

RESEARCH

Open Access



# Investigating the drivers of the spatio-temporal heterogeneity in COVID-19 hospital incidence—Belgium as a study case

Simon Dellicour<sup>1,2\*</sup> , Catherine Linard<sup>3,4</sup>, Nina Van Goethem<sup>5</sup>, Daniele Da Re<sup>6</sup>, Jean Artois<sup>1</sup>, Jérémie Bihin<sup>3</sup>, Pierre Schaus<sup>7</sup>, François Massonnet<sup>6</sup>, Herman Van Oyen<sup>5,8</sup>, Sophie O. Vanwambeke<sup>6</sup>, Niko Speybroeck<sup>9</sup> and Marius Gilbert<sup>1</sup>

## Abstract

**Background:** The COVID-19 pandemic is affecting nations globally, but with an impact exhibiting significant spatial and temporal variation at the sub-national level. Identifying and disentangling the drivers of resulting hospitalisation incidence at the local scale is key to predict, mitigate and manage epidemic surges, but also to develop targeted measures. However, this type of analysis is often not possible because of the lack of spatially-explicit health data and spatial uncertainties associated with infection.

**Methods:** To overcome these limitations, we propose an analytical framework to investigate potential drivers of the spatio-temporal heterogeneity in COVID-19 hospitalisation incidence when data are only available at the hospital level. Specifically, the approach is based on the delimitation of hospital catchment areas, which allows analysing associations between hospitalisation incidence and spatial or temporal covariates. We illustrate and apply our analytical framework to Belgium, a country heavily impacted by two COVID-19 epidemic waves in 2020, both in terms of mortality and hospitalisation incidence.

**Results:** Our spatial analyses reveal an association between the hospitalisation incidence and the local density of nursing home residents, which confirms the important impact of COVID-19 in elderly communities of Belgium. Our temporal analyses further indicate a pronounced seasonality in hospitalisation incidence associated with the seasonality of weather variables. Taking advantage of these associations, we discuss the feasibility of predictive models based on machine learning to predict future hospitalisation incidence.

**Conclusion:** Our reproducible analytical workflow allows performing spatially-explicit analyses of data aggregated at the hospital level and can be used to explore potential drivers and dynamic of COVID-19 hospitalisation incidence at regional or national scales.

**Keywords:** COVID-19, Hospitalisation incidence, Spatial covariates, Temporal covariates, Boosted regression trees, Belgium

## Background

COVID-19 (coronavirus disease 2019), caused by a new coronavirus (severe acute respiratory syndrome coronavirus 2; SARS-CoV-2), was first reported in early

\*Correspondence: simon.dellicour@ulb.ac.be

<sup>1</sup> Spatial Epidemiology Lab (SpELL), Université Libre de Bruxelles, 50 av. FD

Roosevelt, 1050, CP160/12 Bruxelles, Belgium

Full list of author information is available at the end of the article



© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

December 2019 in the province of Hubei in China [1] and followed by a fast and extensive spread throughout the world [2]. On March 12, 2020, the World Health Organisation (WHO) declared COVID-19 a pandemic. The COVID-19 pandemic has resulted in considerable public health, social and economic disruptions [3, 4]. Moreover, the pandemic has threatened the saturation or actually led to the saturation of national hospitalisation systems. In order to mitigate and manage epidemic surges, it is of strategic importance to disentangle the spatio-temporal dynamic and potential drivers of COVID-19 hospitalisation incidence. In this context, one practical limitation can be the lack of access to spatially-explicit health data like, for instance, the geographic origin of hospitalised patients. To circumvent this issue, we here propose an analytical framework based on the definition of hospital catchment areas (HCAs), a concept that has previously been used to study the geographic access to healthcares in the context of the COVID-19 epidemic in Brazil [5]. The overall objective of our study is to present and describe an analytical framework based on the delimitation of such HCA units and that allows to analyse the dynamic of potential drivers of COVID-19 hospitalisation incidence in a spatially-explicit context. As a study case, we illustrate our approach on the study of the COVID-19 epidemic during the year 2020 in Belgium, for which we precisely had access to hospitalisation incidence data at the hospital level.

In Belgium, the first COVID-19 case was officially reported on February 4, 2020, in an asymptomatic Belgian citizen repatriated from Wuhan [6]. Symptomatic cases increased from March onwards, initially related to the return of holidaymakers, mainly from Italy [7]. This was soon followed by local transmission and extensive spread in the country [8]. The Federal Government implemented social-distancing measures and eventually a full lockdown during March–May 2020, which led to the end of the first wave. The summer months showed a lower incidence of positive cases and hospitalisations, in Belgium as in other European countries. However, September 2020 saw a slow but steady growth of the registered cases and new hospitalisations, which resulted in the start of the second wave at the end of September and brought the Federal Government to tighten up the social distancing measures at the end of October. With now more than 20,000 confirmed COVID-19 deaths, Belgium has been highly impacted by the two epidemic waves occurring during spring and fall 2020, with a mortality rate among the highest in the world. While explanatory factors like population density, connectivity, age pyramid or the critical circulation of the virus in nursing homes [9] have been discussed, the exact causes of one the highest COVID-19 mortality in the

world (172 deaths/10,000 people, <https://coronavirus.jhu.edu/data/mortality>) has still to be more thoroughly understood, which could require analytical comparisons between countries. As illustrated by the New York Times in an article [10] highlighting excess mortality likely due to underreporting COVID-19 deaths, such a comparison will however be complicated by differences among countries in terms of reporting of COVID-19 mortality. In Belgium, the COVID-19-induced mortality has been monitored relatively well, as reflected by the high correlation between excess mortality statistics for 2020 and the COVID-19 recorded deaths [11]. Besides a high mortality, with more than 50,000 hospital admissions, Belgium also experienced an important hospital incidence during the two first waves of the COVID-19 epidemic, with marked variations from one hospital to another. During the peak surge, the intensive care unit (ICU) thresholds were locally exceeded, causing the need to supplement ICU beds in emergency. On many occasions, these overflows adversely impacted the fate of critically-ill COVID-19 patients [12]. With several weeks of hindsight and the availability of quality-controlled data, we can now investigate the potential drivers of the spatial–temporal heterogeneity in COVID-19 hospitalisation incidence in Belgium.

In the present study, we specifically aim at (i) comparing the spatial patterns of hospitalisation incidence between the two epidemic waves; (ii) understanding if spatial variations in hospitalisation incidence is related to population, economic, or environmental covariates (population density, number of nearby nursing home residents, median age, percentage of workers working in the primary/secondary/tertiary sectors, median declared income, ratio of urban areas, local concentration of fine particles); and (iii) exploring if temporal variations in hospitalisation incidence might be related to some temporal covariates such as human mobility, climatic factors (temperature, humidity, and solar radiation), or environmental factor (air quality).

## Methods

### Definition and delimitation of hospital catchment areas

The Belgian national public health research institute (Sciensano) collected and aggregated COVID-19 hospitalisation data according to an established hospital surveillance protocol [13]. Although some data are available at the municipality level, other important data sets were only collected per hospital ( $n=103$ ). This forced us to work at the level of HCAs. The hypothesis underlying the concept of HCA is that patients normally choose the hospital closest to their home. While the simplest approximation of spatial accessibility is Euclidean distance, a more realistic way to generate accessibility maps is to

estimate travel times between populations and hospitals. In practice, we used the friction surface developed by Weiss et al. [14] to derive maps of travel time between any 1 km<sup>2</sup> pixel in Belgium and each of the 103 hospitals. Each pixel was assigned the hospital that minimised such travel time, and the set of pixels attributed to a given hospital was defined as its catchment area. In areas where multiple hospitals could be found within the same pixel of the friction surface, catchment areas and associated epidemiological data were aggregated ( $n=2$ ). The resulting HCAs were used as working units in all subsequent analyses based on three different time periods: a time period corresponding to the first (01/03–31/05/2020) and second (01/09–30/11/2020) epidemic waves, as well as a time period covering the entire Belgian epidemic until the end of November 2020 (01/03–30/11/2020). Specifically, we worked with a measure of hospitalisation incidence (HI) computed as the cumulative number of new hospitalisations per 100,000 inhabitants for each HCA and time period.

#### Investigating the drivers of the spatial heterogeneity in COVID-19 hospitalisation incidence

We investigated the extent to which the HI heterogeneity among HCAs could be explained by spatial covariates using three different categories of analyses: (i) visual explorations based on principal component analyses (PCAs), (ii) univariate correlation analyses summarised with a correlogram and estimated through univariate linear regression (ULR) analyses, and (iii) multivariate analyses conducted with multivariate linear regression (MLR) and machine learning approaches. The machine learning approach consisted in using a boosted regression tree (BRT) algorithm [15] allowing the estimation of non-linear relationships between response (HI) and predictive variables (spatial covariates) by generating a collection of sequentially fitted regression trees to optimise the predictions. The PCA was performed with the `dudi.pca` function of the `ade4` R package, and the correlogram visualisation was generated with the `heatmap.2` function of the R package `gplots`. We also performed spatial autocorrelation tests using the `Moran.I` function of the R package `ape`. Regarding BRT analyses, they were carried out using the R package `dismo`. To select the optimal number of trees in the BRT models, we used a cross-validation procedure based on 5 separated folds. All BRT analyses were run and averaged over 10 cross-validated replicates, with a tree complexity set at 5, an initial number of trees set at 10, a learning rate of 0.005, and a step size of 5.

We analysed eleven distinct spatial covariates as drivers of the spatial heterogeneity of HI (Fig. 1): human population density computed from the human population

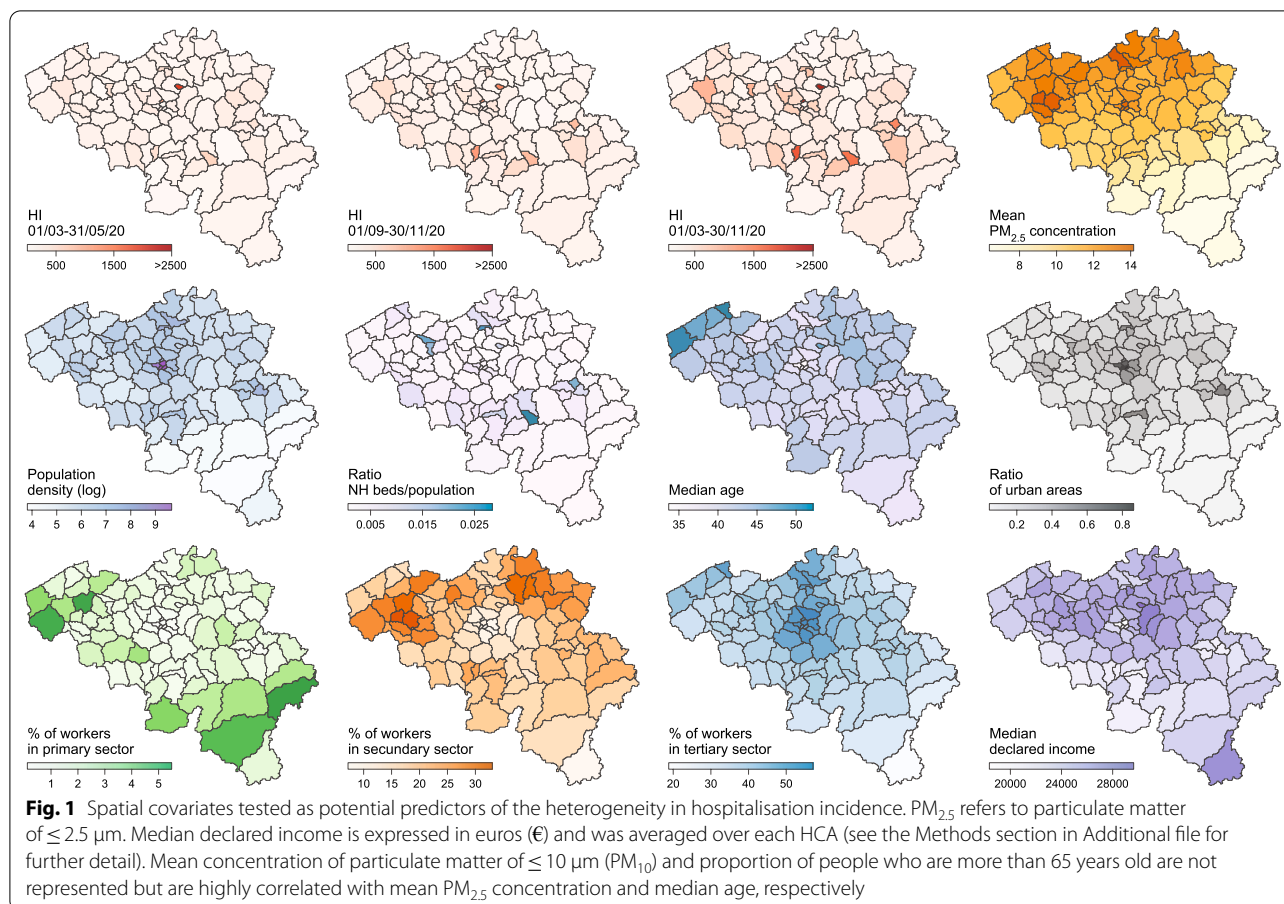
raster made available by the WorldPop project (<https://www.worldpop.org>), the median age and the proportion of people older than 65 years, the median income, the percentages of workers in the primary, secondary or tertiary sector, all provided by the Belgian Federal Public Service Economy (SPF Economie), the ratio nursing home (NH) beds/population computed from the number of NH beds provided at municipal level by the Belgian Federal social care agencies (<https://www.zorg-en-gezondheid.be> for Flanders, <http://sante.wallonie.be> for Wallonia, and <http://www.iriscare.brussels> for Brussels), the concentration of particulate matters  $\leq 2.5$  or  $10 \mu\text{m}$  ( $\text{PM}_{2.5}$  and  $\text{PM}_{10}$ ) provided by the Belgian Interregional Environment Agency, and the proportion of urban areas computed from the European land use raster provided by the Corine Land Cover database (<https://land.copernicus.eu>). Median age, proportion of people >65 years old, median income, percentage of workers in the primary, secondary, or tertiary sectors, and the number of NH beds were initially provided at the municipality level. We transposed these values to the HCA level by weighting the contribution of each municipality according to the number of people living in areas overlapped by both a given municipality and given HCA polygon. In practice, the value  $v_{HCA}$  assigned to a given HCA was computed as follows:

$$v_{HCA} = \sum_i \left( v_{municipality,i} \times \left( \frac{p_{overlap,i}}{p_{HCA}} \right) \right) \quad (1)$$

where  $v_{municipality,i}$  is the covariate value available for the municipality  $i$ ,  $p_{overlap,i}$  is the number of people living in the area overlapped by the HCA and municipality  $i$ , and  $p_{HCA}$  is the total number of people living in the HCA under consideration. Estimates of  $\text{PM}_{2.5}$  and  $\text{PM}_{10}$  concentrations were provided in the form of geo-referenced grids averaged over the entire year 2017. We used the function `exact_extract` from the R package `exactextractr` to extract and assign PM concentration values to each HCA.

#### Investigating the drivers of the temporal variability in COVID-19 hospitalisation incidence

We subsequently investigated the extent to which HCA specific daily hospitalisation ratios (DHRs) could be related to temporal covariates. DHR values were computed on a time window of 7 days. For a given HCA, a DHR was defined by the ratio of total number of persons hospitalised at day  $d$  divided by the total number of persons hospitalised at day  $d-7$ , and are thus by essence temporally autocorrelated (Additional file 1: Fig. S1). We analysed five distinct temporal covariates (Additional file 1: Fig. S1): an index of human mobility,



PM<sub>2.5</sub> concentration, and three climatic variables (temperature, relative humidity, and solar radiation). Index values of human mobility were based on anonymised mobile phone data provided by Proximus, the largest mobile telecommunications company in Belgium. Proximus provided a mobility index per day and municipality computed as the ratio between the number of trips “in” or “out” the municipality, divided by the number of subscriptions for the municipality. The number of trips “in” and “out” were here defined as the number of journeys inside and outside the considered municipality, respectively. Specifically, the number of trips “in” measures the number of mobile phone SIM cards “living” in the considered municipality, and the number of trips “out” measures the number of mobile phone SIM cards not “living” in but “visiting” that municipality. The number of subscriptions is defined as the number of mobile phone SIM cards “living” in a municipality. Following data privacy protection, Proximus only provided a number of trips “in”/“out” when a sufficient number of at least 30 SIM cards were involved in a daily trip between two municipalities. The resulting index values of mobility were transposed from the municipality to the HCA level

using the procedure described above. Estimates of PM<sub>2.5</sub> concentration were obtained from the Belgian Interregional Environment Agency and provided in the form of daily geo-referenced grids. We again used the function `exact_extract` from the R package `exactextractr` to extract and assign PM<sub>2.5</sub> concentration values to assign to each HCA.

Three atmospheric variables were considered to explain the observed variations in DHR. All three variables were retrieved from the ERA5 climate reanalysis [16]. A climate reanalysis consists of outputs from a numerical model of the atmosphere that has been constrained by in-situ measurements like sea-level pressure, near-surface air temperature and wind speed (among others), all obtained from in-situ measuring devices such as thermometers in weather stations or radiosondes operating on weather balloons. The prime advantage of using reanalyses is that they provide gridded information at high temporal frequency (up to hourly), which is not possible from the few weather stations available in Belgium alone. In the context of this study, for which regional scale near-real time estimates of atmospheric variables were required, the ERA5 reanalysis data (~35 km resolution)



appeared as a better choice than the scattered point weather station data. The three variables retrieved from ERA5 are: (i) two-meter air temperature (units: K), (ii) the surface direct short-wave radiation flux (units:  $W m^{-2}$ ), (iii) the relative humidity at 1000 hPa, defined as the ratio between the water vapour pressure and the saturation water vapour pressure (in %).

**Results**

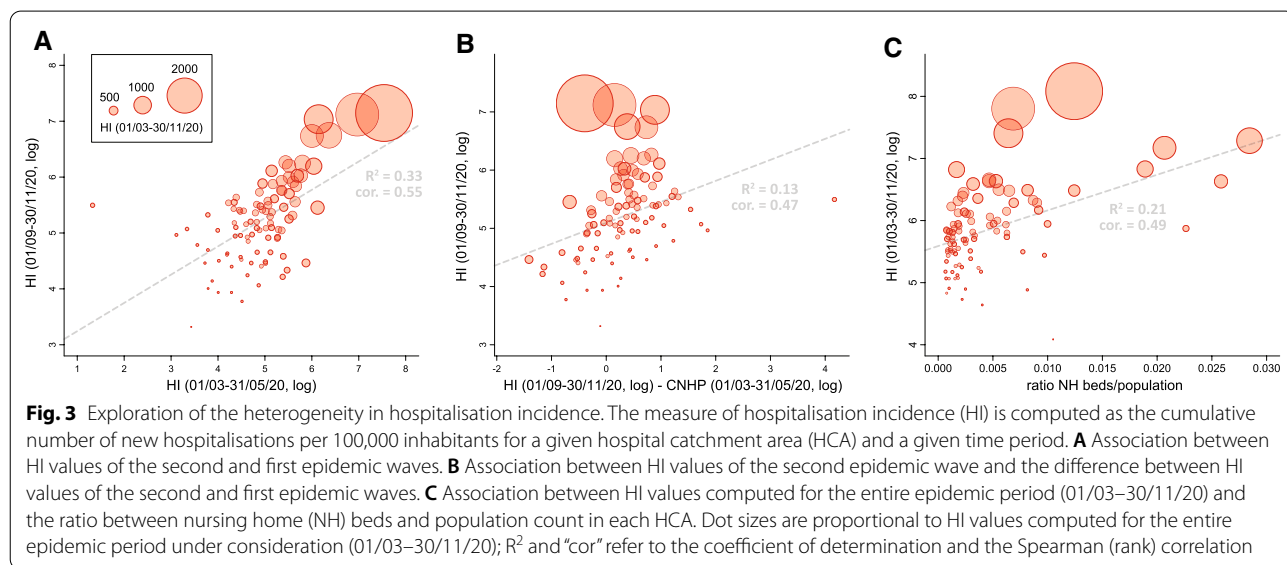
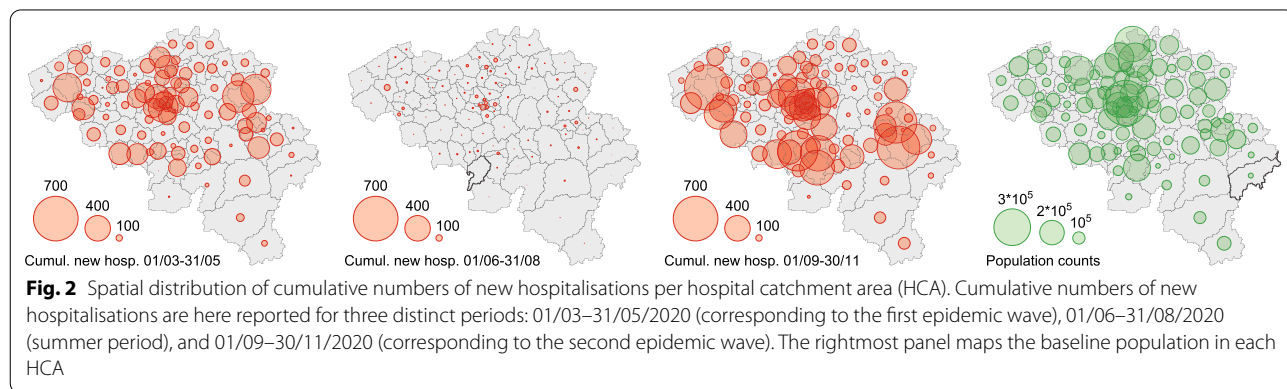
**Comparing the hospitalisation incidence of the two epidemic waves**

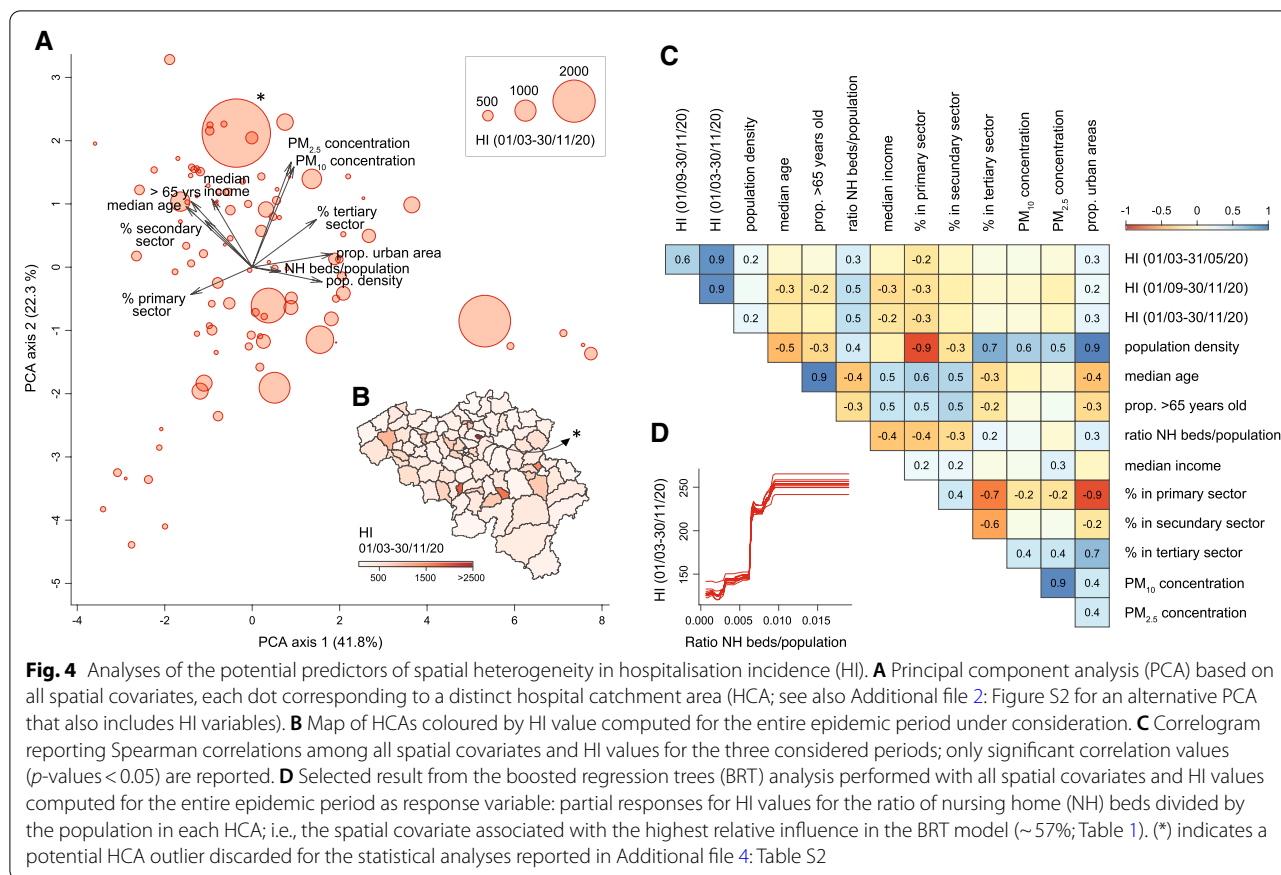
A first visual exploration of the absolute number of cumulated new hospitalisations per HCA and per time period confirmed that hospital admissions were globally higher during the second than the first epidemic wave and that, as expected, hospital admissions were lower in less populated areas of the country (Fig. 2). As a preliminary analysis, we performed Moran’s I tests that did not lead to the detection of significant spatial autocorrelation

for HI measures (for all time periods), meaning that there was no clear tendency for neighbouring HCAs to share similar hospitalisation incidence. Comparisons of HI values indicated that hospitalisation incidence was relatively well correlated between the two epidemic waves (Fig. 3A), but also that the hospitalisation incidence of the second wave tended to be correlated with the HI differences between the two epidemic waves (Fig. 3B). The latter results thus relate to a tendency to observe a higher hospitalisation incidence during the second wave in HCAs that were relatively less affected by the first epidemic wave.

**Investigating the drivers of the spatial heterogeneity in COVID-19 hospitalisation incidence**

Visual explorations based on PCAs (Fig. 4A, Additional file 2: Figure S2) did not yield a clear correlation pattern between HI and spatial predictors, even if we could observe that most HCAs with relatively higher HI





values tended to be associated with higher local population density, proportion of urban areas, and ratio between NH beds and HCA population. In Fig. 4B, we could note that one of the HCA was associated with a particularly high HI value, which was likely due to the relatively small area defined around the hospital included in this area. As the resulting HCA could represent a potential outlier, we repeated the multivariate analyses after having excluded that specific HCA (see below). Overall, these PCA trends were further confirmed by the univariate analyses summarised in our correlogram (Fig. 4C): for all three periods considered (first and second epidemic waves, as well as the entire epidemic period until the end of November, 2020), HI values were positively correlated with the proportion of urban areas and the ratio NH beds/population, but also negatively correlated with the percentage of people working in the primary sector. In our correlogram, population density was, however, not significantly associated with HI during the second epidemic wave (Fig. 4C). In our ULR analyses, only two spatial covariates were associated with a significant coefficient of determination ( $R^2$ ) value for all time periods under

consideration: the proportion of urban areas and the ratio NH beds/population (Table 1).

To investigate the effects that the cross-covariate correlations may have on our results, we further performed multivariate analyses, which confirm the association between HI and the ratio NH beds/population. We first performed MLR analyses, which were only based on spatial covariates for which we got a significant  $R^2$  value from the corresponding ULR analysis. For the ratio NH beds/population, we estimated significant MLR coefficient ( $\beta$ ) values ranging from 0.22 to 0.41 for the first and second epidemic wave, respectively (Table 1). On the contrary, we did not estimate significant MLR coefficients for the proportion of urban areas. It is however important to note that our MLR analyses only explained a relatively limited proportion of HI variability ( $R^2 < 0.25$ ; Table 1). Secondly, we performed BRT analyses to further explore non-linear relationships between HI and the different spatial covariates tested in this study. Our BRT models allowed reaching a relatively high correlation between observed and predicted HI values, with Spearman correlation values ranging between 0.69 and 0.76 for the first and second epidemic wave, respectively (Table 1). These BRT analyses mainly confirmed

**Table 1** Analyses of the potential predictors of spatial heterogeneity in hospitalisation incidence

Spatial covariate	HI (01/03–31/05/2020)			HI (01/09–30/11/2020)			HI (01/03–30/11/2020)		
	MLR R <sup>2</sup> = 0.18*, BRT cor. = 0.69			MLR R <sup>2</sup> = 0.24*, BRT cor. = 0.76			MLR R <sup>2</sup> = 0.23*, BRT cor. = 0.75		
	ULR R <sup>2</sup>	MLR β	BRT RI	ULR R <sup>2</sup>	MLR β	BRT RI	ULR R <sup>2</sup>	MLR β	BRT RI
Population density	0.00	–	0.8%	0.00	–	1.1%	0.00	–	0.9%
Median age	0.00	–	2.6%	0.02	–	6.2%	0.00	–	2.4%
Prop. > 65 years old	0.00	–	6.7%	0.01	–	1.4%	0.00	–	1.3%
Ratio NH beds/population	0.15*	0.22*	48.1%	0.22*	0.41*	55.3%	0.21*	0.31*	56.8%
Median income	0.00	–	3.2%	0.05*	-0.07	5.5%	0.00	–	3.9%
% in primary sector	0.04	–	5.7%	0.04*	0.00	14.3%	0.05*	0.02	14.4%
% in secondary sector	0.02	–	2.2%	0.02	–	1.8%	0.00	–	2.4%
% in tertiary sector	0.05*	0.06	3.0%	0.01	–	2.5%	0.00	–	1.5%
PM <sub>10</sub> concentration	0.01	–	1.6%	0.00	–	3.1%	0.00	–	1.7%
PM <sub>2.5</sub> concentration	0.01	–	0.9%	0.00	–	1.8%	0.00	–	0.9%
Prop. urban areas	0.05*	0.05	25.4%	0.05*	0.09	7.1%	0.06*	0.12	13.7%

This table summarises the results of univariate linear regression (ULR), multivariate linear regression (MLR), and boosted regression trees (BRT) analyses performed to investigate the association between measures of hospitalisation incidence (HI) and various spatial covariates associated with hospital catchment areas (HCAs). We report the following metrics: the coefficient of determination (R<sup>2</sup>) for the ULR analyses, the regression coefficient (β) for the MLR analyses, and the relative influence (RI) associated with each spatial covariate for the BRT analyses. We also report the overall R<sup>2</sup> and Spearman correlation (“cor.”) for each distinct MLR and BRT analysis, respectively. (\*) indicates if a given R<sup>2</sup> or β is significant (p-value < 0.05). See also Additional file 4: Table S1 for the results of the same analyses performed when only considering HI values based on nursing home (NH) residents, and Additional file 4: Table S2 for the results of the same analyses performed after having discarded HCAs of Brussels-Capital region as well as a potential outlier HCA (highlighted in Fig. 4)

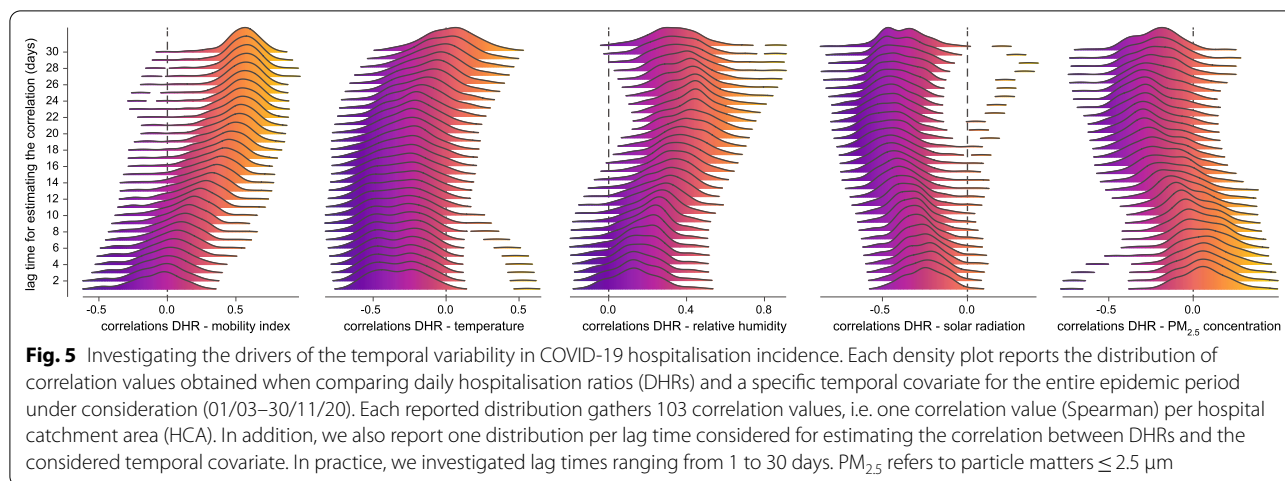
the notable contribution of the ratio NH beds/population in predicting HI values. Specifically, we got BRT relative influence (RI) values of 48 and 55% for the first and second epidemic wave, respectively, as well as of 57% for the entire epidemic period (Table 1). For the proportion of urban areas, RI values were much lower, except when considering the first epidemic wave for which we get a RI value of 25%. BRT response curves illustrate the positive yet not linear estimated relationship between HI and the ratio NH beds/population (Fig. 4D; and see also Additional file 3: Figure S3 for all response curves). We repeated all multivariate analyses only considering hospitalisations of NH residents (Additional file 4: Table S1) and excluding HCAs of Brussels-Capital Region and the potential outlier area highlighted in Fig. 4B (Additional file 4: Table S2). These two alternative sets of analyses confirmed the main results presented in Table 1.

In addition to exploring the association between HI and spatial covariates, we also used our BRT models to analyse the ability of those models to predict observed hospitalisation incidence. Specifically, we estimated the correlation between observed and predicted HI values, the latter being predicted under our different BRT models: the BRT model trained with HI values of the first epidemic wave, the BRT model trained with HI values of the second epidemic wave, the BRT model trained with HI values for the entire epidemic period under consideration, and a fourth BRT model trained with HI values of the second epidemic wave

(01/09–30/11/2020) but also including HI values of the first epidemic wave as a potential predictor in addition to all the other considered spatial predictors (Additional file 4: Table S3). Our results indicate a good ability of this fourth BRT model to predict HI values for the second epidemic wave, which can in practice be useful for predicting upsurge in hospital incidence in case of further epidemic waves.

**Investigating the drivers of the temporal variability in COVID-19 hospitalisation incidence**

A first visual comparison indicates an association between daily hospitalisation incidence and two temporal covariates: temperature and solar radiation (Additional file 1: Fig. S1). For each HCA, we then estimated the correlation between DHR and a temporal covariate while considering different lag times between the two measures (i.e. lag times ranging from 1 to 30 days; Fig. 5). The resulting density plots of correlation values highlight different trends in terms of temporal associations (Fig. 5): (i) DHR only becomes clearly positively correlated with the mobility index when considering a lag time of at least ~15 days, (ii) DHR is negatively correlated with daily temperature but not when considering important lag times (>20 days), (iii) DHR is positively and negatively correlated with relative humidity and solar radiation, respectively, for almost all lag times considered. However, DHR tends to be negatively correlated



with concentration of particles matter ( $\leq 2.5 \mu\text{m}$ ) for lag times higher than 15–20 days.

**Discussion**

Our analyses of the spatial heterogeneity in COVID-19 hospitalisation incidence mainly converge towards the relative importance of a particular spatial covariate: the ratio between the number of NH beds and the corresponding HCA population. In other words, HCAs with relatively higher proportions of NH beds tended to be associated with a higher hospitalisation incidence during the two epidemic waves of 2020. If we assume that the number of NH beds is a good proxy for the number of NH residents and that most hospitalised NH residents were admitted in the hospital of their HCA, this overall result further illustrates the strong impact of COVID-19 within NH communities on the Belgian hospitalisation burden. While it has been already established that nursing homes have to a large degree contributed to the Belgian COVID-19 death toll (>57% of COVID-19 deaths in Belgium are nursing home residents), our results also illustrate their important contribution to the national hospitalisation burden, despite the fact that symptomatic NH residents have not all been hospitalised. According to the weekly nursing home reports released by Sciensano (<https://covid-19.sciensano.be>), a majority of COVID-19 deaths of NH residents (76% since the beginning of the crisis) indeed occurred in nursing homes rather than in hospitals.

Our study highlights that hospitalisation incidence was reasonably correlated between the two epidemic waves of 2020, but also that a BRT model based on spatial covariates and previous hospitalisation incidence can lead to good predictions of hospitalisation incidence (correlation between predicted and observed hospitalisation incidence > 0.85). The latter result should be interpreted with

caution because we also show a tendency for an increase in incidence in HCAs that were relatively less impacted by the first epidemic wave. It is however difficult to interpret this tendency as it is based on the comparison between (i) hospitalisation incidence of the second wave and (ii) the difference in incidence between the second and first waves. These two measures are indeed by definition correlated because both are based on the incidence during the second wave.

Brussels-Capital Region is a particular case because it has several hospitals whose apparently small HCAs could not reflect the areas of origin of hospitalised people. Because of their relatively important capacities, especially its two major university hospitals, hospitals in Brussels also admitted patients from surrounding areas. For all these reasons, we also performed a new set of multivariate analyses without HCAs of Brussels in order to investigate to what extent they might have affected the outcome of our analyses (Additional file 4: Table S2). A comparison of the MLR and BRT results obtained with and without the HCAs of Brussels indicates no effect on our main conclusions.

The literature on hospitalisation incidence related to COVID-19 and spatial drivers is rather sparse. The few ecological studies describing the link between human and environmental drivers and COVID19, focus mostly on cases, mortality or case-fatality rates as outcomes. Also, to our knowledge, our study is the first indicating a significant association between the proportion of NH beds and hospitalisation incidence, even if some studies have shown a significant association between the proportion of older individuals and other COVID-19 indicators [17–19]. Other studies reported associations between various demographic and socio-economic factors [18–22] that are not confirmed in our analyses of hospitalisation incidence at the scale of Belgium. An association



between urban areas and COVID-19 was also observed in Brazil, where the distance from the state capital (São Paulo) was negatively associated with COVID-19 prevalence [22].

In the second part of study, we explored the association between the evolution of hospitalisation incidence through time and several temporal covariates of interest. Our relatively descriptive analyses support the previously discussed links between viral circulation (here reflected by resulting hospitalisations) and weather conditions [23–25]: new hospitalisations were far lower in between the two epidemic waves of 2020, corresponding to the summer period associated with higher temperature and solar radiation. Although it is not straightforward to infer a causality due to several confounding factors that may come into play (e.g. behavioural), several studies have previously discussed such potential direct or indirect links between viral circulation and climatic factors such as temperature or solar radiation [23, 24, 26–28]. When considering a delay of at least ~15 days, our analyses also indicate that the evolution of new hospitalisations is associated with human mobility, which is not necessarily easy to interpret when looking at the overall epidemic period. Finally, our analyses find a negative and delayed association with concentrations of particulate matter, which is not in line with (nor necessarily contradicts) previous studies discussing the potential role of fine particulate matter in spreading the virus [29] or having a negative impact on COVID-19 symptoms [30].

The present study dealt with hospitalisation data only available at hospital level, and therefore used HCAs as units of analysis. This approach has some limitations. First, we assumed that patients were hospitalized in the closest hospital with a COVID-19 unit. However, other factors may influence hospital choices such as its size or reputation, especially in urban areas where the density of hospitals is higher. Second, the definition of HCAs is dependent on the friction surface used to calculate travel times. We used a global friction surface at a spatial resolution of approximately one by one km, which is based on default input data and modelling methods [14]. Such estimations may however vary with local contexts such as varying road speed limits or congestion. Another limitation relates to the way hospitalisations are recorded in Belgium. Some hospitals have several site locations, but COVID-19 patients were aggregated and only recorded at the main site. This may cause spatial discrepancies when a hospital receives COVID-19 patients on secondary sites. In addition, when hospitals were approaching saturation, patient transfers may have been organised resulting in hospitalisations recorded in more distant hospitals.

## Conclusion

Our study proposes a comprehensive analytical framework that can be applied to investigate dynamics and potential drivers of COVID-19 hospitalisation incidence at a regional or national scale. While our approach based on hospital catchment areas presents some limitations (discussed above), it allows performing spatially-explicit analyses of data aggregated at the hospital level. A better epidemiological understanding of the COVID-19 pandemic can serve as a guide in settling the most appropriate public health responses. Because most public health interventions aim at avoiding a saturation of hospitals, understanding the spatio-temporal dynamic as well as the potential drivers of COVID-19 hospitalisation incidence remains of crucial importance. In addition, we also illustrate that it is possible to get relevant predictions of hospitalisation incidence, which could be useful for a better planning of transfers among hospitals and thus avoiding potential saturation and local overflow.

## Abbreviations

BRT: Boosted regression trees; DHR: Daily hospitalisation ratios; HCA: Hospital catchment area; HI: Hospitalisation incidence; MLR: Multivariate linear regression; NH: Nursing home; PCA: Principal component analysis; PM: Particulate matter;  $R^2$ : Coefficient of determination; RI: Relative influence; ULR: Univariate linear regression.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12942-021-00281-1>.

**Additional file 1: Figure S1.** Visual comparison between the evolution of daily new hospitalisations and the evolution of the temporal covariates considered in the present study. All temporal covariates were averaged over a distinct hospital catchment area (HCA), and each curve thus corresponds to a distinct HCA. All variables were also preliminarily treated by a moving average of 7 days. The mobility index is based on mobile phone data (see the text for further detail); the temperature is reported in the Kelvin scale (K); the relative humidity is expressed as a percentage (ratio between vapor partial pressure and saturation vapor partial pressure; 100% meaning an air mass fully charged in humidity); and the solar radiation is reported in Joules per square metre ( $J/m^2$ ).

**Additional file 2: Figure S2.** Principal component analysis (PCA) based on all spatial covariates as well as measures of hospitalisation incidence (HI). Specifically, we here included in the ACP HI values computed for the period corresponding to the first (01/03–31/05/2020) and the second (01/09–30/11/2020) epidemic waves. Each dot corresponds to a distinct hospital catchment area (HCA) and is displayed with an area proportional to the HI value computed for the entire epidemic period under consideration (01/03–30/11/2020).

**Additional file 3: Figure S3.** Response curves estimated for the boosted regression trees (BRT) model trained on measures of hospitalisation incidence (HI) computed for the entire epidemic period under consideration (01/03–30/11/2020).

**Additional file 4: Table S1.** Analyses of the potential predictors of spatial heterogeneity in hospitalisation incidence of nursing home (NH) residents. This table is equivalent to Table 1 and summarises the results of univariate linear regression (ULR), multivariate linear regression (MLR), and boosted regression trees (BRT) analyses performed to investigate the association between measures of hospitalisation incidence (HI) of NH residents and

various spatial covariates associated with hospital catchment areas (HCAs). We report the following metrics: the coefficient of determination ( $R^2$ ) for the ULR analyses, the regression coefficient ( $\beta$ ) for the MLR analyses, and the relative influence (RI) associated with each spatial covariate for the BRT analyses. In addition, we also report the overall  $R^2$  and Spearman correlation ("cor.") for each distinct MLR and BRT analysis, respectively. (\*) indicates if a given  $R^2$  or  $\beta$  is significant ( $p$ -value  $< 0.05$ ). **Table S2.** Analyses of the potential predictors of spatial heterogeneity in hospitalisation incidence, when excluding Brussels-Capital Region and a potential outlier area. This table is equivalent to Table 1 and summarises the results of univariate linear regression (ULR), multivariate linear regression (MLR), and boosted regression trees (BRT) analyses performed to investigate the association between measures of hospitalisation incidence (HI) and various spatial covariates associated with hospital catchment areas (HCAs). For these alternative analyses, we discarded the six HCAs of the Brussels-Capital Region, as well as a potential outlier HCA (marked with an asterisk in Fig. 4). We report the following metrics: the coefficient of determination ( $R^2$ ) for the ULR analyses, the regression coefficient ( $\beta$ ) for the MLR analyses, and the relative influence (RI) associated with each spatial covariate for the BRT analyses. In addition, we also report the overall  $R^2$  and Spearman correlation ("cor.") for each distinct MLR and BRT analysis, respectively. (\*) indicates if a given  $R^2$  or  $\beta$  is significant ( $p$ -value  $< 0.05$ ). **Table S3.** Comparison of the different boosted regression trees (BRT) models trained in the present study. Specifically, we here report the Spearman correlation between hospitalisation incidence predicted under various BRT models (one row = one specific BRT model) and observed hospitalisation incidence. As detailed in the text, the measure of hospitalisation incidence (HI) is computed as the cumulative number of new hospitalisations per 100,000 inhabitants for a given hospital catchment area (HCA) and a given time period. For each comparison between predicted and observed set of HI values, we report both the mean Spearman correlation value, as well as the minimum and maximum Spearman correlation values, all obtained while considering the ten BRT model replicates. (\*) refers to the BRT model trained with HI values of the second epidemic wave (01/09-30/11/2020) but also including HI values of the first epidemic wave (01/03-31/05/2020) as a potential predictor in addition to all the other considered spatial predictors.

#### Acknowledgements

We are grateful to Youri Baeyens and Patrick Lusyne from the SPF Economie for sharing population data, as well as to Philippe Maetz, Charlotte Vanpoucke, and Olav Peeters from the Belgian Interregional Environment Agency (IRCEL—CELINE) for sharing PM concentration data.

#### Authors' contributions

SD, CL, PS, SOV, NS, and MG designed the study; NVG and HVO collected and curated the data. SD, NVG, DDR, JA, JB, PS, FM, HVO, SOV, and NS analysed the data. All authors interpreted the results. SD wrote the first version of the manuscript and all authors edited the final manuscript. All authors read and approved the final manuscript.

#### Funding

SD, DDR, FM, and MG are supported by the *Fonds National de la Recherche Scientifique* (FNRS, Belgium). The research leading to these results has received funding from the European Union's Horizon 2020 project MOOD (grant agreement no. 874850).

#### Availability of data and materials

R script and related files needed to run the analyses are all available at [https://github.com/sdellicour/covid19\\_spell](https://github.com/sdellicour/covid19_spell). The GitHub repository also contains a tutorial for the delimitation of hospital catchment areas.

#### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interest.

#### Author details

<sup>1</sup>Spatial Epidemiology Lab (SpELL), Université Libre de Bruxelles, 50 av. FD Roosevelt, 1050, CP160/12 Bruxelles, Belgium. <sup>2</sup>Department of Microbiology, Immunology and Transplantation, Laboratory for Clinical and Epidemiological Virology, Rega Institute, KU Leuven - University of Leuven, Leuven, Belgium. <sup>3</sup>Institute of Life-Earth-Environment (ILEE), Université de Namur, Rue de Bruxelles 61, 5000 Namur, Belgium. <sup>4</sup>NAMur Research Institute for Life Sciences (NARILIS), Université de Namur, Rue de Bruxelles 61, 5000 Namur, Belgium. <sup>5</sup>Department of Epidemiology and Public Health, Sciensano, Brussels, Belgium. <sup>6</sup>Earth & Life Institute, Georges Lemaître Centre for Earth and Climate Research, UCLouvain, Place Louis Pasteur 3, 1348 Louvain-la-Neuve, Belgium. <sup>7</sup>ICTEAM, UCLouvain, 1348 Louvain-la-Neuve, Belgium. <sup>8</sup>Public Health and Primary Care, Gent University, Gent, Belgium. <sup>9</sup>Institute of Health and Society (IRSS), Université Catholique de Louvain, Clos Chapelle-aux-champs 30, 1200 Brussels, Belgium.

Received: 4 April 2021 Accepted: 1 June 2021

Published online: 14 June 2021

#### References

- Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nat Publishing Group*. 2020;579:270–3.
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020;382:727–33.
- McKee M, Stuckler D. If the world fails to protect the economy, COVID-19 will damage health not just now but also in the future. *Nat Med Nature Publishing Group*. 2020;26:640–2.
- Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. *Lancet Psychiatry Elsevier*. 2020;7:547–60.
- Pereira RHM, Braga CKV, Servo LM, Serra B, Amaral P, Gouveia N, et al. Geographic access to COVID-19 healthcare in Brazil using a balanced float catchment area approach. *Soc Sci Med*. 2021;273:113773.
- Spiteri G, Fielding J, Diercke M, Campese C, Enouf V, Gaymard A, et al. First cases of coronavirus disease 2019 (COVID-19) in the WHO European Region, 24 January to 21 February 2020. *Eurosurveillance*. European Centre for Disease Prevention and Control; 2020. 25:2000178.
- Kinross P, Suetens C, Dias JG, Alexakis L, Wijermans A, Colzani E, et al. Rapidly increasing cumulative incidence of coronavirus disease (COVID-19) in the European Union/European Economic Area and the United Kingdom, 1 January to 15 March 2020. *Eurosurveillance*. European Centre for Disease Prevention and Control; 2020. 25:2000285.
- Dellicour S, Durkin K, Hong SL, Vanmechelen B, MartiCarreras J, Gill MS. A phylodynamic workflow to rapidly gain insights into the dispersal history and dynamics of SARS-CoV-2 lineages. *Mol Biol Evol*. 2020. <https://doi.org/10.1093/molbev/msaa284>.
- Hardy OJ, Dubourg D, Bourguignon M, Dellicour S, Eggerickx T, Gilbert M, et al. A world apart: levels and factors of excess mortality due to COVID-19 in care homes. The case of Wallonia - Belgium. medRxiv. Cold Spring Harbor Laboratory Press; 2020;2020.08.29.20183210.
- Wu J, McCann A, Katz J, Peltier E, Singh KD. The pandemic's hidden toll: half a million deaths. *The New York Times* [Internet]. 2020 Apr 21 [cited 2021 Jan 18]; <https://www.nytimes.com/interactive/2020/04/21/world/coronavirus-missing-deaths.html>
- Bustos Sierra N, Bossuyt N, Braeye T, Leroy M, Moyersoen I, Peeters I, et al. All-cause mortality supports the COVID-19 mortality in Belgium and comparison with major fatal events of the last century. *Arch Public Health*. 2020;78:117.
- Taccone FS, Goethem NV, Pauw RD, Wittebole X, Blot K, Oyen HV, et al. The role of organizational characteristics on the outcome of COVID-19

- patients admitted to the ICU in Belgium. The Lancet Regional Health - Europe [Internet]. Elsevier; 2021 [cited 2021 Jan 21]; [https://www.thelancet.com/journals/lanep/article/PIIS2666-7762\(20\)30019-3/abstract](https://www.thelancet.com/journals/lanep/article/PIIS2666-7762(20)30019-3/abstract)
13. Van Goethem N, Vilain A, Wyndham-Thomas C, Deblonde J, Bossuyt N, Lernout T, et al. Rapid establishment of a national surveillance of COVID-19 hospitalizations in Belgium. *Arch Public Health*. 2020;78:121.
  14. Weiss DJ, Nelson A, Gibson HS, Temperley W, Peedell S, Lieber A, et al. A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*. 2018;553:333.
  15. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. *J Anim Ecol*. 2008;77:802–13.
  16. Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, et al. The ERA5 global reanalysis. *Q J R Meteorol Soc*. 2020;146:1999–2049.
  17. Hoffmann C, Wolf E. Older age groups and country-specific case fatality rates of COVID-19 in Europe USA and Canada. *Infection*. 2020. <https://doi.org/10.1007/s15010-020-01538-w>.
  18. Martins LD, da Silva I, Batista WV, Andrade M de F, Freitas ED de, Martins JA. How socio-economic and atmospheric variables impact COVID-19 and influenza outbreaks in tropical and subtropical regions of Brazil. *Environ Res*. 2020.191:110184.
  19. Sarmadi M, Moghanddam VK, Dickerson AS. L Martelletti Association of COVID-19 distribution with air quality, sociodemographic factors, and comorbidities: an ecological study of US states. *Air Qual Atmos Health*. 2020. <https://doi.org/10.1007/s11869-020-00949-w>.
  20. Buja A, Paganini M, Cocchio S, Scioni M, Rebba V, Baldo V. Demographic and socio-economic factors, and healthcare resource indicators associated with the rapid spread of COVID-19 in Northern Italy: An ecological study. *PLoS ONE*. 2020;15:e0244535.
  21. Rafael R de MR, Neto M, Depret DG, Gil AC, Fonseca MHS, Souza-Santos R, et al. Effect of income on the cumulative incidence of COVID-19: an ecological study. *Rev Latino-Am Enfermagem* [Internet]. Escola de Enfermagem de Ribeirão Preto / Universidade de São Paulo; 2020 [cited 2021 Jan 29];28. [http://www.scielo.br/scielo.php?script=sci\\_abstract&pid=S0104-11692020000100364&lng=en&nrm=iso&tlng=en](http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0104-11692020000100364&lng=en&nrm=iso&tlng=en)
  22. Fortaleza CMCB, Guimarães RB, Almeida GB de, Pronunciante M, Ferreira CP. Taking the inner route: spatial and demographic factors affecting vulnerability to COVID-19 among 604 cities from inner São Paulo State, Brazil. *Epidemiol Infect* [Internet]. Cambridge University Press; 2020 [cited 2021 Jan 29];148. <https://www.cambridge.org/core/journals/epidemiology-and-infection/article/taking-the-inner-route-spatial-and-demographic-factors-affecting-vulnerability-to-covid19-among-604-cities-from-inner-sao-paulo-state-brazil/B7238319C73790554568578182A14C40>
  23. Wu Y, Jing W, Liu J, Ma Q, Yuan J, Wang Y, et al. Effects of temperature and humidity on the daily new cases and new deaths of COVID-19 in 166 countries. *Sci Total Environ*. 2020;729:139051.
  24. Mecnas P, Bastos RT da RM, Vallinoto ACR, Normando D. Effects of temperature and humidity on the spread of COVID-19: A systematic review. *PLoS ONE*. 2020. 15:0238339.
  25. Carlson CJ, Gomez ACR, Bansal S, Ryan SJ. Misconceptions about weather and seasonality must not misguide COVID-19 response. *Nat Commun Nature Publishing Group*. 2020;11:4312.
  26. Kassem AZE. Does temperature affect COVID-19 transmission? *Front Public Health* . 2020: <https://doi.org/10.3389/fpubh.2020.554964/full>
  27. Ahmadi M, Sharifi A, Dorosti S, Jafarzadeh Ghouschi S, Ghanbari N. Investigation of effective climatology parameters on COVID-19 outbreak in Iran. *Sci Total Environ*. 2020;729:138705.
  28. Ratnesar-Shumate S, Williams G, Green B, Krause M, Holland B, Wood S, et al. Simulated sunlight rapidly inactivates SARS-CoV-2 on surfaces. *J Infect Dis*. 2020;222:214–22.
  29. Domingo JL, Marquès M, Rovira J. Influence of airborne transmission of SARS-CoV-2 on COVID-19 pandemic. A review. *Environ Res*. 2020;188:109861.
  30. Pozzer A, Dominicci F, Haines A, Witt C, Münzel T, Lelieveld J. Regional and global contributions of air pollution to risk of death from COVID-19. *Cardiovasc Res*. 2020;116:2247–53.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

