

RESEARCH

Open Access



Using a general practice research database to assess the spatio-temporal COVID-19 risk

Oana Petrof^{1*}, Thomas Neyens^{1,2}, Bert Vaes³, Arne Janssens³ and Christel Faes¹

Abstract

Background In Flanders, general practitioners (GPs) were among the first ones to collect data regarding COVID-19 cases. Intego is a GPs' morbidity registry in primary care with data collected from the electronic medical records from a sample of general practices. The Intego database contain elaborate information regarding patient characteristics, such as comorbidities. At the national level, the Belgian Public Health Institute (Sciensano) recorded all test-confirmed COVID-19 cases, but without other patient characteristics.

Methods Spatio and spatio-temporal analyses were used to analyse the spread of COVID-19 incidence at two levels of spatial aggregation: the municipality and the health sector levels. Our study goal was to compare spatio-temporal modelling results based on the Intego and Sciensano data, in order to see whether the Intego database is capable of detecting epidemiological trends similar to those in the Sciensano data. Comparable results would allow researchers to use these Intego data, and their wealth of patient information, to model COVID-19-related processes.

Results The two data sources provided comparable results. Being a male decreased the odds of having COVID-19 disease. The odds for the age categories (17,35], (35,65] and (65,110] of being a confirmed COVID-19 case were significantly higher than the odds for the age category [0,17]. In the Intego data, having one of the following comorbidities, i.e., chronic kidney disease, heart and vascular disease, and diabetes, was significantly associated with being a COVID-19 case, increasing the odds of being diagnosed with COVID-19.

Conclusion We were able to show how an alternative data source, the Intego data, can be used in a pandemic situation. We consider our findings useful for public health officials who plan intervention strategies aimed at monitor and control disease outbreaks such as that of COVID-19.

Keywords Spatio-temporal methods, COVID-19, Comorbidities

Introduction

COVID-19 is a respiratory disease caused by a highly infectious corona virus, SARS-CoV-2 [1, 2], which has quickly spread across continents. Since the initial

outbreak of the global pandemic, governments and governmental agencies around the world were responsible for the epidemiological follow-up of the COVID-19 epidemic. March 2020 is the date commonly referred to as the start of the epidemic in Belgium [3]. Given the rapid increase in COVID-19 cases, the Belgian government decided to implement a lockdown from March until April 2020, i. e., during the first wave of the epidemic (1 March 2020 until 21 June 2020) [4, 5]. Later, from October until April 2021, a second lockdown was imposed during the second wave (31 August 2020 until 14 February 2021) [4, 5].

*Correspondence:

Oana Petrof
oana.petrof@uhasselt.be

¹ I-Biostat, Hasselt University, Diepenbeek, Belgium

² Leuven Biostatistics and Statistical Bioinformatics Centre (L-BioStat), KU Leuven, Leuven, Belgium

³ Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium



© The Author(s) 2024. **Open Access** 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.

It is crucial to have a reliable source of data to follow up the epidemic, where complex analyses such as spatio-temporal models can detect local outbreaks or local hotspots of the epidemic. In Belgium, the leading scientific institution in the epidemiology of infectious diseases, Sciensano is responsible for monitoring the epidemic evolution and assessing its consequences on the Belgian population health [6]. They did set up a new database during the pandemic, to collect information about the daily COVID-19 data new cases, hospitalizations, ICU admitted patients and deaths [7], together with their age, gender, and residential postal codes. Although this has been a very important database, it lacks further individual level information. Multiple comorbidities are associated with the COVID-19 disease progression [8–10], while the elderly population is more susceptible to the COVID-19 disease [8, 11]. Comorbidities often increase the probability of infection and represent a risk factor for COVID-19 patients [8, 10, 12]. Underlying diseases, such as hypertension, cardio-vascular diseases, diabetes, and asthma, have been reported as risk factors for COVID-19, increasing the mortality rate [9, 10, 12, 13]. Information on the comorbidities of the patients could not be collected in the federal database, but this information can easily be collected by general practitioners (GPs), since they follow up their patients in time. The Intego network represents a GP morbidity registry, implemented in Flanders, in which GPs continuously register information about their patients' sociodemographic variables, diagnoses, clinical parameters, laboratory results and medication prescriptions [14]. Other European countries like the United Kingdom [15, 16] or the Netherlands [17] implemented a similar GP registry, given its added value.

The objective of this paper was to validate the Intego database by comparing the spatio-temporal analysis results with those based on the the federal COVID-19 data. The advantage of using the Intego study is the availability of individual level information data, which includes details of the patients, like comorbidities. In our study, results from the national COVID-19 database are considered as the “gold standard” showing the real spatio-temporal distribution of all COVID-19 cases in Flanders. We investigated if a GP morbidity registry could provide similar and, thus, reliable results in a pandemic situation, further adding important information on the association between COVID-19 and comorbidity.

Methods

Data sources

In the current study two data sources on COVID-19 were analyzed: (i) confirmed positive COVID-19 cases based on the Intego database, a registration network for participating family practices in Flanders [14] and (ii) reported

test-confirmed cases in Belgium based on data provided by the Belgian Public Health Institute (Sciensano). The study period was chosen to cover the first two COVID-19 waves, from March 1st until November 30th 2020, before the start of the nationwide vaccination campaign in Belgium.

The Intego morbidity registration network collects data from GPs using the medical software program CareConnect® (Corilus, Namur, Belgium). On 31 December 2019, there were 16,722 active GP's and 2,209 GP's in training in Belgium, with 2,65 GPs per practice in Belgium [18]. During the study period, the Intego program included data from 104 public GP practices spread throughout Flanders, covering patients with varying socio-demographic and socio-economic backgrounds. The Intego GPs registered new COVID-19 diagnoses. We estimate the incidence of registered patients with a confirmed positive COVID-19 diagnosis as the ratio between the number of patients that tested positive and the patient population. We assume that the patient population is actually the yearly contact group [19], which consists of those patients who contacted their general practitioner at least once from December 1st 2019 until November 30th 2020. The Intego database includes for each patient the residential information (postal code), age, gender and presence/absence of the following comorbidities: asthma, chronic liver disease, chronic lung disease, chronic neurological disorder, chronic kidney disease, heart and vascular disease, diabetes, hypertension, immunodeficiency and obesity. All International Classification of Primary Care (ICPC) codes can be found in the Appendix, Table 4.

As a second data source, we used data from the Belgian population health institute (Sciensano, <https://epistat.wiv-isp.be/covid/>), responsible for the epidemiological follow-up of the COVID-19 epidemic in Belgium. They collect daily numbers on confirmed COVID-19 cases, hospitalizations and mortality. We used summary data on the number of cases per age group, gender and postal code residential information, who were diagnosed with COVID-19 between March 1st and November 30th 2020.

We defined the COVID-19 incidence as the proportion of COVID-19 positive cases divided by the population size. In this case, the Flanders' population in 2019 is used to describe the “population at risk” per area.

The analysis was done at two different levels of areal aggregation: at the municipality level and at the health sector level. The 300 municipalities in Flanders are included in our analysis.

Health sectors, denoted as primary care zones (PCZs), create the basis for effective integrated care and services in a locality, each serving a community of approximately 75 000 - 125 000 inhabitants [20]. The role of these

organizations include aligning the organization and provision of high-quality care and support; supporting local social policy; supporting profession-specific associations; supporting primary care professionals and the organization of multi- and interdisciplinary collaboration; and cooperating on the Flemish health objectives relating to prevention and propose their own objectives [20]. There are 59 health sectors in Flanders which were included in our study.

Looking at the observed incidences on Fig. 1 as well as in the Appendix on Figs. 11 and 12, we can see that overall trends are similar, but the maps based on the national database are much smoother as compared to the observed incidences based on the Intego database. This could be expected as the Intego database contains only information for a fraction of the Flemish population.

$$\begin{aligned} \text{logit}(\pi_{itga}) &= \log(\pi_{itga}/(1 - \pi_{itga})) = \\ &= \beta_0 + \beta_1 * \text{gender}_i + \beta_2 * \text{agegroup}_{2i} + \beta_3 * \text{agegroup}_{3i} + \beta_4 * \text{agegroup}_{4i} + \\ &+ u_i + v_i + \gamma_t + \delta_{it}, \end{aligned} \tag{2}$$

Models

Disease mapping methods play an important role in understanding the spatial pattern of diseases and discovering areas characterized by unusually high or low risk [21–24]. These methods are commonly used for areal data, usually based on administrative boundaries, which are discrete in nature, as counts of diseases or deaths in each area [23]. We developed two models: (1) a validation model which uses only the age, gender, and residential information similar to the information available from the federal database, and (2) a comorbidity model using additional patient level information from the Intego database.

Validation model

Let's assume the total number of confirmed COVID-19 cases Y_{itga} in area i during month t with gender g in age group a have a binomial distribution

$$Y_{itga} \sim \text{Binomial}(\pi_{itga}, n_{itga}), \tag{1}$$

where the parameter n_{itga} denotes the number of people at risk in area i with gender g in age group a , which is constant during the entire time period analysed. Spatial and spatio-temporal models with various combinations of covariates and their interactions as well as different random effects for the spatial and temporal effects are used as depicted in Table 2, in Appendix. The interactions between different spatial and temporal random effects were mainly developed by Knorr-Held [25], but a good overview and applications can be found in the book

written by Blangiardo & Cameletti [23]. They include spatially structured and unstructured random effects, a temporally structured effect (see Table 2 in Appendix). We used model selection to find the best fitting model, via the WAIC (Watanabe-Akaike Information Criterion, [26]) and the DIC (Deviance Information Criterion, [27]). However, the best fitting models, model 15 and 16, included the interaction of the covariates and the categorical time effects. The interpretation of these parameters did not make sense from a medical and statistical point of view. Scientific rationale and expert opinion were used to develop the second best fitting model for our data (model 18).

We introduced the spatio-temporal model found to be the best fitting model for our study. The probability π_{itga} is modelled via a logit link, as follows:

where $gender$ denotes a binary variable taking the value 1 when a patient is a male and 0 when a patient is a female; $agegroup_2$, $agegroup_3$, and $agegroup_4$ are dummy variables which indicate whether patients belong to the age groups 17-35, 35-65, and +65, respectively. The random effects term $\mathbf{u} = (u_1, u_2, \dots, u_n)$ accounts for the spatially correlated heterogeneity (CH) and the random effects term $\mathbf{v} = (v_1, v_2, \dots, v_n)$ for the uncorrelated heterogeneity (UH) at the areal level. The unstructured random effect is defined as $v_i \sim N(0, \sigma_v^2)$. The CH random effect is based on the full conditional distribution of the random effects u_i , an intrinsic conditional autoregressive (CAR) random effects term, as introduced by [28], as follows:

$$[u_i | u_j, j \neq i, \sigma_u^2] \sim N(\bar{u}_i, \sigma_i^2), \tag{3}$$

where

$$\bar{u}_i = \frac{1}{\sum_j w_{i,j}} \sum_j u_j w_{i,j}$$

$$\sigma_i^2 = \frac{\sigma_u^2}{\sum_j w_{i,j}}$$

and σ_u^2 is a variance parameter that controls the degree of smoothing, with the adjacency-based weights $w_{i,j}$. A binary indicator, is used based on sharing boundaries, with $w_{i,j} = 1$ if the areas i and j were adjacent and 0 otherwise. The parameter γ_t represents the temporally structured random effect, modeled dynamically using a random walk of order 1, defined as:

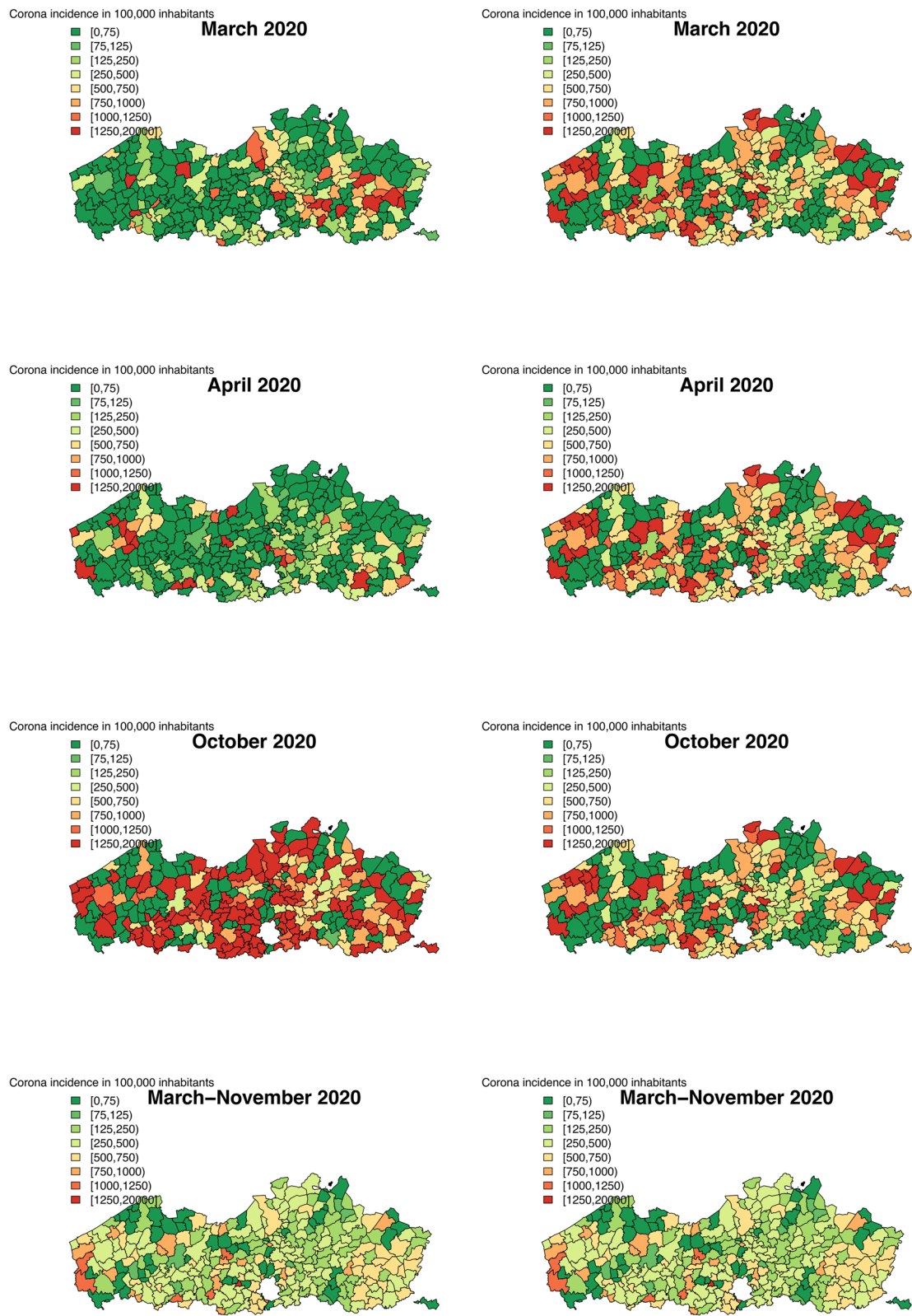


Fig. 1 Incidence of positive COVID-19 cases in a population of 100,000 inhabitants using the Intego data (left) and the Sciensano data (right)

$$\gamma_t | \gamma_{t-1} \sim \text{Normal}(2\gamma_{t-1}, \sigma^2). \tag{4}$$

The parameter δ_{it} denotes the spatio-temporal interaction between an uncorrelated heterogeneity (UH) random effect and an unstructured time random effect. It represents unobserved covariates for each pixel (i, t) that do not have any structure in space and time [25], capturing variation which cannot be explained by the main effects of a model.

Comorbidities' model

The analysis for the Intego data is done at both municipality and health sector level including comorbidities, using the model described by Eq. (2) adapted as follows:

$$\begin{aligned} \text{logit}(\pi_{itgac}) = & \beta_0 + \beta_1 * \text{male}_i + \beta_2 * \text{agegroup}_{2i} + \beta_3 * \text{agegroup}_{3i} + \beta_4 * \text{agegroup}_{4i} + \\ & + \beta_5 * \text{comorbidity}_i + u_i + v_i + \gamma_t + \delta_{it}, \end{aligned} \tag{5}$$

where *comorbidity* denotes a binary variable taking the value 1 when a comorbidity is present and 0 when a comorbidity is not present. The interpretation of the other parameters remains unchanged. Each individual comorbidity is included in the analysis in either:

- a univariate analysis, in which the significance of every comorbidity effect is tested separately, and
- a multivariate analysis, in which all significant comorbidities are included together in the final model.

Parameter estimation

Spatial and spatio-temporal models are traditionally analyzed using Bayesian methods using Markov chain Monte Carlo (MCMC) simulation methods [23, 29, 30]. While the Bayesian approach can easily account for uncertainty in the estimates/predictions, can flexibly cope with complex model specifications and can easily deal with missing data, its main limitation is the computational burden [23]. Especially for large data collections, characterized by high spatial and temporal resolutions, the computing time needed to perform Bayesian inference via MCMC could take from hours or even days to compute accurate posterior marginals. To overcome this issue, we use the integrated nested Laplace approximations (INLA) [31]. This alternative estimation method gained popularity as an approximation procedure via numerical integration for Bayesian computing.

The analysis is performed using the R-INLA package [31] in R [32].

Results

Validation of the Intego database

The municipality analysis' results are presented in this section, while the health sector analysis' results are shown in the Appendix section. Both analyses lead to similar general conclusions. Significant effects were found for gender and age categories, with similar trends found as based on the national collected database data (Table 1). Based on the Intego data, the odds of being a confirmed COVID-19 case for males are 0.94 (0.90; 0.98) times the odds for females, when controlling for all other variables. In terms of percent change, we can say that the odds of being a confirmed COVID-19 case for males are 5.65% (2%; 10%) lower than the odds for females. The odds of being a confirmed COVID-19 case for the age

categories (17,35], (35,65] and (65,110] are, respectively, 106% (91%; 121%), 83% (70%; 95%) and 10% (1.8%; 19.8%) higher than the odds for the age category [0,17], controlling for all other variables in the model.

Based on the Sciensano data, the odds of being a confirmed COVID-19 case for males are 17.54% (17.24%; 18.58%) lower than the odds for females. The odds for the age categories (17,35], (35,65] and (65,110] of being a confirmed COVID-19 case are 188% (187%; 196%), 135% (133%; 140%) and 141% (135%; 143%) higher than the odds for the age category [0,17], controlling for all other variables, which differs substantially from the results based on the Intego data.

Comorbidities' results

Three comorbidities had important, i.e., "significant", effects based on the univariate analysis, thus were included in the multivariate analysis. The odds for the patients diagnosed with chronic kidney disease, heart and vascular disease and diabetes of being a confirmed COVID-19 case are 30% (11%; 49%), 17% (9%; 26%) and 20% (10%; 31%) higher than the odds of a patient without other comorbidities, controlling for all other variables.

The two random effects, the spatially correlated heterogeneity (CH) and the uncorrelated heterogeneity (UH), are not uniquely identifiable and only their sum is well identified by the data [33]. We interpret for every model the sum of these spatial random effects, while avoiding interpreting them separately. We observe a larger effect of the random effects components in the Intego analysis, as compared to the Sciensano analysis. For the Intego

Table 1 Parameter estimates at the municipality level for the Intego and Sciensano data

	Intego			Sciensano		
	Mean	2.50%	97.50%	Mean	2.50%	97.50%
Validation model						
$\hat{\beta}_0$	0.0010	0.0009	0.0011	0.0007	0.0007	0.0007
$\hat{\beta}_1$	0.9453	0.9076	0.9833	0.8209	0.8142	0.8276
$\hat{\beta}_2$	2.0650	1.9175	2.2150	2.9192	2.8740	2.9645
$\hat{\beta}_3$	1.8293	1.7042	1.9567	2.3683	2.3331	2.4035
$\hat{\beta}_4$	1.1077	1.0183	1.1986	2.3939	2.3551	2.4328
\hat{v}_i	0.1078	0.0422	0.1912	0.0007	0.0002	0.0014
\hat{u}_i	0.1270	0.0253	0.2581	0.0587	0.0447	0.0732
$\hat{\gamma}_t$	0.6738	0.2166	1.3050	2.5216	0.8301	4.6852
$\hat{\delta}_{it}$	0.3552	0.2907	0.4243	0.2562	0.2374	0.2744
Comorbidity model						
$\hat{\beta}_0$	0.0010	0.0009	0.0011			
$\hat{\beta}_1$	0.9371	0.8996	0.9749			
$\hat{\beta}_2$	2.0584	1.9113	2.2080			
$\hat{\beta}_3$	1.7821	1.6592	1.9071			
$\hat{\beta}_4$	0.9855	0.8999	1.0727			
$\hat{\beta}_5$ Chronic kidney disease	1.3026	1.1156	1.4937			
$\hat{\beta}_6$ Heart and vascular disease	1.1760	1.0929	1.2600			
$\hat{\beta}_7$ Diabetes	1.2087	1.1065	1.3122			
\hat{v}_i	0.0986	0.0323	0.1801			
\hat{u}_i	0.1470	0.0348	0.2975			
$\hat{\gamma}_t$	0.7086	0.1761	1.4036			
$\hat{\delta}_{it}$	0.3574	0.2846	0.4274			

analysis, the variations of the random effects component were close for both analyses, with or without the comorbidities' effects. The addition of the comorbidities as covariates in our model did not explain part of the variation.

The temporal structured effects over time, together with the 95% credible intervals, are plotted for Intego and Sciensano data, based on the two aggregation levels, municipality and health sector level (Fig. 5, in [Appendix](#)). Two waves of the pandemic can easily be observed in the months of March - April, 2020 and September - November, 2020. A slight disagreement was observed for the first wave between the two data sources. While the results based on the Sciensano database do not capture the cluster of increased risk present in the eastern part of Flanders, the Intego database results reflect the starting point of the epidemic in Belgium. Starting from May until September, a constant time trend effect can be observed for the probability of being diagnosed with

COVID-19. This probability tremendously increased during the second Belgian wave, which lasted from September until November 2020, with both data sources being in agreement. It is clear that the results do not reflect the same incidence for this time period. More heterogeneity is present on the Intego compared to Sciensano results. Likely, this is the case, because the Intego database represents a subsample of the total COVID-19 cases, which the Sciensano database should contain.

We calculated the predicted number of COVID-19 confirmed cases in a population of 100,000 inhabitants for both data sources and both levels of aggregation. Figures 2, 3 and 4 show the predicted number of cases per municipality per month, while Figs. 6, 7, 8, 9 and 10 in [Appendix](#) show the predicted number of cases per health sector per month. To make the predictions comparable between the two data sources, the Flanders' population in 2019 is used to calculate the predicted incidence per area

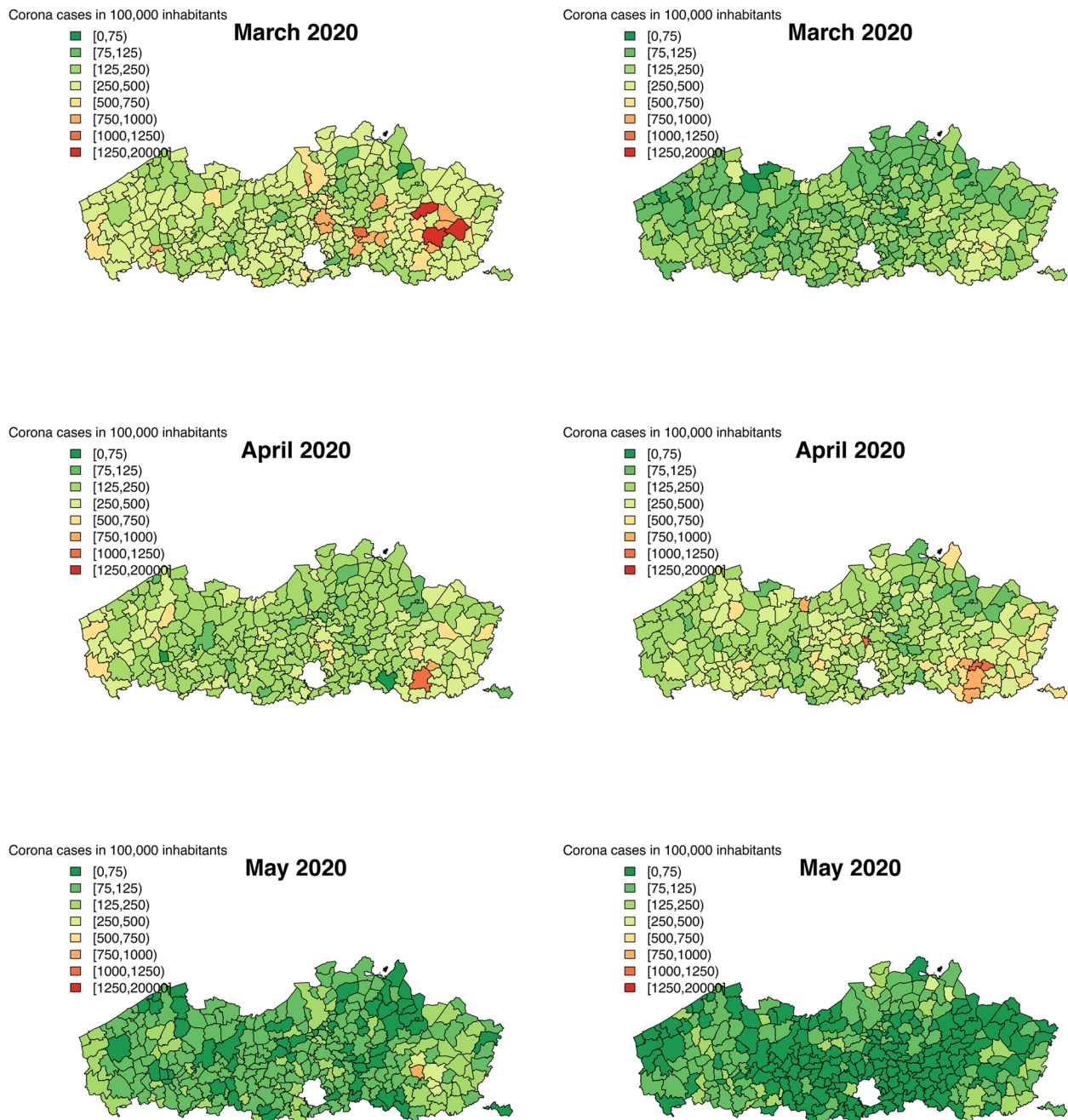


Fig. 2 Predicted number of corona cases in a population of 100,000 inhabitants using the Intego data (left) and the Sciensano data (right)

and month. Over the entire time period, it seems that the two data sources provide similar predicted number of cases per area and month, with slight differences due to variation. The two waves of the pandemic are present on these maps as well. For the first Belgian wave (March - April, 2020), a similar disagreement is present between

the results of the Intego and Sciensano data analyses, as was seen for the temporal structured random effects. We note that this map of model-based predictions is smoother than those of the observed incidences (Figs. 1, 11 and 12 in Appendix), a direct effect of the modelling procedure that captures the mean spatial trend in incidences.

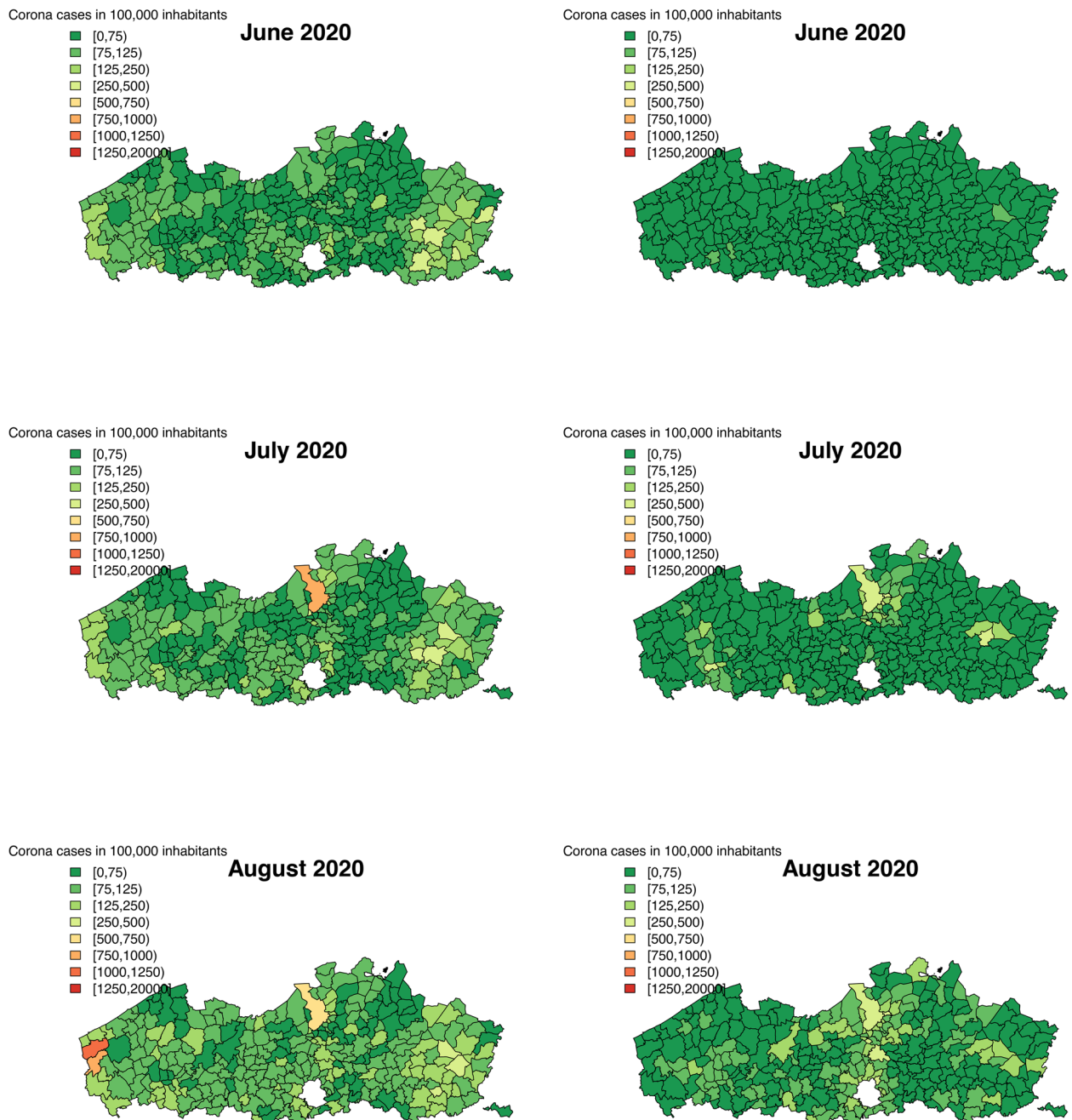


Fig. 3 Predicted number of corona cases in a population of 100,000 inhabitants using the Intego data (left) and the Sciensano data (right)

The maps of the predicted number of COVID-19 cases clearly show us the two pandemic waves, March-April 2020 and September-November 2020. During the first wave, a cluster of increased number of cases was located around Limburg province, in the Eastern part of Flanders. This part of the country was the first one to be

affected by the COVID-19 disease. Over the entire time period, the results of the two analyses for the two databases are in agreement, with a small deviation during the first wave. This difference can be attributed to the difference in the coding system between the two databases in the beginning of the epidemic. From May until August,

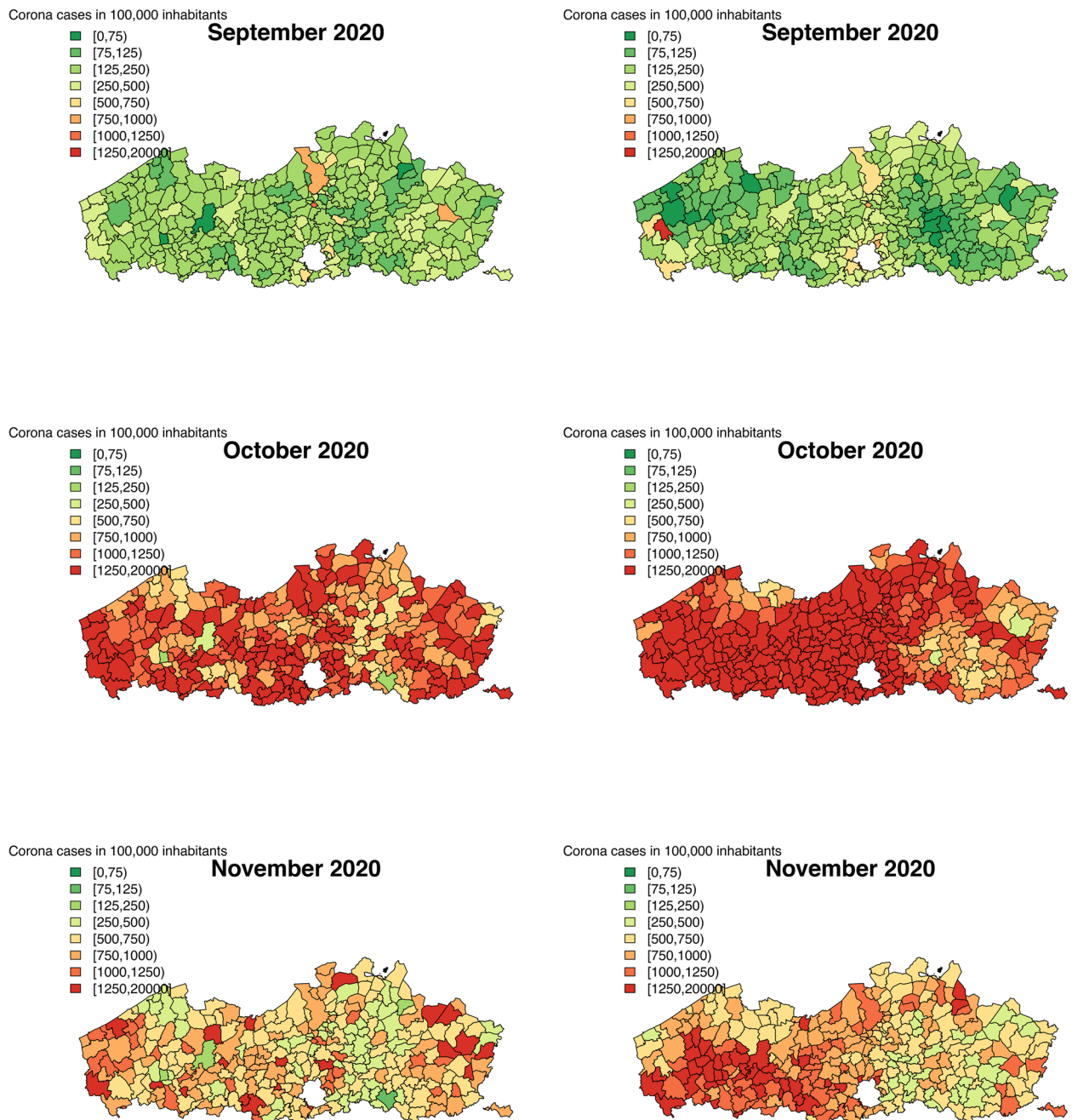


Fig. 4 Predicted number of corona cases in a population of 100,000 inhabitants using the Intego data (left) and the Sciensano data (right)

2020, the number of cases was decreasing over the entire study region, most likely as a consequence for the strict lockdown measures imposed by the Belgian and Flemish government during the first pandemic wave. In September, the number of cases started to increase again, reaching a peak in October. Note that, likely due to a smaller

sample size in the Intego database, we observe more variability of the maps compared to Sciensano results.

Age and gender have significant effects on the probability to have COVID-19. Males have a slightly lower probability to get COVID-19 compared to females. The effect for gender is larger based on the Sciensano database. A

larger effect based on the Sciensano database can be seen for the age groups as well. The sample size of Intego data is much smaller (9,467 patients) compared to Sciensano data (235,066 individuals).

Discussion

In this manuscript, a spatio-temporal model was used to investigate the distribution of COVID-19 cases in Flanders, Belgium based on two different databases. The Sciensano database was assumed to reflect the true spread of COVID-19 disease. Over the entire time period, the results of the two analyses for the two databases were in agreement, with a small deviation during the first wave. This difference can be attributed to the difference in the coding system between the two databases at the start of the Flemish epidemic. Using data from the Intego database, we examined if a GP morbidity registry could be used in a pandemic situation and we investigated if specific comorbidities had an influence on the COVID-19 disease progression. Our findings are in agreement with other studies, showing that patients with at least one of these comorbidities are more likely to be diagnosed with COVID-19 disease [8, 12, 34]. These results suggest that age, gender and comorbidities represent risk factors for critical patients with COVID-19 [8, 11].

The results of the health sector analyses presented in the [Appendix](#) section of our study are in agreement with the results at the municipality level. The health sector analyses' results are important for policy makers since the COVID-19 policies are made at this level. Moreover, especially model-based results are useful for policy-makers, and often more than maps of observed incidences, as model-based estimates provide insight in mean trends across the region of interest, here Flanders.

Our study has a number of limitations. The amount of GPs included in the Intego project is relatively small compared to the total number of GPs in Flanders. Progress is being made since every year new GPs are joining this database, thus more data about more patients are collected. Every area has a different number of practices and every practice has a different number of GPs included in the Intego project. Moreover, the number of tests performed per area varies, but their availability might also differ temporally, due to changes in the production capacity of these tests.

For instance, during the first pandemic wave, due to limited capacity, only a fraction of suspected Belgian COVID-19 patients has been tested to confirm SARS-CoV-2 infection [35]. Moreover, it should be noted that patients with a comorbidity may have been tested more often by their GP's, as compared to a person without comorbidities. We did not correct for these factors in our analyses and future work should be done to correct for these factors.

The population at risk for the two analysis is different. In the Sciensano database, it consists of all Flanders residents, while in the Intego database, it only consists of individuals seeking treatment of any kind within a year. We assume here that the patients seeking treatment are a random subset of the total population of Flanders, but that assumption may not be valid, as socioeconomically deprived individuals are known to adopt tendencies to visit GPs that are different from those of the rest of the population.

The most important advantage of using the Intego database is the ability to use a GP morbidity registry for a cost-effective and timely investigation of the dynamics in an epidemic. Since the Intego database collects data on many pathogens, it can be used to investigate the disease risk of other epidemics, such as flu, gastro-enteritis, RSV infections. Thus, we can use a GP morbidity registry as an alternative to predict and monitor an epidemic, with the advantage of the availability of detailed patient information. Next to these advantages, the Intego database is continuously updated, data are added on a daily basis, and more GPs and practices start to use the database, effectively adding more patients into it.

As part of future work, a joint analysis of both databases will be investigated which may result in more precise estimates, as the simultaneous modelling of both diseases allows both processes to borrow information from each other. This can be implemented via the use of, e.g., correlated random-effects structures [36, 37], which allow for shared stochasticity.

In conclusion, we were able to show how an alternative data source, the Intego data, can be used in a pandemic situation. We consider our findings useful for public health officials who plan intervention strategies aimed at bringing disease outbreaks such as the COVID-19 outbreak under control as well as to monitor disease outbreaks.

Appendix

Table 2 Model selection results based on WAIC and DIC values. Covariates effects: gender (G), age groups (A), comorbidity (C), categorical time effect per month (M) and their interactions. Random effects: spatially correlated heterogeneity (u_i), spatially uncorrelated heterogeneity (v_i), temporally structured effect modeled dynamically using a random walk of order 1 (γ_{t1}), temporally structured effect modeled dynamically using a random walk of order 2 (γ_{t2}), temporally structured effect modeled dynamically using an autoregressive random effect (γ_{t3}), unstructured temporal effect θ_i , spatio-temporal interaction between a spatially uncorrelated heterogeneity random effect and an unstructured time random effect (δ_{it}) and spatio-temporal interaction between a spatially uncorrelated heterogeneity random effect and a temporally structured effect modeled dynamically using a random walk of order 1 (ϕ_{it}). Blangiardo & Cameletti [23] provide explanations and applications of various spatio-temporal interactions

Model	Covariate effects									Random effects							WAIC	DIC	
	G	A	C	M	G*A	G*C	A*C	G*M	A*M	C*M	u_i	v_i	γ_{t1}	γ_{t2}	γ_{t3}	θ_i			δ_{it}
1	✓	✓	✓	✓								✓					✓	22436.06	22360.80
2	✓	✓	✓	✓							✓	✓					✓	22423.97	22347.64
3	✓	✓	✓									✓	✓					22966.20	22901.82
4	✓	✓	✓									✓		✓				22966.23	22901.80
5	✓	✓	✓									✓			✓			22966.17	22901.82
6	✓	✓	✓								✓	✓	✓			✓	✓	21247.00	21180.36
7	✓	✓	✓		✓						✓	✓	✓			✓	✓	21242.44	21176.52
8	✓	✓	✓			✓					✓	✓	✓			✓	✓	21238.16	21170.81
9	✓	✓	✓				✓				✓	✓	✓			✓	✓	21223.14	21155.51
10	✓	✓	✓		✓	✓	✓				✓	✓	✓			✓	✓	21232.70	21163.13
11	✓	✓	✓					✓			✓	✓	✓				✓	21200.66	21132.34
12	✓	✓	✓						✓		✓	✓	✓				✓	20679.03	20637.31
13	✓	✓	✓							✓	✓	✓	✓				✓	21096.19	21036.21
14	✓	✓	✓				✓	✓			✓	✓	✓				✓	20647.02	20605.71
15	✓	✓	✓				✓	✓	✓		✓	✓	✓				✓	20627.03	20589.03
16	✓	✓	✓				✓	✓	✓	✓	✓	✓	✓			✓	✓	20625.85	20586.26
17	✓	✓	✓								✓	✓	✓			✓	✓	21234.85	21168.21
18	✓	✓	✓								✓	✓	✓				✓	21235.67	21169.81
19	✓	✓	✓									✓	✓				✓	27145.73	27344.24

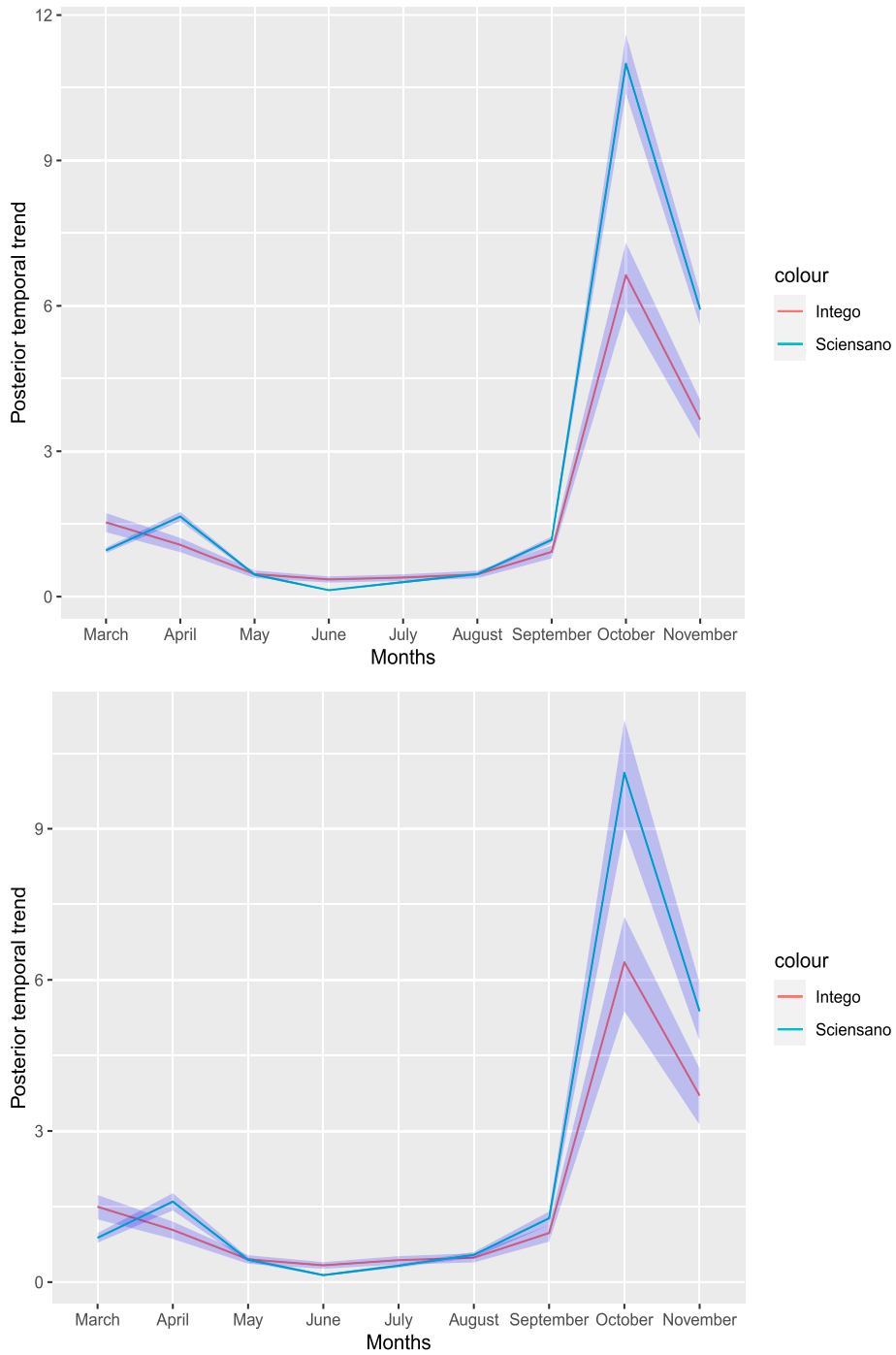


Fig. 5 The temporal structured random effect with 95% credibility intervals based on the municipality level (top) and based on the health sector level (bottom)

Table 3 Parameter estimates (odds ratios) at the health sector level for the Intego and Sciensano data with and without comorbidities

	Intego				Sciensano			
	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%
$\hat{\beta}_0$	0.0010	0.0001	0.0009	0.0011	0.0006	0.0000	0.0005	0.0006
$\hat{\beta}_1$	0.9453	0.0193	0.9076	0.9833	0.8198	0.0034	0.8131	0.8265
$\hat{\beta}_2$	2.0650	0.0759	1.9175	2.2150	2.9268	0.0231	2.8815	2.9722
$\hat{\beta}_3$	1.8293	0.0644	1.7042	1.9567	2.3598	0.0179	2.3248	2.3949
$\hat{\beta}_4$	1.1076	0.0460	1.0183	1.1986	2.3798	0.0197	2.3412	2.4184
$\hat{\nu}_i$	0.0932	0.0381	0.0315	0.1690	0.0545	0.0155	0.0315	0.16900
\hat{u}_i	0.1540	0.0735	0.0427	0.3002	0.0016	0.0032	0.0427	0.3002
$\hat{\gamma}_t$	0.6978	0.3579	0.2070	1.4112	1.2672	0.6164	0.2070	1.4112
$\hat{\delta}_{it}$	0.3559	0.0335	0.2918	0.4227	0.1985	0.0138	0.2918	0.4227
$\hat{\beta}_0$	0.0012	0.0001	0.0010	0.0013				
$\hat{\beta}_1$	0.9356	0.0191	0.8982	0.9733				
$\hat{\beta}_2$	2.0756	0.0763	1.9272	2.2265				
$\hat{\beta}_3$	1.7716	0.0628	1.6494	1.8958				
$\hat{\beta}_4$	0.9710	0.0433	0.8868	1.0567				
$\hat{\beta}_5$ (Chronic nierziekte)	1.3285	0.0981	1.1383	1.5228				
$\hat{\beta}_6$ (Harten vaatziekten)	1.1921	0.0431	1.1080	1.2772				
$\hat{\beta}_7$ (Diabetes)	1.2153	0.0527	1.1126	1.3193				
$\hat{\nu}_i$	0.1560	0.0426	0.0843	0.2425				
\hat{u}_i	0.0020	0.0045	0.0001	0.0069				
$\hat{\gamma}_t$	0.6789	0.3376	0.1879	1.3507				
$\hat{\delta}_{it}$	0.3028	0.0316	0.2425	0.3657				

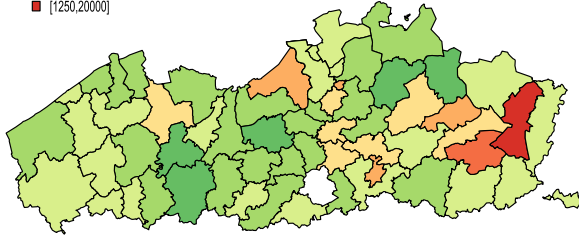
Table 4 The International Classification of Primary Care (ICPC) codes per comorbidity

Comorbidity	ICPC codes
Asthma	R96
Chronic liver disease	D97
Chronic lung disease	R79, R84, R95
Chronic neurological disorder	N74, N86, N87, N88
Chronic kidney disease	U99
Diabetes	T89, T90
Heart and vascular disease	K74, K75, K76, K77, K78, K80, K82, K83, K84, K89, K90, K91, K92
Hypertension	K86, K87
Immunodeficiency	B72, B73, B74, B90
Obesity	T82

March 2020

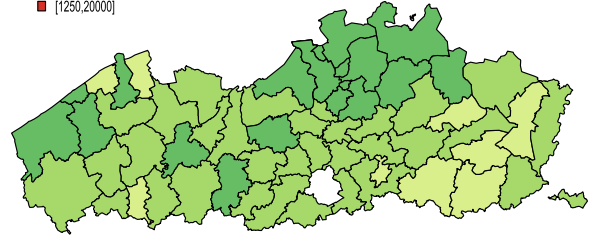
Corona cases in 100,000 inhabitants

- [0,75)
- [75,125)
- [125,250)
- [250,500)
- [500,750)
- [750,1000)
- [1000,1250)
- [1250,20000]



Corona cases in 100,000 inhabitants

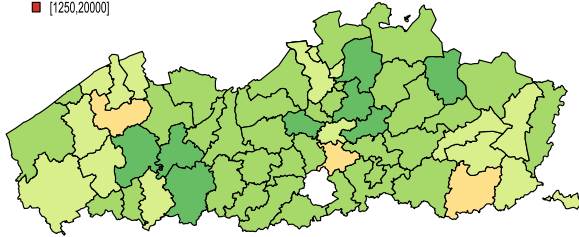
- [0,75)
- [75,125)
- [125,250)
- [250,500)
- [500,750)
- [750,1000)
- [1000,1250)
- [1250,20000]



April 2020

Corona cases in 100,000 inhabitants

- [0,75)
- [75,125)
- [125,250)
- [250,500)
- [500,750)
- [750,1000)
- [1000,1250)
- [1250,20000]



Corona cases in 100,000 inhabitants

- [0,75)
- [75,125)
- [125,250)
- [250,500)
- [500,750)
- [750,1000)
- [1000,1250)
- [1250,20000]

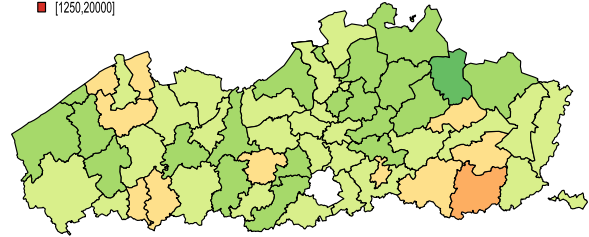
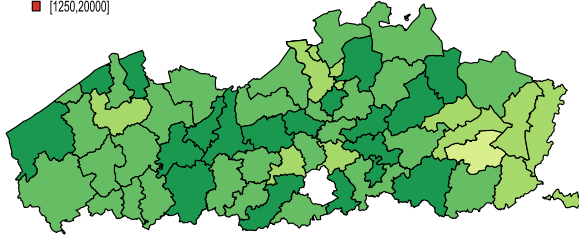
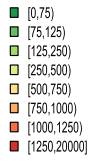


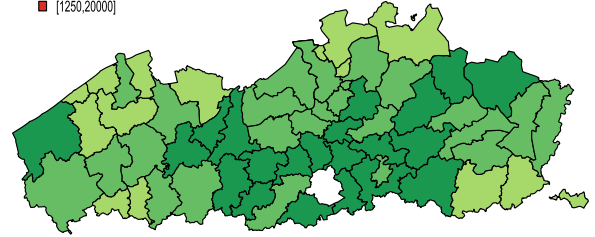
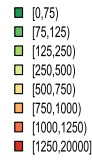
Fig. 6 Predicted number of corona cases in a population of 100,000 inhabitants using the Intego data (left) and the Sciensano data (right) at the health sector level

May 2020

Corona cases in 100,000 inhabitants

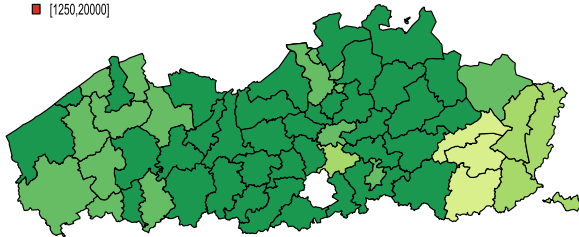
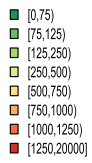


Corona cases in 100,000 inhabitants



June 2020

Corona cases in 100,000 inhabitants



Corona cases in 100,000 inhabitants

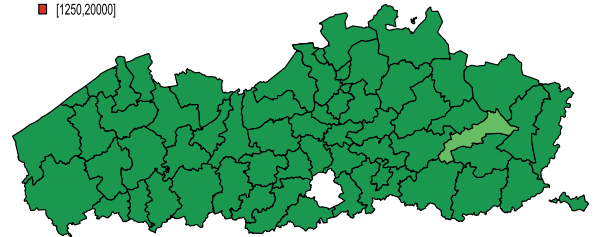
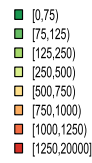
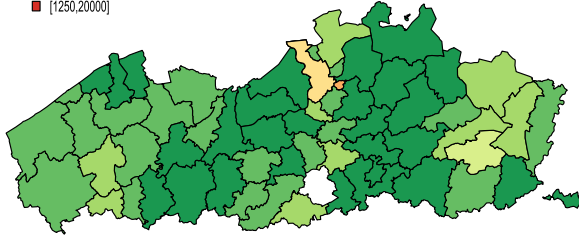


Fig. 7 Predicted number of corona cases in a population of 100,000 inhabitants using the Intego data (left) and the Sciensano data (right) at the health sector level

July 2020

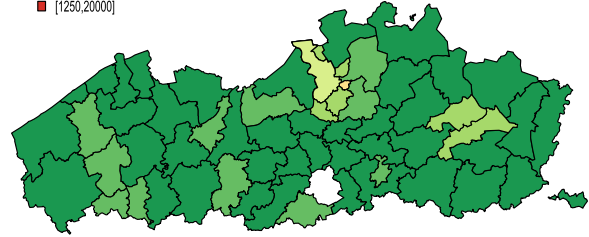
Corona cases in 100,000 inhabitants

- [0,75)
- [75,125)
- [125,250)
- [250,500)
- [500,750)
- [750,1000)
- [1000,1250)
- [1250,20000]



Corona cases in 100,000 inhabitants

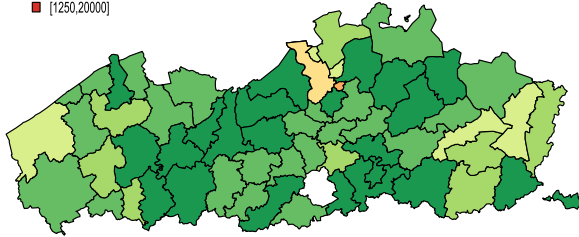
- [0,75)
- [75,125)
- [125,250)
- [250,500)
- [500,750)
- [750,1000)
- [1000,1250)
- [1250,20000]



August 2020

Corona cases in 100,000 inhabitants

- [0,75)
- [75,125)
- [125,250)
- [250,500)
- [500,750)
- [750,1000)
- [1000,1250)
- [1250,20000]



Corona cases in 100,000 inhabitants

- [0,75)
- [75,125)
- [125,250)
- [250,500)
- [500,750)
- [750,1000)
- [1000,1250)
- [1250,20000]

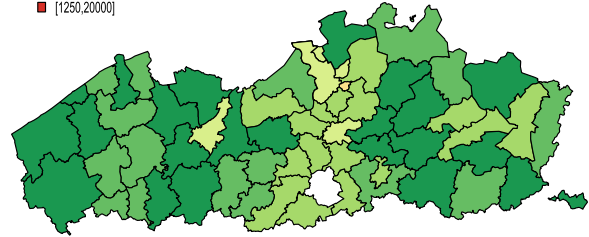
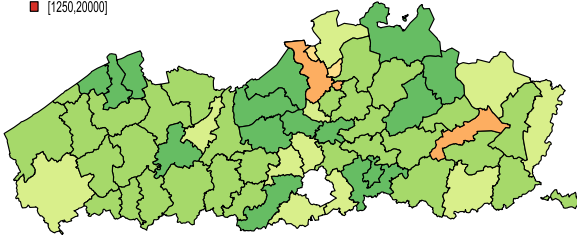
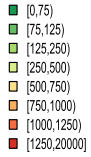


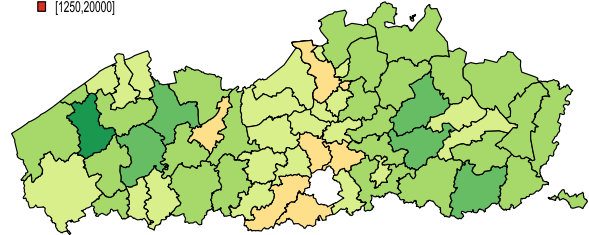
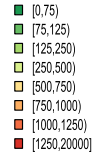
Fig. 8 Predicted number of corona cases in a population of 100,000 inhabitants using the Intego data (left) and the Sciensano data (right) at the health sector level

September 2020

Corona cases in 100,000 inhabitants

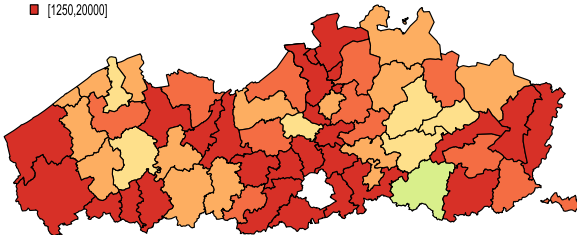
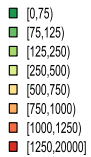


Corona cases in 100,000 inhabitants



October 2020

Corona cases in 100,000 inhabitants



Corona cases in 100,000 inhabitants

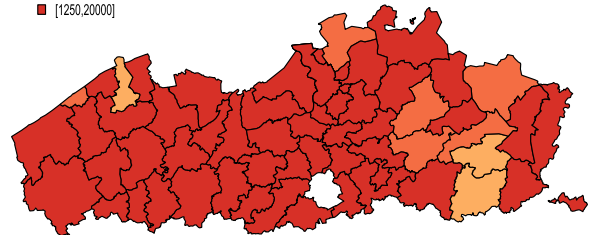
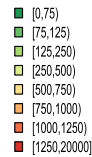
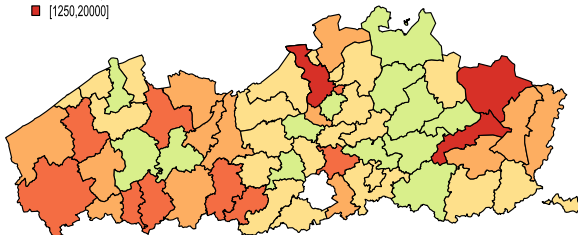
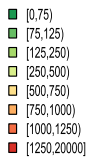


Fig. 9 Predicted number of corona cases in a population of 100,000 inhabitants using the Intego data (left) and the Sciensano data (right) at the health sector level

November 2020

Corona cases in 100,000 inhabitants



Corona cases in 100,000 inhabitants

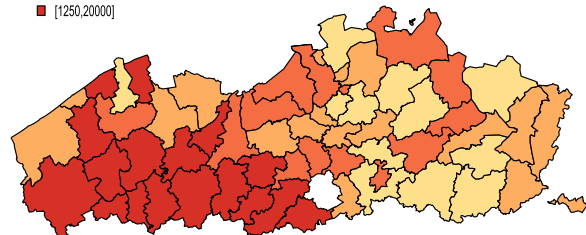
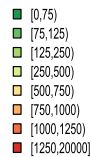


Fig. 10 Predicted number of corona cases in a population of 100,000 inhabitants using the Intego data (left) and the Sciensano data (right) at the health sector level

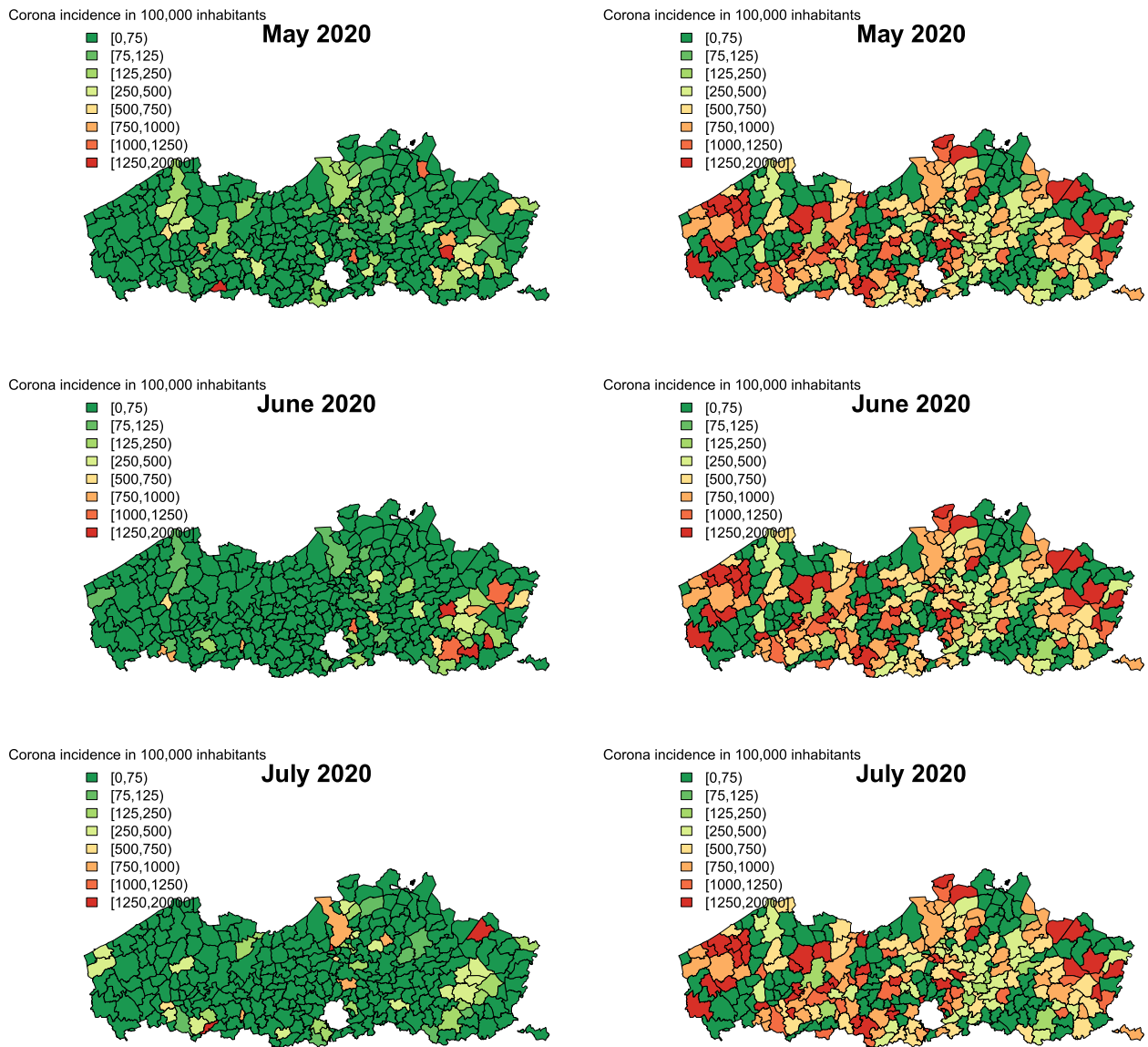


Fig. 11 Incidence of positive COVID-19 cases in a population of 100,000 inhabitants using the Intego data (left) and the Sciensano data (right)

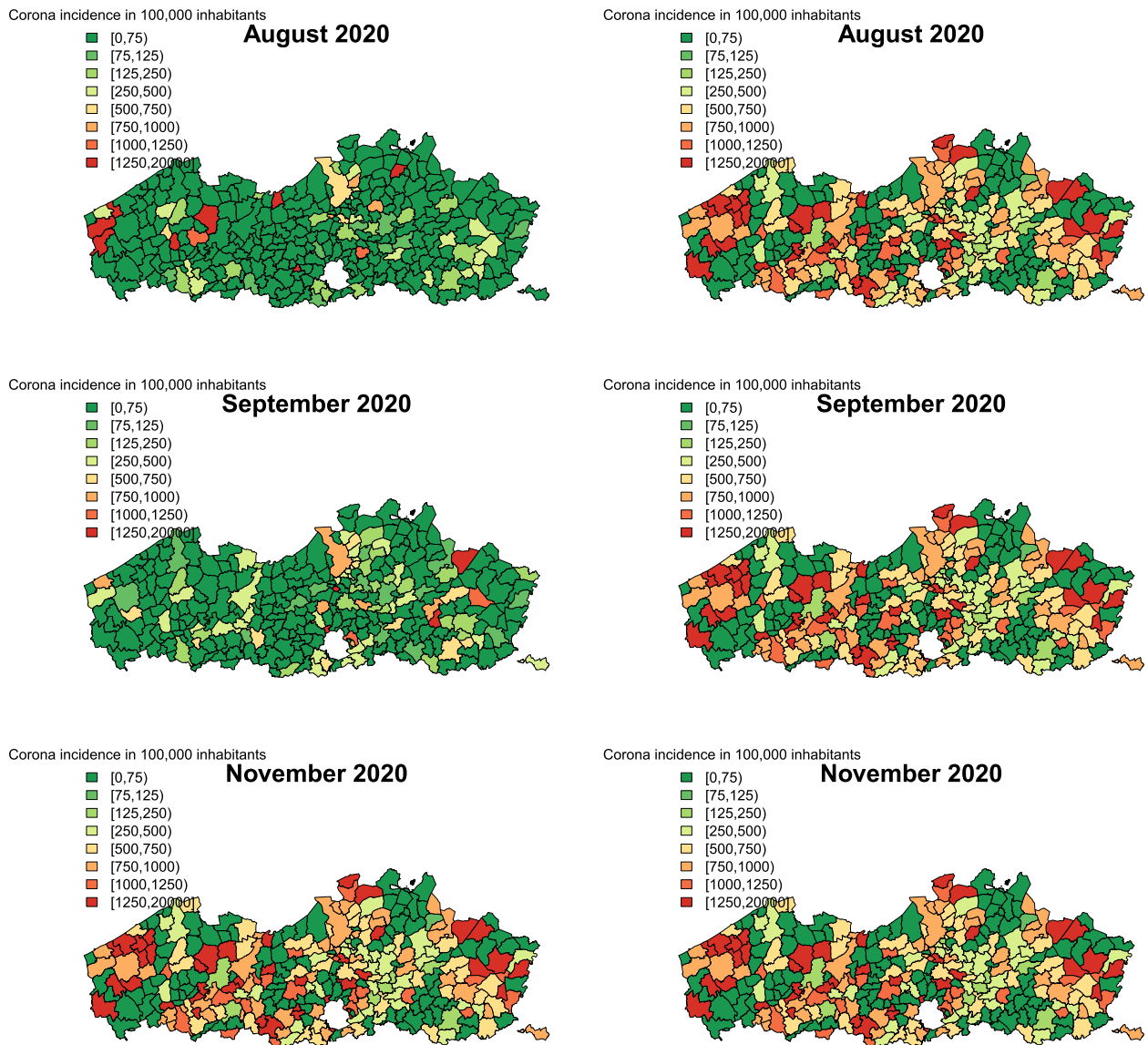


Fig. 12 Incidence of positive COVID-19 cases in a population of 100,000 inhabitants using the Intego data (left) and the Sciensano data (right)

Acknowledgements

Not applicable.

Authors' contributions

All authors contributed to the study conception and design. The data was provided by BV. Data analysis was performed by OP, AJ, CF and TN. The methods were chosen under the guidance of CF and TN. The first draft of the manuscript was written by OP and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. The authors confirm that all MAP/IMAGE/ILLUSTRATIONS depicted in the figures are our own.

Funding

TN gratefully acknowledges funding by the Internal Funds KU Leuven (project number 3M190682).

Availability of data and materials

The datasets generated and analysed during the current study are not publicly available due to inclusion of protected health information but can be made available subsequent to further de-identification upon reasonable request to the corresponding author (OP).

Declarations

Ethics approval and consent to participate

The Intego project is approved by the ethical review board of the Research Ethics Committee of the UZ Leuven University Hospitals (number ML 1723) and by the Belgian Privacy Commission (number SCSZG/13/079). Intego was waived the need for individual informed consent, but operates under an opt-out procedure for patients who do not wish their data to be included. This was approved by the aforementioned ethical review board. The study methods

were performed in accordance with the relevant guidelines and regulations. Patient data were pseudonymized before use.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 3 August 2022 Accepted: 8 May 2024

Published online: 21 May 2024

References

- Baloch S, Baloch MA, Zheng T, Pei X. The coronavirus disease 2019 (COVID-19) pandemic. *Tohoku J Exp Med.* 2020;250(4):271–8.
- Johns Hopkins Staff. What Is Coronavirus? Johns Hopkins medicine website. 2021. <https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus>. Accessed 18 Oct 2021.
- COVID-19 surveillance. <https://www.healthbelgium.be/en/health-status/54-infectious-diseases#read-more>. Accessed 19 Oct 2021.
- Wikipedia. COVID-19 pandemic in Belgium. Wikipedia, the free encyclopedia website. 2021. https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Belgium#October_to_April_2021_%E2%80%93_return_to_lockdown. Accessed 19 Oct 2021.
- Peeters I, Vermeulen M, Bustos Sierra N, Renard F, VanderHeyden J, Scohy A, Braeye T, Bossuyt N, Haarhuis F, Proesmans K, Vernemmen C, Vanhaverbeke M. Surveillance of COVID-19 mortality in Belgium, epidemiology and methodology during 1st and 2nd wave (March 2020 - 14 February 2021). Brussels: Sciensano; 2021. <https://covid-19.sciensano.be/fr/covid-19-situation-epidemiologique>. Report number: D/2021/14.440/57.
- Sciensano & Coronavirus. What role does Sciensano play? Sciensano website. 2021. <https://www.sciensano.be/en/health-topics/coronavirus/role>. Accessed 20 Oct 2021.
- Bustos Sierra N, Bossuyt N, Braeye T, Haarhuis F, Peeters I, Proesmans K, Renard F, Scohy A, Vanhaverbeke M, Vermeulen M, Vernemmen C, Van der Heyden J. Excess mortality during the first and second waves of the COVID-19 epidemic in Belgium (data from 10 March 2020 to 14 February 2021). Brussels: Sciensano; 2021. <https://epistat.wiv-isp.be/momo/>. Report number : D/2021/14.440/62.
- Sanyaolu A, Okorie C, Marinkovic A, et al. Comorbidity and its Impact on Patients with COVID-19. *SN Compr Clin Med.* 2020;2:1069–76. <https://doi.org/10.1007/s42399-020-00363-4>.
- Guan WJ, Liang WH, He JX, et al. Cardiovascular comorbidity and its impact on patients with COVID-19. *Eur Respir J.* 2020;55:2001227. <https://doi.org/10.1183/13993003.01227-2020>.
- Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Cao B. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet.* 2020;395(10229):1054–62.
- Liu K, Chen Y, Lin R, Han K. Clinical features of COVID-19 in elderly patients: a comparison with young and middle-aged patients. *J Infect.* 2020;80(6):e14–8.
- Ejaz H, Alsrhani A, Zafar A, Javed H, Junaid K, Abdalla AE, Younas S. COVID-19 and comorbidities: deleterious impact on infected patients. *J Infect Public Health.* 2020;13(12):1833–1839.
- Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Peng Z. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan. *China JAMA.* 2020;323(11):1061–9.
- Truyers C, Goderis G, Dewitte H, vanden Akker M, Buntinx F. The Intego database: background, methods and basic results of a Flemish general practice-based continuous morbidity registration project. *BMC Med Inform Decis Making.* 2014;14(1):1–9.
- Clinical Practice Research Datalink. <https://cprd.com/>. Accessed 20 Oct 2021.
- The Health Improvement Network. <https://www.the-health-improvement-network.com/?hsLang=en>. Accessed 10 Oct 2021.
- Hasselaar J. Nivel Primary Care Database. 2021. <https://www.nivel.nl/en/nivel-zorgregistraties-eerste-lijn/nivel-primary-care-database>. Accessed 19 June 2022.
- Vaes B, Vos B, Foidart M, De Schreye R, Schrans D, Philips H, Aertgeerts B, Doggen K. Burden of COVID-19 on primary care in Belgium: a prospective nationwide observational study from March to August 2020. *Arch Public Health.* 2020;80(1):250.
- Bartholomeeusen S, Kim CY, Mertens R, Faes C, Buntinx F. The denominator in general practice, a new approach from the Intego database. *Fam Pract.* 2005;22(4):442–7.
- Creating 21st century primary care in Flanders and beyond. Copenhagen: WHO Regional Office for Europe; 2019. Licence: CC BY-NC-SA 3.0 IGO. <https://www.who.int/belgium/publications/item/9789289054164>. Accessed 19 Oct 2021.
- Pasutto C, Wakefield JC, Best NG, Richardson S, Bernardinelli L, Staines A, Elliott P. Statistical issues in the analysis of disease mapping data. *Stat Med.* 2000;19(17–18):2493–519.
- Lawson AB. Bayesian disease mapping: hierarchical modeling in spatial epidemiology. CRC Press; 2009.
- Blangiardo M, Cameletti M. Spatial and spatio-temporal Bayesian models with R-INLA. John Wiley & Sons; 2015.
- Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. *Environ Health Perspect.* 2004;112(9):998–1006.
- Knorr-Held L. Bayesian modelling of inseparable space-time variation in disease risk. *Stat Med.* 2000;19(17–18):2555–67.
- Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J Mach Learn Res.* 2010;11(Dec):3571–94.
- Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Ser B Stat Methodol.* 2002;64(4):583–639.
- Besag J, Kooperberg C. On conditional and intrinsic autoregressions. *Biometrika.* 1995;82(4):733–46.
- Casella G, George EI. Explaining the Gibbs sampler. *Am Stat.* 1992;46(3):167–74.
- Gilks WR, Richardson S, Spiegelhalter D, editors. Markov Chain Monte Carlo in Practice. 1st ed. Chapman and Hall/CRC; 1995. <https://doi.org/10.1201/b14835>.
- Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J R Stat Soc Ser B Stat Methodol.* 2009;71(2):319–92.
- R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2020. <https://www.R-project.org/>.
- Eberly LE, Carlin BP. Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. *Stat Med.* 2000;19(17–18):2279–94.
- Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, Zhou Y. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *Int J Infect Dis.* 2020;94:91–5.
- Neyens T, Faes C, Vranckx M, Pepermans K, Hens N, Van Damme P, Molenberghs G, Aerts J, Beutels P. Can COVID-19 symptoms as reported in a large-scale online survey be used to optimise spatial predictions of COVID-19 incidence risk in Belgium? *Spat Spatio-Temporal Epidemiol.* 2020;35:100379.
- Neyens T, Lawson AB, Kirby RS, Faes C. The bivariate combined model for spatial data analysis. *Stat Med.* 2016;35(18):3189–202.
- Vranckx M, Faes C, Molenberghs G, Hens N, Beutels P, Van Damme P, Aerts J, Petrof O, Pepermans K, Neyens T. A spatial model to jointly analyze self-reported survey data of COVID-19 symptoms and official COVID-19 incidence data. *Biom J.* 2022;65(1):2100186.

Publisher's Note

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