

# Usefulness of ecological mobility and socio-economic indicators in SARS-CoV-2 infection modelling: a French case study

1. Nicolas ROMAIN-SCELLE, MD<sup>1,2,3,\*</sup>

Corresponding author: [nicolas.romain-scelle@chu-lyon.fr](mailto:nicolas.romain-scelle@chu-lyon.fr)

ORCID: 0000-0002-9036-411X

2. Benjamin RICHE, PhD<sup>3</sup>

3. Thomas BENET, MD, PhD<sup>4</sup>

4. Muriel RABILLOUD, MD, PhD<sup>1,2,3</sup>

## Affiliations:

1. Université Lyon 1, Lyon, France

2. Laboratoire de Biométrie et Biologie Evolutive - UMR CNRS 5558, Villeurbanne, France

3. Service de Biostatistiques-Bioinformatique, Hospices Civils de Lyon, Lyon, France

4. Auvergne-Rhône-Alpes Regional Office, Santé publique France, Lyon, France.

Word count: 3,904

Keywords: COVID-19, Spatial statistics, Epidemiology, SARS-CoV-2, Emerging diseases

## Abstract:

*Introduction* Following its emergence in January 2020, SARS-CoV-2 diffusion occurred for a year with only non-pharmaceutical interventions (NPIs) available as mitigation tools. We aimed to assess the predictive capability of census-based indicators on the infection risk by SARS-CoV-2 in the French Auvergne-Rhône-Alpes region to assist NPIs allocation at the neighbourhood level. *Methods* We aggregated all counts of biologically confirmed cases of SARS-CoV-2 infection at the neighbourhood level between May 2020 and February 2021. 10 census-based ecological covariates were evaluated as predictors of case incidence using a Poisson regression with conditional autoregressive (CAR) spatial effects. Benefits of CAR effects and covariates on model fit were evaluated using pseudo-R<sup>2</sup> and Moran's I statistics. *Results* 438,992 infection cases over 5,410 neighbourhoods among 7,917,997 inhabitants were analysed. The association between covariates and case incidence was inconstant across time and space. Spatial correlation was estimated at high levels. Spatial CAR effects were necessary to improve on the pseudo-R<sup>2</sup> and the Moran's I statistics compared to the null model (intercept only). *Conclusion* The ecological covariates assessed were

31 insufficient to adequately model the distribution of cases at the neighbourhood level. Excess  
32 incidence was found mainly in metropolitan areas before the epidemic wave peak.

## 33 Introduction

34 The emergence of the SARS-CoV-2 virus and the COVID-19 disease in early 2020 has stressed  
35 the necessity of readiness in pandemic response, as no pharmaceutical measure or pre-existent  
36 immunity against the pathogen can be expected in the onset of an epidemic. Among all immediately  
37 available public health resources available to manage an emergent infectious disease outbreak,  
38 contact tracing have proven effective in mitigating reproduction rates, with varying successes  
39 depending on exhaustiveness, reactivity and resource allocation [1]. However, contact tracing is a  
40 labour-intensive activity, with limited scalability in short delays, as individuals cannot be quickly  
41 tasked from one job to another if prior training is required. As such, there is a necessity for policy  
42 makers to plan an early control in the case of emergent disease with pandemic potential, as contact  
43 tracing can be easily overwhelmed as the number of cases increase [2].

44 The first wave of the pandemic in France occurred between the first imported case in January  
45 24<sup>th</sup> 2020 and a near complete recess of viral circulation with a minimum infection incidence rate  
46 during the first week of June at 3,169 confirmed cases per 100,000 person-week. This inaugural wave  
47 impacted French regions varyingly: the north-east, and the Paris region of France were particularly  
48 marred with a high number of cases (up to 74 new hospital admissions per 100,000 person-week)  
49 compared to a more moderate impact in the north, south-east and centre regions, and a low impact  
50 on the Atlantic coast and associated in-land areas (10/100,000 person-week in the Brittany region).

51 Using the weekly infection cases incidence rate, a second wave period can be determined  
52 around the peak, which occurred the last week of October 2020 (335,204 cases in France). The  
53 incidence rate started to increase by the beginning of June 2020, and decreased back to a  
54 moderately high level of 72,758 cases in the first week of December 2020 before a slow increase to  
55 128,276 weekly cases by the first week of March 2021. This second wave was also marked by  
56 unequal regional impacts, with the south-eastern part of France experiencing the highest incidence  
57 rates in infection cases and hospital admissions (49 hospital admissions/100,000 person-week in the  
58 Auvergne-Rhône-Alpes region). All epidemiologic indicators are provided by the French Public Health  
59 Agency (Santé publique France) and available online [3]

60 Prognostic factors of COVID-19 disease, morbidity and mortality are an extensively studied  
61 topic [4]. Supplementary improvements could be obtained from a better understanding of the risk  
62 factors associated with the infection itself, as this understanding would allow for preventive public

63 health actions aimed toward at-risk individuals and populations, such as more intense contact-tracing  
64 effort and dedicated procurement of masks.

65 It is established that infectious diseases do not affect all equally. Beyond the potential  
66 existence of genetic risk factors, socio-economic determinants are known predictors of incidence of  
67 infection. In the viral world of infectious disease, influenza infections are associated with lower  
68 literacy levels, unemployment rates, and home ownership as shown by retrospective works  
69 conducted in US cities in 1918-1919 [5–8].

70 Several socio-economic factors are already proven or highly suspected of being associated  
71 with the risk of infection by SARS-CoV-2. A literature review from Khanijahani et al. (2021)[9] has  
72 shown that a low education level, poverty or deprivation, household overcrowding and low income  
73 are associated with an increase of SARS-CoV-2 infection incidence rate, however with some diverging  
74 results. Studies done specifically in France at the individual level has shown that the presence of a  
75 child attending school is a risk factor [10]. The same study found the education level as a risk factor,  
76 but with a higher risk of infection among individuals with both low and high education level.

77 As such, the present study aims at identifying relevant socio-economic predictors of an  
78 excess risk of infection by SARS-CoV-2 at a population level in order to optimise the use of public  
79 health interventions, guiding them toward more at-risk populations, therefore more effectively  
80 mitigate emergent infectious diseases.

## 81 **Materials and methods**

82 We conducted a retrospective, ecological analysis of the count of incident cases of infection  
83 by SARS-CoV-2 within the Auvergne-Rhône-Alpes French region, to assess the association between  
84 infection risk and socio-economic indicators from census data. The study region accounted for 12% of  
85 the total metropolitan French population, with 8.1 million inhabitants in 2022. Our study was  
86 conducted for a time period between May 13<sup>th</sup>, 2020 and February 14<sup>th</sup>, 2021, covering most of the  
87 second wave. Aggregated counts of confirmed cases of infection were used as an outcome.

### 88 **National information system for infection cases**

89 Infection cases data was extracted from the dedicated national infection case information  
90 system (SI-DEP). This system was deployed in France from May 13<sup>th</sup> 2020 in order to centralize all  
91 tests and results of biological analysis searching for SARS-CoV-2 infection, independently of  
92 technique used. Upload of tests results is mandatory for all the French medical analysis laboratories,  
93 making the system exhaustive. For each test carried out, the system collects the test result and the  
94 following information regarding the patient: age, sex, main postal address. The data are anonymized

95 and aggregated before the extraction for analysis, providing a count of people tested and infected by  
96 spatial unit. Due to both technical limitations at the time of extraction and a changing definition of  
97 the count of individual tested over the study period, only the counts of confirmed cases were  
98 available for this study.

### 99 Case definition

100 A case was defined by the confirmed infection of an individual by the SARS-CoV-2 virus by  
101 specific RT-PCR test or antigenic test. In case of repeatedly positive biological analyses, an individual  
102 was considered as a new case if the repeated test was positive more than 60 days after the previous  
103 one to exclude test positivity associated to viral material persistence at the sampling site.

### 104 Spatial unit

105 The IRIS (*Ilôt de Regroupement pour l'Information Statistique*), an infra-municipal spatial unit,  
106 was used for this ecological study as the aggregation scale for the case count of SARS-CoV-2 infection  
107 and the covariates. Each mainland French town (*commune*) counting 15,000 residents or more is  
108 split between multiples IRIS, with towns between 5,000 and 15,000 inhabitants being split if  
109 necessary. The composition target for each IRIS is a resident population of 1,500 and a coherence of  
110 habitation type.

### 111 Temporal unit

112 This study was conducted over four time periods, between May 13<sup>th</sup> 2020 and February 14<sup>th</sup>  
113 2021, defined empirically to capture four phases of the second wave corresponding respectively to  
114 extremely low case count following the first wave and lockdown, rapid case count growth, incidence  
115 peak and decrease, stabilization at an elevated level of weekly incidence. Data aggregation over time  
116 was necessary to limit the number of spatial units with case counts censored to prevent patients'  
117 identification. The four periods were defined as follow:

- 118 • May 13<sup>th</sup> to July 26<sup>th</sup> 2020 (weeks 20 to 30), period P1, said "Low incidence",
- 119 • July 27<sup>th</sup> to October 25<sup>th</sup> 2020 (weeks 31 to 43), period P2, said "Growth",
- 120 • October 26<sup>th</sup> to December 13<sup>th</sup> 2020 (weeks 44 to 50), period P3, said "Peak and decrease",
- 121 • December 14<sup>th</sup> 2020 to February 14<sup>th</sup> 2021 (weeks 2020-51 to 2021-06), period P4, said  
122 "Stabilization"

### 123 Inclusion and exclusion criteria

124 All 5,410 IRIS belonging to the Auvergne-Rhône-Alpes region of mainland France were  
125 included. Fifteen IRIS for which the number of inhabitants was 0 or near 0 were excluded of the

126 analysis (industrial/commercial areas with marginal habitations). All IRIS for which census-based  
127 indicators were unavailable were excluded from the analysis (341 units).

### 128 Socio-economic, mobility and population density indicators

129 The socio-economic indicators potentially associated with the distribution of infection cases  
130 were the proportion of migrants, the proportion of unemployed individuals, the proportion of single  
131 person homes, the proportion of households without a child, the proportion of car ownership, the  
132 proportion of individuals without a high school diploma, and the proportion of overpopulated homes  
133 (French National Institute of Statistics and Economic Studies, INSEE). Two other indicators were  
134 included to account for population mobility in volume and nature, i.e. the proportion of individuals  
135 working outside of their residency town, and the proportion of individuals travelling to work by car.  
136 Their distributions were described in Table 1, defined and represented using maps in Supplementary  
137 Material S1. To control for the urbanisation degree of the spatial units, a combined indicator in four  
138 levels measuring population density and urbanization, was included in all the models. A detailed  
139 definition of this indicator is provided in Table 1. This indicator was grouped into two levels for the  
140 purpose of the present analysis (Very high-High vs. Low-Very low). All mentioned indicators were  
141 obtained from the 2017 population census data. The socio-economic indicators were available at the  
142 IRIS level. The mobility and population density were available at the town level. All indicators are  
143 censored at the 1<sup>st</sup> and 99<sup>th</sup> centiles to control for extreme data issuing from low population areas.

### 144 Statistical analysis

145 To quantify the association between the socio-economic indicators and case incidence, four  
146 models described in Table 2 were carried out for each period. The first model (M1), serving as  
147 reference, was a generalized linear model with the counts of cases following a Poisson distribution,  
148 estimating the mean incidence rate across the region (intercept only). A second GLM model was  
149 carried out (M2), with the introduction of the studied covariates. A conditional autoregressive (CAR)  
150 random effect was added in the third (M3) and fourth models (M4), to account for the spatial  
151 autocorrelation of the case counts. The CAR random effect proposed by Leroux and MacNab [11,12]  
152 was used based on its capacity to provide estimates of all parameters with minimum bias in scenarios  
153 of both high and low spatial correlation compared to other CAR effects [13].

154 The CAR random effect follows a multinomial distribution, with two parameters to estimate,  
155 i.e. a spatial correlation coefficient  $\rho$  measuring the strength of the dependence among neighbouring  
156 spatial units, and a parameter  $\tau^2$  weighting the variance of the spatial effect component around its  
157 expectancy.

158 Parameters' estimate was obtained using a Bayesian approach, with weakly informative  
159 priors. All  $\beta$  parameters related to a continuous covariate and corresponding rate ratios are given for  
160 one standard deviation increase of the covariate. Incidence rate ratios (IRR), defined as the ratio  
161 between the predicted incidence rate within one spatial unit and the mean predicted incidence rate  
162 across the study region, were presented using maps. Values above 1 therefore indicate an excess  
163 incidence.

164 The neighbourhood structure between spatial units was defined on a travel-time-based  
165 metric. All IRIS within 60 minutes of each other were considered neighbours, with a weight defined  
166 as the inverse of the computed travel time separating each other. Normalization of the weight matrix  
167 imply a mean variance for the random effect of 1 when  $\tau^2$  is equal to 1.

168 Each model was evaluated by computing the Moran index of the residuals to assess whether  
169 the model was successful in accounting for the spatial autocorrelation. To assess goodness-of-fit, the  
170 observed case count in each IRIS was compared to the posterior distribution of the predicted count.  
171 A good fit was defined by an observed count in the interval defined by the 5<sup>th</sup> and 95<sup>th</sup> centiles of the  
172 posterior distribution. We used the pseudo- $R^2$  proposed by McFadden [14] to measure and compare,  
173 for each period, the variance explained by the models.

174 A sensitivity analysis was conducted to assess the validity of modelling over the entire study  
175 region. We fitted the full model in each of the 12 *Départements* (French administrative subdivision)  
176 composing the study region.

177 Detailed models' specification, address of the violation of the case independence hypothesis,  
178 weight matrix computation, and estimation procedure are reported in Supplementary Material 2. All  
179 analysis were conducted using R software (4.1.2). Travel times were computed using the r5r package  
180 [15] to interface with the Conveyal R5 routing engine [16]. Geographic data was extracted from the  
181 OpenStreetMap dataset. The model parameters were estimated using the package CARBayes [17].

## 182 Results

183 The study region is composed of 5,410 IRIS, with 5,069 for which all variables were available,  
184 and 7,917,997 individuals at risk (Table 1). The total number of cases observed during the study  
185 period was 562,376, of which 438,992 cases could be attributed to one IRIS and used in this study  
186 (78%).

187 Unless stated otherwise, all results regarding parameters estimates are extracted from the  
188 full model (M4). Across all periods, all socio-economic indicators were significantly associated with  
189 the rate of confirmed cases of SARS-CoV-2 infection (Table 3). Those associations were not

190 statistically significant for all periods, nor were they homogeneous. The effect of the proportion of  
191 low education level was dependent of the observation period: during the Growth period, it was  
192 associated with an incidence rate reduction of 10% for an increase of 11.6% of the proportion of low  
193 education level (rate ratio (RR): 0.90, 95% credibility interval (CI): [0.88;0.93]), but during the Peak  
194 and decrease, and Stabilization periods the estimated RR was above 1 (respectively 1.08, 95% CI:  
195 [1.05;1.10] and 1.06, 95% CI:[1.04;1.09]). The estimated RR for unemployment, statistically  
196 significant only during the Peak and decrease period, was under 1, corresponding to a decrease of 4%  
197 of the incidence rate for an increase of 5.7% of the proportion of unemployed individuals (RR: 0.96,  
198 95% CI: [0.93;0.99]). The proportion of migrants was the only indicator found to be associated with  
199 the risk of infection during the Low incidence period, corresponding to an increase of 15% of the  
200 incidence rate for an increase of 7.2% of the proportion of migrants (RR: 1.15, 95% CI: [1.04;1.29]).  
201 The increase of the proportion of individuals living alone and of car ownership were both associated  
202 with a significant incidence rate reduction during the Growth period and a significant increase during  
203 the Peak and decrease period.

204           Among mobility-related indicators, the proportion of individuals working outside their  
205 residency town was found positively associated with an increased risk of infection, significantly so  
206 during the Growth and Stabilization periods (RR: 1.04, 95% CI: [1.02;1.07] in both cases). The  
207 proportion of individuals travelling to work by car was positively associated with the incidence rate  
208 during the Peak and decrease period (RR: 1.06, 95% CI: [1.02;1.10]), but appeared negatively  
209 associated with the incidence rate during the other periods (not significantly).

210           A very high or high population density index was found to be associated with an increase of  
211 the incidence rate in comparison to a low or very low index, although statistical significance was  
212 reached only for the Growth and Stabilization periods (respectively RR: 1.29, 95% CI: [1.20;1.39] and  
213 RR: 1.09, 95% CI: [1.02;1.16]).

214           In the full model the spatial correlation coefficients were estimated at 0.49 and 0.60 for the  
215 Low incidence and Stabilization periods respectively, and 0.76 and 0.89 for the Growth and Peak and  
216 decrease periods, indicating a varying strength of dependence of the incidence rates between  
217 neighbours across periods. The variance weighting parameter was low for the Growth, Peak and  
218 decrease and Stabilization periods, between 0.26 and 0.34, but was found significantly higher during  
219 the Low incidence period (1.34, 95% CI: [1.15;1.56]).

220           Models' residuals for the Low incidence and Stabilization periods were not spatially  
221 correlated, per their respective Moran indices (maximum: 0.08) (Table 4). For the Growth period,  
222 both the inclusion of a spatial effect and the inclusion of covariates were able to mitigate the residual

223 autocorrelation indicating a spatial structure of the covariate distribution. The explained variance for  
224 the Growth period was also noticeably improved by the introduction of covariates (M2), with the  
225 pseudo-R<sup>2</sup> reaching 0.15. For the Peak and decrease and Stabilization periods, the proportion of  
226 variance explained improved significantly only with the addition of a spatial effect to the model, with  
227 no relevant difference between M3 and M4. During the Low incidence period, the pseudo-R<sup>2</sup> showed  
228 marginal improvement between M3 and M4, showing a small benefit from the covariates addition.  
229 Model fitting was satisfactory for M3 and M4 only during the last three periods (Supplementary  
230 Material S3).

231 Cartography of the predicted IRRs by the full model (M4) for each period is provided in Figure  
232 1 (high resolution version with metropolitan areas foci available in Supplementary Material S4). The  
233 “Low incidence” period was marked by a sharp distinction between urban centres exhibiting excess  
234 incidence and the rest of the region (Figure 1A). The “Growth” period displayed IRRs above 1 within  
235 and around urban areas mostly, showing a development of the second wave in a limited number of  
236 populated areas (Figure 1B). The maps for the “Peak and decrease” and “Stabilization” periods  
237 expressed a progressive diffusion of the elevated IRRs across the study region, with the last period  
238 showing a dispersed IRR spatial distribution (Figures 1C and 1D). A longitudinal gradient was  
239 observed in the first three periods, with IRRs under 1 preferentially in the western part of the study  
240 region and IRRs over 1 in the eastern part.

241 The sensitivity analysis revealed varying direction of association between the covariates and  
242 infection incidence across spatial subunits (Supplementary Material S5), and lower than average  
243 estimates of the spatial correlation in the Alps (northeastern part of the study region).

## 244 Discussion

245 This research found several associations between ecological factors and infection risk.  
246 Population concentration was found to be the most preeminent ecological characteristic to explain  
247 the case distribution, with a higher population density associated with an increased risk of infection,  
248 even more so during the build-up to the second wave peak in November 2020.

249 The results found regarding the socio-economic and mobility indicators are less clear, with  
250 only three indicators maintaining a relatively constant and significant effect across study periods: the  
251 proportion of immigrants, the proportion of individuals working outside their residency town, and  
252 the proportion of families without a child. The protective effect of families without a child may be  
253 related to the family size (i.e. number of family sources) or the lack of Covid-19, thus limiting  
254 household transmissions [18]. Our results are concordant with other findings that identified higher



255 risk of Covid-19 infection among immigrants [19]. The remaining indicators were either significant for  
256 a single period, or with changing directions relative to the risk of infection.

257           The association between education level and infection risk at an ecological level suggests a  
258 less at-risk behaviour within units with a higher proportion of high education individuals during the  
259 Growth period. This may be associated with the spatial distribution of elevated IRRs during that  
260 time, as higher education levels tend to be observed in urban areas. This could also be associated  
261 with a higher mobility and social activities of more educated individuals, especially during the  
262 summer and long vacation periods in a similar fashion as what Plümper and Neumayer [20] reported  
263 during the first COVID-19 wave in Germany.

264           A behavioural mechanism may also be at play regarding unemployment rate. This indicator is  
265 considered as a risk factor for worst health outcomes and is a component of several deprivation  
266 indexes [21,22]. However, in our study the unemployment rate was found to be associated to a lower  
267 risk of infection during the Peak and decrease period. The unemployment may be acting as a proxy  
268 for social isolation in this work more than a deprivation measure, as discussed by Scarpone et al.  
269 [23]. It is in contrast with the role of single-person households, associated with an excess of risk  
270 solely during the Peak and decrease period, when this indicator is expected to be a protective one,  
271 since it captures the absence of household contamination risks.

272           The changing estimated role of home overcrowding is surprising, as this indicator is firmly  
273 associated with deprivation and worst health outcomes [9,24–27], and is a proxy of promiscuity. As  
274 such, the excess risk found for the Low incidence period is coherent with previous literature, and  
275 suggest the interest of directing prevention and infection control resources toward those highly  
276 vulnerable households in the early phase of an epidemic wave.

277           In this study, a limited association between the socio-economic indicators and the rate of  
278 SARS-CoV-2 cases was found at the ecological level. Although several indicators were found to be  
279 associated with the incidence rate at the IRIS level, it is dubious that any form of inference can be  
280 conducted regarding the risk at the individual level. Previous studies on COVID-19 and other  
281 infectious diseases clearly suggests that the socio-economic level of an individual is associated with  
282 the risk of infection [6,7,10,28,29]. Those results are in line with the more general association  
283 between socio-economic position and multiple health outcomes such as all-cause mortality or  
284 prevalence of chronic diseases [30–32].

285           The present findings suggest that one must proceed with care in interpreting the results of  
286 ecological studies involving an infectious disease and socio-economic indicators when working with  
287 small-area units (around 1,000 inhabitants). As shown in the present study, models which did not

288 account for the spatial autocorrelation of incidence rates failed to sufficiently explain the observed  
289 distribution of cases. In contrast, the overall strong spatial correlation associated with a low variance  
290 weighting that was found suggests that during the second wave of the COVID-19 pandemic in the  
291 study region, the stronger determinant of incidence rate in an IRIS was the incidence rates in its close  
292 neighbours, measurably above all the socio-economic indicators studied in this research. From a  
293 planning perspective, this result indicates the necessity of a cartographic approach to the allocation  
294 of infection control resources or implementation of measures to limit contacts (i.e. curfew) in an  
295 emergent pandemic situation more than a socio-economic-based approach at the population level.  
296 The present literature remains scarce regarding small-area modelling of infectious disease, limiting  
297 our ability to compare the present results appropriately. We identified a handful of authors reporting  
298 on small-area analysis about COVID-19 incidence, but without providing a quantitative estimation of  
299 the explanatory power of the analysis [33–35]. One study conducted in New York found a significant  
300 association between testing for infection and socio-economic position with an estimated  $R^2$  around  
301 0.3, but with spatial units ranging in population between  $10^3$  and  $10^5$  [36], compromising  
302 comparability with our work.

303 Our study has several limitations. First, for reasons of personal data protection, we were  
304 unable to conduct an analysis standardized for age and sex. More than 30% of cases observed during  
305 the entire study period would have been censored. Estimates of age-related risk of infection suggests  
306 a upward trend in risk with increasing age [37], indicating a potential excess of cases in units where  
307 the elderly are overrepresented, and the opposite in areas inhabited by a younger population. With  
308 regards to the sex-related risk, other studies found variable results, with a tendency toward a higher  
309 risk of positivity among males compared to females [38,39].

310 Our second caveat is the absence of availability of the counts of individuals tested for SARS-  
311 CoV-2 infection. Indeed, assuming a homogeneous testing rate across all spatial unit is a strong  
312 hypothesis [40]. Moreover, similar work on French data has established a variation of the testing rate  
313 according to socio-economic level, with a decrease in testing with the increase in deprivation [35]. It  
314 is therefore reasonable to suspect the counts of infection cases to be underestimated in more  
315 deprived areas, which could lead to bias the estimates of the association between incidence rates  
316 and socio-economic indicators.

317 In conclusion, although this study identified several socio-economic indicators associated  
318 with the risk of SARS-CoV-2 infection, they did not prove sufficient to properly explain the spread of  
319 the virus during the second wave of the pandemic in one French region. Response to an emerging  
320 disease with human to human transmission should be organised around proper cartographic data to

321 quickly identify excesses of cases in low incidence phases and to mitigate the impact of untargeted  
322 civil liberties restrictions.

## 323 Statements & Declarations

### 324 Funding

325 The authors declare that no funds, grants, or other support were received during the  
326 preparation of this manuscript.

### 327 Competing interests

328 The authors have no relevant financial or non-financial interests to disclose.

### 329 Author Contributions

330 All authors contributed to the study conception and design. Material preparation, data  
331 collection and analysis were performed by Nicolas ROMAIN-SCELLE. The first draft of the manuscript  
332 was written by Nicolas ROMAIN-SCELLE and all authors commented on previous versions of the  
333 manuscript. All authors read and approved the final manuscript.

### 334 Ethics approval

335 We used only aggregated and statistically censored data for the purpose of this analysis. No  
336 ethic committee was solicited for the purpose of this study.

### 337 Data accessibility

338 The infection cases data that support the findings of this study are available from the French Health  
339 Data Hub within the SI-DEP database. Restrictions apply to the availability of these data, which were  
340 used under licence for this study. Data may be requested by the following procedure:  
341 <https://www.demarches-simplifiees.fr/commencer/soumission-d-un-projet-de-recherche-etude-ou->  
342 [evalu.](https://www.demarches-simplifiees.fr/commencer/soumission-d-un-projet-de-recherche-etude-ou-)

343 The census data that support the findings of this study are openly available on the INSEE website at  
344 the following links: <https://www.insee.fr/fr/statistiques/4799268>,  
345 <https://www.insee.fr/fr/statistiques/4799323>, <https://www.insee.fr/fr/statistiques/4799252>,  
346 <https://www.insee.fr/fr/statistiques/4799305>.

347 The administrative layout data that support the findings of this study are openly available on the  
348 INSEE website at the following link: <https://www.insee.fr/fr/information/2017499>

349 The geographical data that support the findings of this study are openly available on the Geofabrik  
350 website at the following links: <https://download.geofabrik.de/europe/france.html>

351 Codes and specific datasets (with the exception of the infection cases dataset) are available on  
352 request to the corresponding author.

- 353 1. **Gardner BJ, Kilpatrick AM.** Contact tracing efficiency, transmission heterogeneity, and  
354 accelerating COVID-19 epidemics. *PLoS computational biology* 2021; **17**: e1009122.
- 355 2. **Lash RR, et al.** COVID-19 Case Investigation and Contact Tracing in the US, 2020. *JAMA network*  
356 *open* 2021; **4**: e2115850.
- 357 3. **Santé publique France.** *Indicateurs*: cartes, données et graphiques. *Géodes*.  
358 (<https://geodes.santepubliquefrance.fr/#view=map2&c=indicator>). Accessed 1 June 2023.
- 359 4. **Zhang J, et al.** Risk and Protective Factors for COVID-19 Morbidity, Severity, and Mortality.  
360 *Clinical Reviews in Allergy & Immunology* 2023; **64**: 90–107.
- 361 5. **Chowell G, Viboud C.** Pandemic influenza and socioeconomic disparities: Lessons from 1918  
362 Chicago. *Proceedings of the National Academy of Sciences National Academy of Sciences*, 2016;  
363 **113**: 13557–13559.
- 364 6. **Grantz KH, et al.** Disparities in influenza mortality and transmission related to sociodemographic  
365 factors within Chicago in the pandemic of 1918. *Proceedings of the National Academy of Sciences*  
366 *National Academy of Sciences*, 2016; **113**: 13839–13844.
- 367 7. **Hutchins SS, et al.** Protection of Racial/Ethnic Minority Populations During an Influenza  
368 Pandemic. *American Journal of Public Health American Public Health Association*, 2009; **99**:  
369 S261–S270.
- 370 8. **Mamelund S-E, Shelley-Egan C, Rogeberg O.** The association between socioeconomic status and  
371 pandemic influenza: protocol for a systematic review and meta-analysis. *Systematic Reviews*  
372 2019; **8**: 5.
- 373 9. **Khanijahani A, et al.** A systematic review of racial/ethnic and socioeconomic disparities in  
374 COVID-19. *International Journal for Equity in Health* 2021; **20**: 248.
- 375 10. **Galmiche S, et al.** *Etude des facteurs sociodémographiques, comportements et pratiques*  
376 *associés à l'infection par le SARS-CoV-2 (ComCor)* [report]. Institut Pasteur; Caisse Nationale  
377 d'Assurance Maladie; IPSOS; Institut Pierre Louis d'Epidémiologie et de Santé Publique  
378 (IPLESP); Santé Publique France, 2021.
- 379 11. **Leroux BG, Lei X, Breslow N.** *Estimation of Disease Rates in Small Areas: A new Mixed Model for*  
380 *Spatial Dependence*. In: Halloran ME, Berry D, eds. *Statistical Models in Epidemiology, the*  
381 *Environment, and Clinical Trials*. New York, NY: Springer, 2000, p. 179–191.
- 382 12. **MacNab YC.** Hierarchical Bayesian Modeling of Spatially Correlated Health Service Outcome and  
383 Utilization Rates. *Biometrics* 2003; **59**: 305–315.
- 384 13. **Lee D.** A comparison of conditional autoregressive models used in Bayesian disease mapping.  
385 *Spatial and Spatio-temporal Epidemiology* 2011; **2**: 79–89.
- 386 14. **McFadden D.** *Conditional logit analysis of qualitative choice behavior*. *Frontiers in econometrics*.  
387 New York: Academic Press, 1974, p. 105–142.
- 388 15. **Rafael H. M. Pereira, et al.** r5r: Rapid Realistic Routing on Multimodal Transport Networks with  
389 R5 in R. *Findings* 2021; Published online: 2021.doi:10.32866/001c.21262.

- 390 16. **Contributors**. *Conveyal R5 Routing Engine. GitHub Repo*. 2022(<https://github.com/conveyal/r5>).  
391 Accessed 27 July 2022.
- 392 17. **Lee D**. CARBayes: An R Package for Bayesian Spatial Modeling with Conditional Autoregressive  
393 Priors. *Journal of Statistical Software* 2013; **55**: 1–24.
- 394 18. **Madewell ZJ, et al**. Household Secondary Attack Rates of SARS-CoV-2 by Variant and Vaccination  
395 Status: An Updated Systematic Review and Meta-analysis. *JAMA Network Open* 2022; **5**:  
396 e229317.
- 397 19. **Gosselin A, et al**. Higher risk, higher protection: COVID-19 risk among immigrants in France—  
398 results from the population-based EpiCov survey. *European Journal of Public Health* 2022; **32**:  
399 655–663.
- 400 20. **Plümper T, Neumayer E**. The pandemic predominantly hits poor neighbourhoods? SARS-CoV-2  
401 infections and COVID-19 fatalities in German districts. *European Journal of Public Health* 2020;  
402 **30**: 1176–1180.
- 403 21. **Pornet C, et al**. Construction of an adaptable European transnational ecological deprivation  
404 index: the French version. *Journal of Epidemiology and Community Health* 2012; **66**: 982–989.
- 405 22. **Rey G, et al**. Ecological association between a deprivation index and mortality in France over the  
406 period 1997 – 2001: variations with spatial scale, degree of urbanicity, age, gender and cause of  
407 death. *BMC Public Health* 2009; **9**: 33.
- 408 23. **Scarpone C, et al**. A multimethod approach for county-scale geospatial analysis of emerging  
409 infectious diseases: a cross-sectional case study of COVID-19 incidence in Germany. *International*  
410 *Journal of Health Geographics* 2020; **19**: 32.
- 411 24. **Nkosi V, et al**. Overcrowding and health in two impoverished suburbs of Johannesburg, South  
412 Africa. *BMC public health* 2019; **19**: 1358.
- 413 25. **Singh A, et al**. Housing Disadvantage and Poor Mental Health: A Systematic Review. *American*  
414 *Journal of Preventive Medicine* 2019; **57**: 262–272.
- 415 26. **Swope CB, Hernández D**. Housing as a determinant of health equity: A conceptual model. *Social*  
416 *Science & Medicine (1982)* 2019; **243**: 112571.
- 417 27. **Ahmad K, et al**. Association of poor housing conditions with COVID-19 incidence and mortality  
418 across US counties. *PloS One* 2020; **15**: e0241327.
- 419 28. **Urbán R, Király O, Demetrovics Z**. Who complies with coronavirus disease 2019 precautions and  
420 who does not? *Current Opinion in Psychiatry* 2021; **34**: 363–368.
- 421 29. **Niedzwiedz CL, et al**. Ethnic and socioeconomic differences in SARS-CoV-2 infection: prospective  
422 cohort study using UK Biobank. *BMC Medicine* 2020; **18**: 160.
- 423 30. **Kivimäki M, et al**. Association between socioeconomic status and the development of mental  
424 and physical health conditions in adulthood: a multi-cohort study. *The Lancet Public Health*  
425 Elsevier, 2020; **5**: e140–e149.

- 426 31. **Foster H, et al.** Understanding the influence of socioeconomic status on the association between  
427 combinations of lifestyle factors and adverse health outcomes: a systematic review protocol.  
428 *BMJ Open* British Medical Journal Publishing Group, 2021; **11**: e042212.
- 429 32. **Mackenbach JP, et al.** Determinants of the magnitude of socioeconomic inequalities in  
430 mortality: A study of 17 European countries. *Health & Place* 2017; **47**: 44–53.
- 431 33. **Arauzo-Carod J-M, Domènech A, Gutiérrez A.** Do local characteristics act in a similar way for the  
432 first two waves of COVID-19? Analysis at intraurban level in Barcelona. *Journal of Public Health*  
433 2021; **43**: 455–461.
- 434 34. **Fernández-Martínez NF, et al.** Socioeconomic differences in COVID-19 infection, hospitalisation  
435 and mortality in urban areas in a region in the South of Europe. *BMC Public Health* 2022; **22**:  
436 2316.
- 437 35. **Vandentorren S, et al.** The effect of social deprivation on the dynamic of SARS-CoV-2 infection in  
438 France: a population-based analysis. *The Lancet Public Health* Elsevier, 2022; **7**: e240–e249.
- 439 36. **Douglas JA, Subica AM.** COVID-19 treatment resource disparities and social disadvantage in New  
440 York City. *Preventive Medicine* 2020; **141**: 106282.
- 441 37. **Goldstein E, Lipsitch M, Cevik M.** On the Effect of Age on the Transmission of SARS-CoV-2 in  
442 Households, Schools, and the Community. *The Journal of Infectious Diseases* 2021; **223**: 362–369.
- 443 38. **Prinelli F, et al.** Sex- and gender-related differences linked to SARS-CoV-2 infection among the  
444 participants in the web-based EPICOVID19 survey: the hormonal hypothesis. *Maturitas* 2022;  
445 **158**: 61–69.
- 446 39. **Vahidy FS, et al.** Sex differences in susceptibility, severity, and outcomes of coronavirus disease  
447 2019: Cross-sectional analysis from a diverse US metropolitan area. *PLOS ONE* Public Library of  
448 Science, 2021; **16**: e0245556.
- 449 40. **Padellini T, et al.** Time varying association between deprivation, ethnicity and SARS-CoV-2  
450 infections in England: A population-based ecological study. *The Lancet Regional Health. Europe*  
451 2022; **15**: 100322.

452

453

454

## Tables

455 Table 1: Description of the demographic and socio-economic characteristics of the IRIS (infra-  
456 municipal spatial units) of the French region Auvergne-Rhône-Alpes

Characteristics	N <sup>a</sup> = 5,069
Number of individuals at risk, median (IQR)	1,149 (444, 2,403)
Population density index <sup>b</sup> (4 levels) n (%)	
Very low	863 (17)
Low	2,249 (44)
High	1,053 (21)
Very high	904 (18)
% of immigrants, median (IQR)	4 (2, 9)
% of individuals unemployed, median (IQR)	9.8 (7.3, 13.6)
% of single person homes, median (IQR)	31 (25, 39)
% of individuals working outside their residency town, median (IQR)	59 (37, 72)
% of individuals travelling to work by car, median (IQR)	83 (73, 89)
% of families without a child, median (IQR)	50 (45, 57)
% of car ownership, median (IQR)	93 (87, 96)
% of low education, median (IQR)	55 (48, 62)
% of overpopulated homes, median (IQR)	1.60 (0.00, 3.48)

457 <sup>a</sup> number of spatial units with available data

458 <sup>b</sup>The density index is defined with respect to the proportion of the population of a town living  
459 in an urban centre (50,000 people living within a contiguous area with a population density above  
460 1,500 ind/km<sup>2</sup> or more), an urban cluster (5,000 people, 300 ind/km<sup>2</sup> or more), or an intermediary  
461 rural aggregate (300 people, 25 ind/km<sup>2</sup> or more): 50% or more of the population is within a urban  
462 centre: Very high ; 50% or more of the population lives within a urban centre or a urban cluster: High ;  
463 50% or more of the population lives outside a centre, cluster or aggregate: Very low ; Otherwise: Low

464

465 Table 2: Statistical models used to analyse the count of COVID-19 cases per IRIS (infra-communal  
466 spatial units) of the French region Auvergne-Rhône-Alpes during the second epidemic wave of  
467 COVID-19, with their respective labels and detailed composition

Label	Model <sup>a</sup>	Description
M1: Null	$Y_k \sim \text{Poisson}(\lambda_k) ; \ln(\lambda) = \mu$	Intercept only: $\mu$
M2: Covariates only	$Y_k \sim \text{Poisson}(\lambda_k) ; \ln(\lambda) = \mu + X\beta$	X: vector of covariates, $\beta$ : vector of regression parameters
M3: Spatial effect only	$Y_k \sim \text{Poisson}(\lambda_k) ; \ln(\lambda) = \mu + \phi$	Intercept and spatial effect ( $\phi$ )
M4: Full	$Y_k \sim \text{Poisson}(\lambda_k) ; \ln(\lambda) = \mu + X\beta + \phi$	Covariates and spatial effect

468 <sup>a</sup>  $Y_k$ : count of cases in IRIS k

469 Table 3: Effect of the demographic and socio-economic characteristics of the IRIS (infra-municipal  
 470 spatial units) of the French region Auvergne-Rhône-Alpes on the rate of SARS-CoV-2 infection (rate  
 471 ratio estimates) during the four periods of the second epidemic wave: results of Poisson models with  
 472 inclusion of covariates and a conditional autoregressive (CAR) random effect (models M4)

Variable	Period			
	Low incidence	Growth	Peak and decrease	Stabilization
Mean predicted weekly incidence <sup>a</sup>	0.01 [0.00;0.01]	0.68 [0.66;0.70]	2.47 [2.41;2.54]	1.28 [1.24;1.31]
Very High or High population density index (ref. Low or Very low)	1.24 [0.99;1.56]	<b>1.29</b> <b>[1.20;1.39]</b>	1.05 [0.98;1.11]	<b>1.09</b> <b>[1.02;1.16]</b>
<b>Socio-economic indicators</b> (for 1 standard deviation increase)				
% of low education (SD: 11.6%)	0.96 [0.87;1.05]	<b>0.90</b> <b>[0.88;0.93]</b>	<b>1.08</b> <b>[1.05;1.10]</b>	<b>1.06</b> <b>[1.04;1.09]</b>
% of single person homes (SD: 10.6%)	0.98 [0.88;1.09]	<b>0.93</b> <b>[0.90;0.97]</b>	<b>1.07</b> <b>[1.04;1.10]</b>	0.98 [0.95;1.01]
% of car ownership (SD: 10.4%)	1.08 [0.92;1.27]	<b>0.89</b> <b>[0.85;0.93]</b>	1.01 [0.97;1.05]	1.00 [0.96;1.04]
% of immigrants (SD: 7.2%)	<b>1.15</b> <b>[1.04;1.29]</b>	<b>1.06</b> <b>[1.02;1.10]</b>	<b>1.04</b> <b>[1.01;1.08]</b>	1.01 [0.98;1.05]
% of individuals unemployed (SD: 5.7%)	1.11 [0.98;1.25]	1.01 [0.98;1.05]	<b>0.96</b> <b>[0.93;0.99]</b>	1.02 [0.98;1.05]
% of individuals working outside their residency town (SD: 20.6%)	1.09 [0.99;1.19]	<b>1.04</b> <b>[1.02;1.07]</b>	1.01 [0.99;1.04]	<b>1.04</b> <b>[1.02;1.07]</b>
% of individuals travelling to work by car (SD: 15.9%)	0.86 [0.73;1.01]	0.99 [0.94;1.03]	<b>1.06</b> <b>[1.02;1.10]</b>	0.99 [0.95;1.02]
% of families without a child (SD: 9.9%)	0.95 [0.85;1.05]	<b>0.94</b> <b>[0.91;0.97]</b>	<b>0.96</b> <b>[0.93;0.99]</b>	0.99 [0.97;1.02]
% of overpopulated homes (SD: 2.8%)	<b>1.15</b> <b>[1.03;1.28]</b>	<b>0.96</b> <b>[0.94;0.99]</b>	<b>0.97</b> <b>[0.95;0.99]</b>	0.98 [0.95;1.01]
<b>Spatial effect parameters</b>				
Spatial correlation coefficient ( $\rho \in (0,1)$ )	0.485 [0.300;0.656]	0.761 [0.680;0.832]	0.888 [0.835;0.931]	0.603 [0.529;0.669]
Variance weighting parameter ( $\tau^2 \in \mathbb{R}_+$ )	1.342 [1.150;1.564]	0.340 [0.319;0.363]	0.256 [0.239;0.274]	0.284 [0.268;0.302]

473 *\*The point estimate and the 95% credible interval is given for each model parameter and each study*  
 474 *period. Apart from the base risk and the spatial effect parameters, estimates significantly different*  
 475 *than their neutral values are emboldened.*

476 <sup>a</sup> *Incidence rate of SARS-CoV-2 infection in an IRIS of 1,000 inhabitants, low population density and*  
 477 *mean covariates level (/1,000 person-week).*

478

479 Table 4: Moran indices and McFadden pseudo-R<sup>2</sup>, for each model fitted on the complete dataset, by  
 480 epidemic period

Period	Model	McFadden R <sup>2</sup> <sup>a</sup>	Moran I <sup>b</sup>
Low incidence	Null		0,048
	Covariates only	0,044	0,017
	Spatial effect only	0,416	0,076
	Full	0,441	0,027
Growth	Null		0,410



	Covariates only	0,150	0,139
	Spatial effect only	0,688	0,088
	Full	0,687	0,044
Peak and decrease	Null		0,230
	Covariates only	0,057	0,186
	Spatial effect only	0,680	0,045
	Full	0,680	0,037
Stabilization	Null		0,052
	Covariates only	0,029	0,042
	Spatial effect only	0,565	0,042
	Full	0,567	0,026

481

482 <sup>a</sup> Pseudo-R<sup>2</sup> are computed using the corresponding null model as reference.

483 <sup>b</sup> All Moran indices are significant (at a threshold of 0.05).

484

485

486

## Figures

487 Figure 1: predicted incidence rate ratio of SARS-CoV-2 infection for the IRIS (infra-municipal spatial  
488 unit) of the French region Auvergne-Rhône-Alpes during the second epidemic wave: model with  
489 covariates and spatial effect (M4). From A to D: Low incidence, Growth, Peak and decrease, and  
490 Stabilization periods.

491

492

## 493            Statements & Declarations

### 494            Funding

495            The authors declare that no funds, grants, or other support were received during the  
496 preparation of this manuscript.

### 497            Competing interests

498            The authors have no relevant financial or non-financial interests to disclose.

### 499            Author Contributions

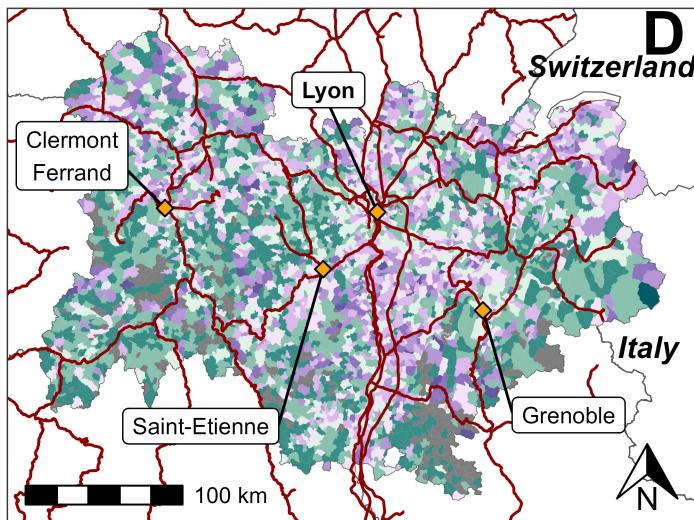
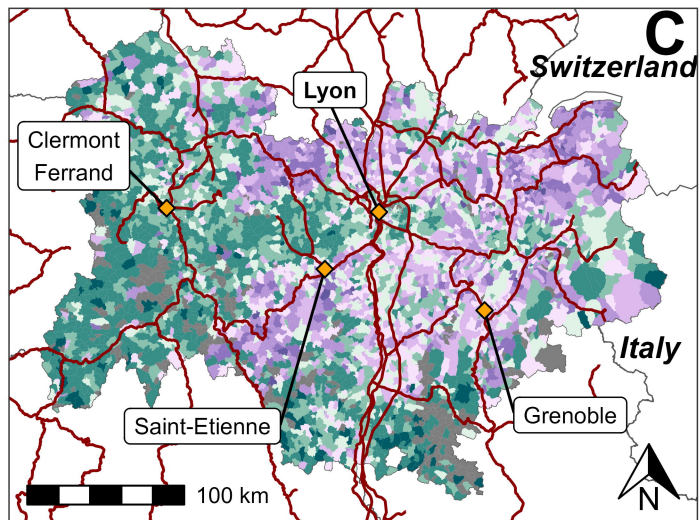
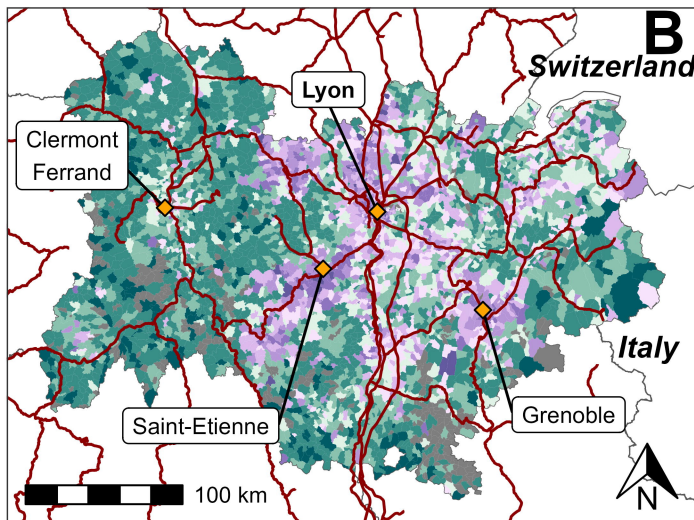
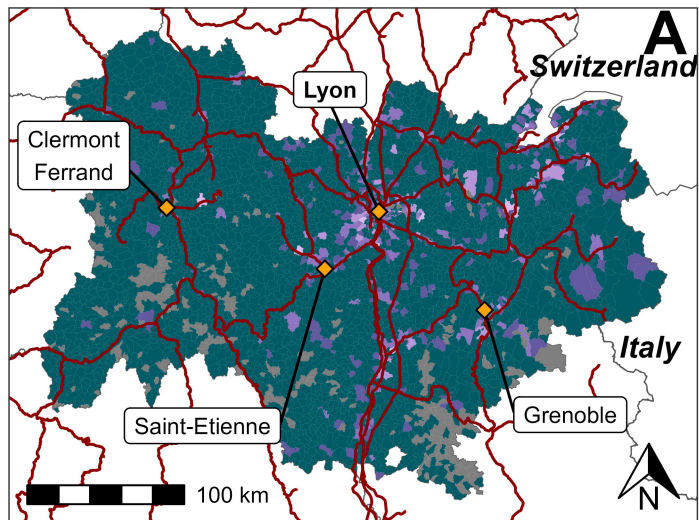
500            All authors contributed to the study conception and design. Material preparation, data  
501 collection and analysis were performed by Nicolas ROMAIN-SCELLE. The first draft of the manuscript  
502 was written by Nicolas ROMAIN-SCELLE and all authors commented on previous versions of the  
503 manuscript. All authors read and approved the final manuscript.

### 504            Ethics approval

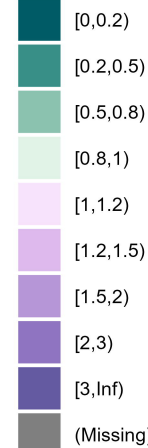
505            We used only aggregated and statistically censored data for the purpose of this analysis. No  
506 ethic committee was solicited for the purpose of this study.

### 507            Acknowledgments

508            We thank the DATA direction at Santé publique France for providing us with the *ad hoc*  
509 dataset, and Dr. Christine Saura from Santé publique France for her input during the elaboration of  
510 this study.



Incidence rate ratio



Railways