, Alimohammadi A, Shirzadi MR, Malek MR, 2015. Geographic information system-based analysis of the spatial and spatio-temporal distribution of zoonotic cutaneous leishmaniasis in Golestan Province, north-east of Iran. Zoonoses Public Health 62:18-28.

Mollalo A, Khodabandehloo E 2016. Zoonotic cutaneous leishmaniasis in northeastern Iran: a GIS-based spatio-temporal multicriteria decision-making approach. Epidemiol Infect 144:2217-29.

Nakaya T, Fotheringham AS, Brunsdon C, Charlton M, 2005. Geographically weighted Poisson regression for disease association mapping. Stat Med 24:2695-717.

Nakaya T. 2014. GWR4 User Manual: Windows application for geographically weighted regression modelling [Internet]. Available from: https://gwr.maynoothuniversity.ie/wp-content/uploads/2013/04/GWR4_Manual.pdf

Nassar M, Daoud A, Nso N, Medina L, Ghernautan V, Bhangoo H, Nyein A, Mohamed M, Alqassieh A, Soliman K, Alfishawy M, Sachmechi I, Misra A, 2021. Diabetes Mellitus and COVID-19: Review Article. Diabetes Metab Syndr 15:102268.

Oshan TM, Li Z, Kang W, Wolf LJ, Fotheringham AS, 2019. mgwr: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. ISPRS Int J Geo-Inf 8:269.

Pedro SA, Ndjomatchoua FT, Jentsch P, Tchuenche JM, Anand M and Bauch CT. 2020 Conditions for a Second Wave of COVID-19 Due to Interactions Between Disease Dynamics and Social Processes. Front Phys doi: 10.3389/fphy.2020.574514.

Poliart A, Kirakoya-Samadoulougou F, Ouédraogo M, Collart P, Dubourg D, Samadoulougou S, 2021. Using geographically weighted Poisson regression to examine the association between socio-economic factors and hysterectomy incidence in Wallonia, Belgium. BMC Womens Health 21:373.

Qiu Y, Chen X, Shi W, 2020. Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China. J Popul Econ 33:1127-1172.

Shawky M., Abdullah K, Alkhattab S, Adham S, Peter A. 2021. Sociodemographic determinants of COVID-19 incidence rates in Oman: Geospatial modelling using multiscale geographically weighted regression (MGWR). Sustain Cities Soc 65:102627.

Slocum, TA .1999. Thematic Cartography and Visualization. Prentice-Hall.

Specchia ML, Di Pilla A, Sapienza M, Riccardi MT, Cicchetti A, Damiani G, Group IR, 2021. Dealing with COVID-19 Epidemic in Italy: Responses from Regional Organizational Models during the First Phase of the Epidemic. Int J Environ Res Public Health 18:5008.

Srivastava A, Chowell G. 2020. Understanding Spatial Heterogeneity of COVID-19 Pandemic Using Shape Analysis of Growth Rate Curves. medRxiv [Preprint]. doi: 10.1101/2020.05.25.20112433.

Sun P, Lu X, Xu C, Sun W, Pan B. 2020. Understanding of COVID-19 based on current evidence. J Med Virol 2020;92:548-551.

Takele K, Zewotir T, Ndanguza D, 2019. Understanding correlates of child stunting in Ethiopia using generalized linear mixed models. BMC Public Health 19:626.

Ufficio Statistica Regionale Regione Emilia Romagna. 10 April 2019. Available from: https://sociale.regione.emilia-romagna.it/notizie/2019/aprile/la-fotografia-dei-cittadini-stranieri-a-bologna-150-le-nazionalita (in Italian).

Umpierrez GE, Isaacs SD, Bazargan N, You X, Thaler LM, Kitabchi AE.2002. Hyperglycemia: an independent marker of in-hospital mortality in patients with undiagnosed diabetes. J Clin Endocrinol Metab 87:978-82.

Wang Y and Teunis P. 2020. Strongly heterogeneous transmission of COVID-19 in mainland China: Local and regional variation.

Front Med 7:329.
Wang W, Li D, 2017. Structure identification and variable selection in geographically weighted regression models. J Statist Computat Simulat 87:2050-2068.
WHO, 2020a. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). Available from: https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf
WHO, 2020b. Coronavirus disease (COVID-19) outbreak. Available from: https://www.who.int/westernpacific/health-topics/coronavirus
Worldometers, "COVID-19 CORONAVIRUS PANDEMIC," 2021. Accessed: 07-April-2022. Available: https://www.worldometers.info/coronavirus/
Yang T-C, Teng HW, Haran M. 2009. The impacts of social capital on infant mortality in the U.S.: A spatial investigation. Applied Spatial Analysis and Policy 2:211-227.
Yang TC, Matthews SA, 2012. Understanding the non-stationary associations between distrust of the health care system, health conditions, and self-rated health in the elderly: a geographically weighted regression approach. Health Place 18:576-85.
Young, L.J., Gotway, C.A., 2010. Using geostatistical methods in the analysis of public health data: the final frontier. geoENV VII-Geostatistics for Environmental Applications 16, 89–98.
Zeleke AJ, Moscato S, Miglio R, Chiari L, 2022. Length of stay analysis of COVID-19 hospitalizations using a count regression model and quantile regression: a study in Bologna, Italy. Int J Environ Res Public Health 19:2224.
Zhang H, Liu Y, Chen F, Mi B, Zeng L, Pei L, 2021.The effect of sociodemographic factors on COVID-19 incidence of 342 cities in China: a geographically weighted regression model analysis. BMC Infect Dis 21:428.
Zhao F, Park N, 2004. Using geographically weighted regression models to estimate annual average daily traffic. In: Transportation Research Record 1879. Washington, DC: Transportation Research Board, National Research Council: 99-107.
Zheng X, Qin G, Tu D, 2017. A generalized partially linear mean-covariance regression model for longitudinal proportional data, with applications to the analysis of quality of life data from cancer clinical trials. Stat Med 36:1884-1894.
Zhou YB, Wang QX, Liang S, Gong YH, Yang MX, Chen Y, Nie SJ, Nan L, Yang AH, Liao Q, Yang Y, Song XX, Jiang QW, 2015. Geographical variations in risk factors associated with HIV infection among drug users in a prefecture in Southwest China. Infect Dis Poverty 4:38.