

Predicting COVID-19 hospitalizations: The importance of healthcare hotlines, test positivity rates and vaccination coverage

Vera van Zoest^{a,b,*}, Karl Lindberg^{a,c}, Georgios Varotsis^c, Frank Badu Osei^d, Tove Fall^c

^a Department of Information Technology, Uppsala University, P.O. Box 337, Uppsala 751 05, Sweden

^b Department of Systems Science for Defence and Security, Swedish Defence University, P.O. Box 27805, Stockholm 115 93, Sweden

^c Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala 751 85, Sweden

^d Faculty of Geo-Information Science and Earth Observation, University of Twente, P.O. Box 217, Enschede 7500 AE, the Netherlands

ARTICLE INFO

Keywords:

COVID-19

Negative binomial regression

Spatio-temporal modeling

Time series

Prediction

ABSTRACT

In this study, we developed a negative binomial regression model for one-week ahead spatio-temporal predictions of the number of COVID-19 hospitalizations in Uppsala County, Sweden. Our model utilized weekly aggregated data on testing, vaccination, and calls to the national healthcare hotline. Variable importance analysis revealed that calls to the national healthcare hotline were the most important contributor to prediction performance when predicting COVID-19 hospitalizations. Our results support the importance of early testing, systematic registration of test results, and the value of healthcare hotline data in predicting hospitalizations. The proposed models may be applied to studies modeling hospitalizations of other viral respiratory infections in space and time assuming count data are overdispersed. Our suggested variable importance analysis enables the calculation of the effects on the predictive performance of each covariate. This can inform decisions about which types of data should be prioritized, thereby facilitating the allocation of healthcare resources.

1. Background

The COVID-19 pandemic had a major impact on our daily lives in recent years. To mitigate the spread of infection, Public Health Agencies across the globe implemented control measures, such as border restrictions and recommendations for social distancing. These measures shared one common goal: to slow down the spread of infection while alleviating the pressure on overloaded healthcare systems and preventing deaths attributed to COVID-19. A key indicator of pressure on the healthcare system is the number of hospital beds occupied by patients with severe COVID-19. Predicting hospital occupancy trends in both space and time is crucial for resource allocation and planning temporary capacity increases (e.g. field hospitals).

Numerous attempts have been made to predict short-term and long-term COVID-19-associated hospitalizations. A common approach is the use of a compartmental model, such as SEIR models, where the total population is divided into different so-called ‘compartments’ (e.g. exposed, infected, hospitalized, recovered) (Gerlee et al., 2021). These models predict the probabilities of movement between compartments and the number of hospitalizations under certain assumptions such as infectivity, incubation period, and virulence. Several variations of this

model exist, including the addition of a compartment to account for isolated infected populations (Reno et al., 2020), and the use of polynomial regression to correct for model errors (Gatto et al., 2021). Time series models have also been used to predict hospitalizations based on previous trends and temporal autocorrelation (Chelo et al., 2021; Perone, 2021). Wesner et al. (2021) built a Bayesian non-linear regression model with priors based on hospitalization data from another area that had already completed a disease curve. Other research teams have incorporated external regressors in the models, such as wastewater samples (Galani et al., 2022), Google search term activity, and health chatbot scores (Turk et al., 2021). A separate line of work has focused on predictions at patient-level, e.g. forecasting the need for hospitalization, intensive care, and respiratory support, as well as mortality rates based on patient-level characteristics such as age, gender, comorbidities, and socioeconomic status (Patricio et al., 2021; Simpson et al., 2020; Wolstenstein-Betech et al., 2020). By summing the predicted needs of individual patients, the needs at hospital-level can be extrapolated.

Most previous works have focused on predicting the total number of hospitalizations within a single hospital or a delimited region, which is highly relevant for local planning and resource allocation. In addition to the total number of hospitalizations in a region, it would also be

* Corresponding author at: Department of Information Technology, Uppsala University, P.O. Box 337, Uppsala 751 05, Sweden.

E-mail address: vera.van.zoest@it.uu.se (V. van Zoest).

<https://doi.org/10.1016/j.sste.2024.100636>

Received 29 June 2023; Received in revised form 6 December 2023; Accepted 16 January 2024

Available online 17 January 2024

1877-5845/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

beneficial to predict the spatial variability of hospitalizations within a larger region served by multiple hospitals. These predictions would be of importance for scheduling healthcare staff, supplies, hospital bed capacity, and patient transfers to other hospitals. They can also assist public health authorities in adjusting their infection control strategies. A high number of hospitalizations in a certain subregion (e.g. municipality) compared to other subregions may be a sign of lacking early-stage measures such as testing or vaccination. Fine-scale spatio-temporal predictions can provide further insights into potential differences due to socioeconomic disparities between different municipalities and the need for local interventions, e.g. by enhancing communication on social distancing and vaccination, and increasing local testing capacity.

The main objective of this study is to develop methods to predict COVID-19-related hospitalizations in space and time within the borders of Uppsala County, Sweden. Predictions were made a week ahead in time for each of the county's eight municipalities. We use a set of covariates as external regressors and evaluate their individual effects on predictive performance. Besides providing a prediction model with a high spatio-temporal resolution, we aim to guide future researchers, healthcare agencies, and policymakers in selecting the most effective variables in the event of a pandemic.

2. Methods

2.1. Data

In this study, we focus on Uppsala County, Sweden which has a population of nearly 390,000 inhabitants (SCB, 2020) spread across eight different municipalities: Enköping, Håbo, Heby, Knivsta, Tierp, Uppsala, Älvkarleby, and Östhammar. Only Enköping and Uppsala contain a hospital with inpatient care; patients from other municipalities are referred to one of these two hospitals. Patients were considered as COVID-19 patients based on their PCR test results, although the primary reason for hospitalization may not have been related to their infection. We used data from week 26, 2020 to week 41, 2021, collected as part of the CRUSH Covid project (van Zoest et al., 2022). We considered the following variables as potential external regressors (covariates) in the model to predict hospitalizations at the municipality level in week t : number of hospitalizations in the previous week (week $t - 1$), positivity rate in PCR tests in week $t - 1$, number of COVID-19 tests in week $t - 1$, number of COVID-related calls to emergency line 112 in week $t - 1$, number of COVID-related calls to healthcare advice hotline 1177 in week $t - 1$, the proportion of the population aged 50+ years with at least two vaccinations against COVID-19 taken at least two weeks ago in week $t - 1$, and municipality (categorical). COVID-19 related calls to the national healthcare hotline 1177 were available from week 48, 2020 and we used vaccination data from week 12, 2021 onwards. A time lag of two weeks after the second dose was applied to vaccination data to allow for immunity development. Table 1 provides an overview of the covariates used in this study.

2.2. Prediction model

Our variable of interest is the number of COVID-19 related hospitalizations per week, counted as the number of bed-days occupied by individuals residing in the municipality. We make use of a negative binomial regression model to account for overdispersion in the counts. Thus, we consider the number of hospitalizations $y_{s,t}$ in municipality s in week t to arise from a negative binomial (NB) distribution with mean $\lambda_{s,t}$ and variance $\lambda_{s,t} + \lambda_{s,t}^2/\theta_t$, where θ_t is the dispersion parameter (Hilbe, 2014):

$$y_{s,t} \sim NB\left(\lambda_{s,t}, \lambda_{s,t} + \frac{\lambda_{s,t}^2}{\theta_t}\right). \quad (1)$$

Here, $\lambda_{s,t}$ equals the population P_s in area s multiplied by the

Table 1

Overview of covariate data used.

| Variable | Description | Data type |
|---|---|---|
| Municipality | The name of the municipality corresponding to municipality s . | Fixed spatial-dependent categorical |
| Test positivity (week $t - 1$) | The percentage of all COVID-19 RT-PCR tests that were positive in municipality s during week $t - 1$. | Spatial and time-dependent continuous numerical |
| Number of COVID-19 RT-PCR tests (week $t - 1$) | The total number of COVID-19 RT-PCR tests in week $t - 1$ among inhabitants of municipality s . | Spatial and time-dependent continuous numerical |
| Number of hospitalizations (week $t - 1$) | The total number of hospitalizations, defined as number of hospital bed-days during week $t - 1$, occupied by patients who tested positive for COVID-19 and who were living in municipality s . | Spatial and time-dependent continuous numerical |
| COVID-19-related calls to 1177 (week $t - 1$) | Number of calls to the 1177 Healthcare Advice Line from people reporting COVID-19-related symptoms in municipality s and week $t - 1$. | Spatial and time-dependent continuous numerical |
| Number of calls assessed as suspected COVID-19 by the ambulance personnel (week $t - 1$) | Number of ambulance calls to the 112 emergency line in municipality s and week $t - 1$ that were assessed as 'suspected COVID-19' by the ambulance personnel. | Spatial and time-dependent continuous numerical |
| Vaccinated proportion of population aged 50+ (dose 2 at least 2 weeks ago) | Proportion of individuals aged 50+ residing within the administrative borders of municipality s who have completed their first full vaccination series consisting of two doses of an EMA-approved vaccine against COVID-19 at least two weeks before week t . | Spatial and time-dependent continuous numerical |

hospitalized proportion of the population $p_{s,t}$ in area s in week t :

$$\lambda_{s,t} = P_s * p_{s,t} \quad (2)$$

which can be rewritten as:

$$\log(\lambda_{s,t}) = \log(P_s) + \log(p_{s,t}) \quad (3)$$

The population P_s is the offset of the model and considered to be static during the period under study. We consider a multiple linear regression model to predict $\log(p_{s,t})$ for area s in week t :

$$\log(p_{s,t}) = \beta_{0,t-1} + \sum_{k=1}^K \beta_{k,t-1} x_{k,s,t-1} + \beta_{y,t-1} y_{s,t-1} \quad (4)$$

where $\beta_{0,t-1}$ is the intercept in week $t - 1$, $x_{k,s,t-1}$ are the values for covariates $k \in \{1 \dots K\}$ in week $t - 1$ in municipality s , which are multiplied by their respective coefficients $\beta_{k,t-1}$, $y_{s,t-1}$ are the lagged hospitalizations from the week $t - 1$ with coefficient $\beta_{y,t-1}$. The set of covariates includes the variables from Table 1 as well as interaction effects between municipality and one-week lagged number of hospitalizations, and interaction effects between municipality and number of COVID-19 RT-PCR tests. A one-week lag is selected based on visual inspection of the partial autocorrelation function plots for hospitalizations. Combining Eqs. (3) and (4), we can thus predict $\log(\lambda_{s,t})$ as:

$$\log(\widehat{\lambda_{s,t}}) = \log(P_s) + \widehat{\beta}_{0,t-1} + \sum_{k=1}^K \widehat{\beta}_{k,t-1} x_{k,s,t-1} + \widehat{\beta}_{y,t-1} y_{s,t-1} \quad (5)$$

where $\hat{\beta}_{0,t-1}$, $\hat{\beta}_{k,t-1}$ and $\hat{\beta}_{y,t-1}$ are the coefficients estimated using Eq. (4). The coefficients are estimated using Maximum Likelihood using the MASS package (version 7.3–60) in R (version 4.3.1) (Ripley et al., 2022). Our interest is in the odds ratios (OR) for the different covariates, defined as $OR_{k,t-1} = \exp(\hat{\beta}_{k,t-1})$, and their variability over time.

2.3. Performance evaluation

To evaluate the model's out-of-sample performance over time, we used the Root Mean Squared Error (RMSE) computed iteratively for each week t over all areas s defined as $RMSE_t$:

$$RMSE_t = \sqrt{\frac{\sum_{s=1}^{N_s} (\hat{y}_{s,t} - y_{s,t})^2}{N_s}} \quad (6)$$

where $\hat{y}_{s,t}$ is the predicted number of hospitalizations and $y_{s,t}$ is the observed number of hospitalizations in municipality s during week t , and N_s is the total number of municipalities. We employ a moving window approach, with new data incorporated weekly allowing us to use all previous data for model training. The $RMSE_t$ is updated accordingly, computed using data from all areas s , using validation data from week t . The RMSE is calculated over out-of-sample predictions, thus only using data from week 1 to $t-1$ to predict the number of hospitalizations in week t . Lower values of $RMSE_t$ indicate better performance.

To evaluate the importance of individual covariates in the model, we removed individual covariates from the model one at a time, and evaluated the impact on the prediction performance using the RMSE values.

3. Results

Table 2 shows summary statistics characterizing the eight municipalities in Uppsala County. Table 3 shows the mean and standard deviation of the variables included in the model, for the entire study period, for each of the eight municipalities. Fig. 1 shows the time series for the estimated coefficients $\hat{\beta}_{k,t}$ for the different covariates k in the model, except for the categorical variable municipality and its interaction effect with number of tests, for which the figures are included in the Supplementary Materials (Figs. S1 and S2). Fig. S3 in the Supplementary Materials shows the time series of the beta coefficients for the interaction effects between municipality and one-week lagged hospitalizations, i.e. the covariate covering spatio-temporal variability. The 95 % confidence intervals for the estimates are shown in gray, and a dashed line indicates zero (null hypothesis). Overlap between the gray area and the dashed line for certain weeks t indicates that the covariate had no significant ($\alpha = 0.05$) association with the prediction of the log hospitalized proportion of the population. When the confidence intervals for the estimates do not overlap with the dashed line, the covariate has a significant effect on the predictions, which can either be positive or negative. Notably, we observe that test positivity clearly and significantly contributed to the prediction of hospitalizations during the first 25 weeks of 2021. A strong decrease in the beta coefficient is visible in week 12, 2021, the moment in which vaccination data enters the model.

Table 2

Summary statistics characterizing the eight municipalities in Uppsala County.

| Municipality | Population | Proportion female | Mean age | Proportion aged 50+ | Population density (pop./km ²) | Proportion foreign background | Proportion higher educated | Driving distance to closest hospital (driving time) |
|--------------|------------|-------------------|----------|---------------------|--|-------------------------------|----------------------------|---|
| Häbo | 22,213 | 0.49 | 40.0 | 0.36 | 155.7 | 0.22 | 0.30 | 30.8 km (26 min.) |
| Älvkarleby | 9,614 | 0.49 | 43.7 | 0.44 | 43.5 | 0.19 | 0.24 | 92.8 km (65 min.) |
| Knivsta | 19,515 | 0.49 | 36.8 | 0.29 | 70.2 | 0.19 | 0.45 | 22.3 km (21 min.) |
| Heby | 14,247 | 0.49 | 43.8 | 0.44 | 12.3 | 0.16 | 0.22 | 51.6 km (50 min.) |
| Tierp | 21,426 | 0.49 | 43.3 | 0.42 | 14.0 | 0.17 | 0.23 | 83 km (67 min.) |
| Uppsala | 235,573 | 0.51 | 38.9 | 0.32 | 108.9 | 0.29 | 0.53 | 22 km (27 min.) |
| Enköping | 47,053 | 0.49 | 41.3 | 0.38 | 40.3 | 0.22 | 0.30 | 10.9 km (12 min.) |
| Östhammar | 22,347 | 0.49 | 45.5 | 0.47 | 15.2 | 0.12 | 0.24 | 70 km (61 min.) |

The number of tests performed seems to have a slight positive association with hospitalizations, which increases with time. In addition, the lagged hospitalizations have a positive association, indicating temporal autocorrelation. This positive association only becomes significant after week 25, when test positivity is no longer a significant covariate. Meanwhile, the vaccinated proportion of the population aged 50+ starts to negatively affect hospitalizations after week 25, 2021, with a very narrow confidence interval. This corresponds with approximately 39 % (range 32–48 % depending on the municipality) of the population aged 50+ being fully vaccinated, i.e. having received a second dose at least 2 weeks ago – the approximate time needed for the body to build up immunity to the virus following a second dose (Feikin et al., 2022). Contrary to expectations, calls to the national healthcare hotline 1177 seem to have no significant effect; but this is likely caused by high collinearity with number of tests. However, as the prediction performance strongly decreases after removing one of the two correlated variables, we decided to keep both variables in the prediction model. The number of calls to emergency line 112 related to COVID-19 symptoms is only significantly different from zero between week 15 and week 25 ($\alpha = 0.05$), likely because the variability is captured by other variables in the model. Fig. S3 in the Supplementary Materials shows the spatio-temporal variability covered by the interaction effect between municipality and one-week lagged hospitalizations. In most cases, it is significantly different from zero (baseline Enköping municipality), indicating significant spatio-temporal variability.

Fig. 2 shows the predicted versus the observed number of hospitalizations over time for each of the eight municipalities. In general, the predictions appear to follow the same pattern as the actual hospitalizations. Fig. 3 shows the difference between the predicted and actual number of hospital beds occupied by COVID-19 patients from all municipalities in the County, converted from weekly total to daily average. Values above zero indicate overprediction, values below zero indicate underprediction.

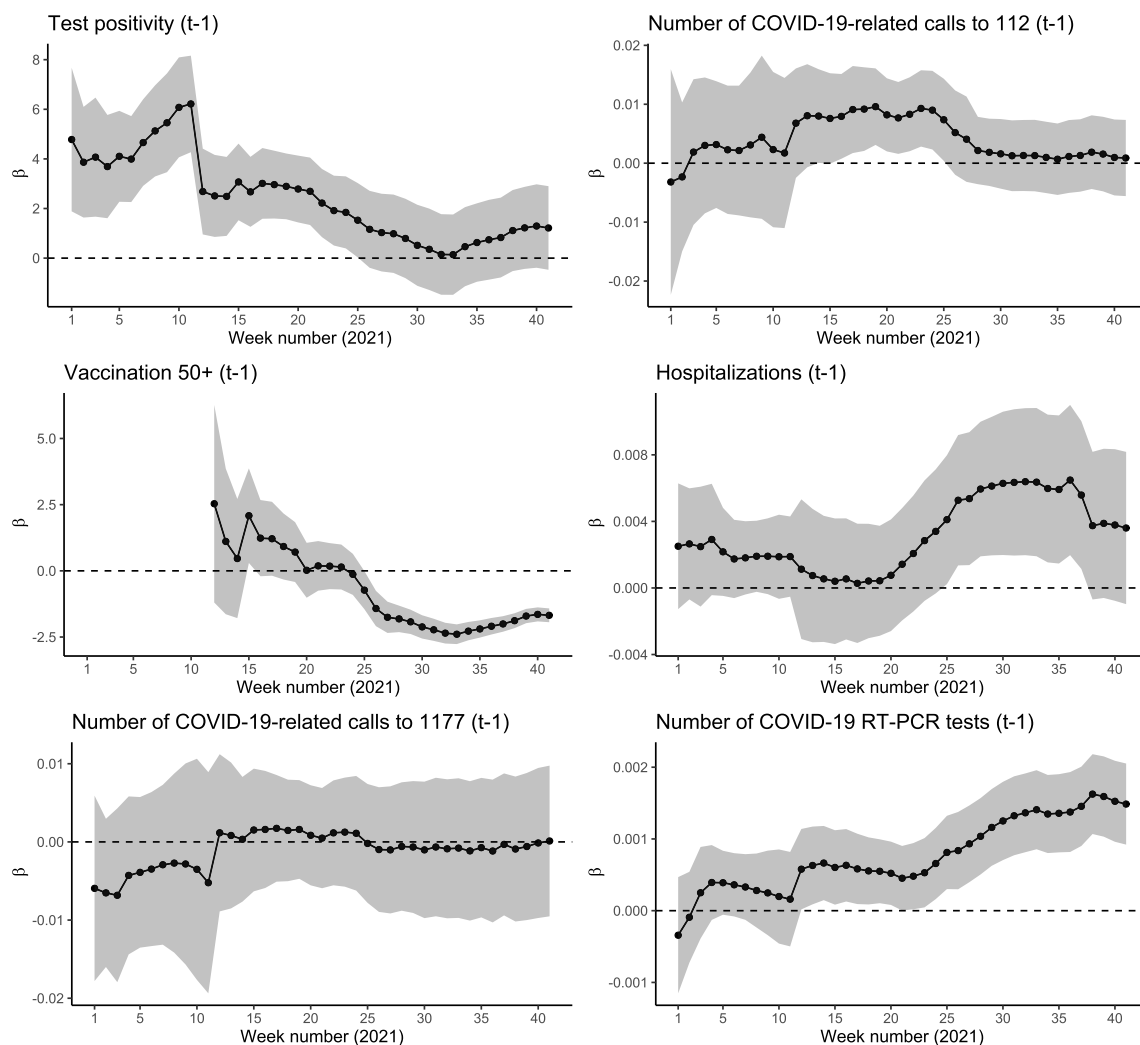
Table 4 shows the average RMSE value for the full model, as well as the average RMSE when removing one variable at a time, an indication of variable importance.

Fig. 4 shows the RMSE values over time. Here, we can also evaluate variable importance over time. Higher RMSE indicates lower performance. Thus, it can clearly be observed that removing a variable like the number of COVID-19 related calls to healthcare hotline 1177 has a major impact on prediction performance, especially in the peak of the pandemic. The interaction effects also have the biggest impact during the peak of the pandemic.

Fig. 5 shows the spatial variability in the observed and predicted number of hospitalizations per 1000 inhabitants during the peak of the third wave of the COVID-19 pandemic (week 17, 2021). The model effectively captures a large part of the spatial variability in the data. However, for Östhammar municipality, located on the northeast side of the region, the predictions overestimate the observed values. As can be derived from Fig. 2, there seems to be a systematic bias for a longer period of time, in which the predictions for Östhammar overestimate the observed values.

Table 3Mean (std. dev.) of the variables included in the model, averaged over the weeks in the study period, except for vaccinated individuals².

| Municipality | Hospitalizations # bed days/week ¹ | Positivity rate | RT-PCR tests ¹ # | Calls to 112 ¹ # | Calls to 11,77 ¹ # | Vaccinated individuals 50+ ² |
|--------------|--|-----------------|--------------------------------|--------------------------------|----------------------------------|---|
| Håbo | 0.92 (0.99) | 0.08 (0.06) | 18.13 (11.63) | 0.3 (0.22) | 0.23 (0.15) | 0.35 |
| Älvkarleby | 1.43 (1.66) | 0.08 (0.07) | 16.40 (10.02) | 0.46 (0.31) | 0.22 (0.20) | 0.48 |
| Knivsta | 0.54 (0.53) | 0.07 (0.06) | 18.40 (10.93) | 0.28 (0.14) | 0.19 (0.13) | 0.41 |
| Heby | 0.83 (0.88) | 0.05 (0.05) | 17.12 (11.39) | 0.42 (0.21) | 0.21 (0.15) | 0.48 |
| Tierp | 1.21 (1.23) | 0.07 (0.05) | 18.67 (11.15) | 0.47 (0.3) | 0.22 (0.13) | 0.32 |
| Uppsala | 1.03 (0.9) | 0.07 (0.05) | 19.65 (10.69) | 0.32 (0.11) | 0.26 (0.12) | 0.39 |
| Enköping | 1.11 (1.2) | 0.07 (0.06) | 19.46 (10.26) | 0.43 (0.2) | 0.28 (0.11) | 0.36 |
| Östhammar | 1.33 (1.56) | 0.06 (0.05) | 17.49 (9.38) | 0.43 (0.19) | 0.12 (0.10) | 0.43 |

¹ Per 1000 inhabitants.² Proportion of the population aged 50+, which had received a second vaccination at least 2 weeks ago, in week 25, 2021, when the vaccination variable in the model had a significant impact on decreasing hospitalizations.**Fig. 1.** Time series of the β -coefficients of the model for all covariates except municipality. The gray area shows the confidence interval of the estimates. The dashed line indicates the zero-effect line.

4. Discussion

In this paper we propose a spatio-temporal, regression-based model for predicting the number of hospitalizations in Uppsala County during the second and third wave of the COVID-19 pandemic in Sweden. The set of covariates used in the model reflected the spatial, temporal and spatio-temporal variations in hospitalizations in the different municipalities within the region.

Our study, which focused on the COVID-19 hospitalizations in a county in Sweden, found calls to healthcare hotline 1177 for symptoms related to COVID-19 to be the most important predictor for hospitalizations. Healthcare hotline data was also found a useful predictor in another study, using a different methodology in a different region in Sweden (Spreco et al., 2022). Positivity rates, share of PCR tests that were positive, were also an important predictor in our study. Since positivity rate data is only available once community testing is available,

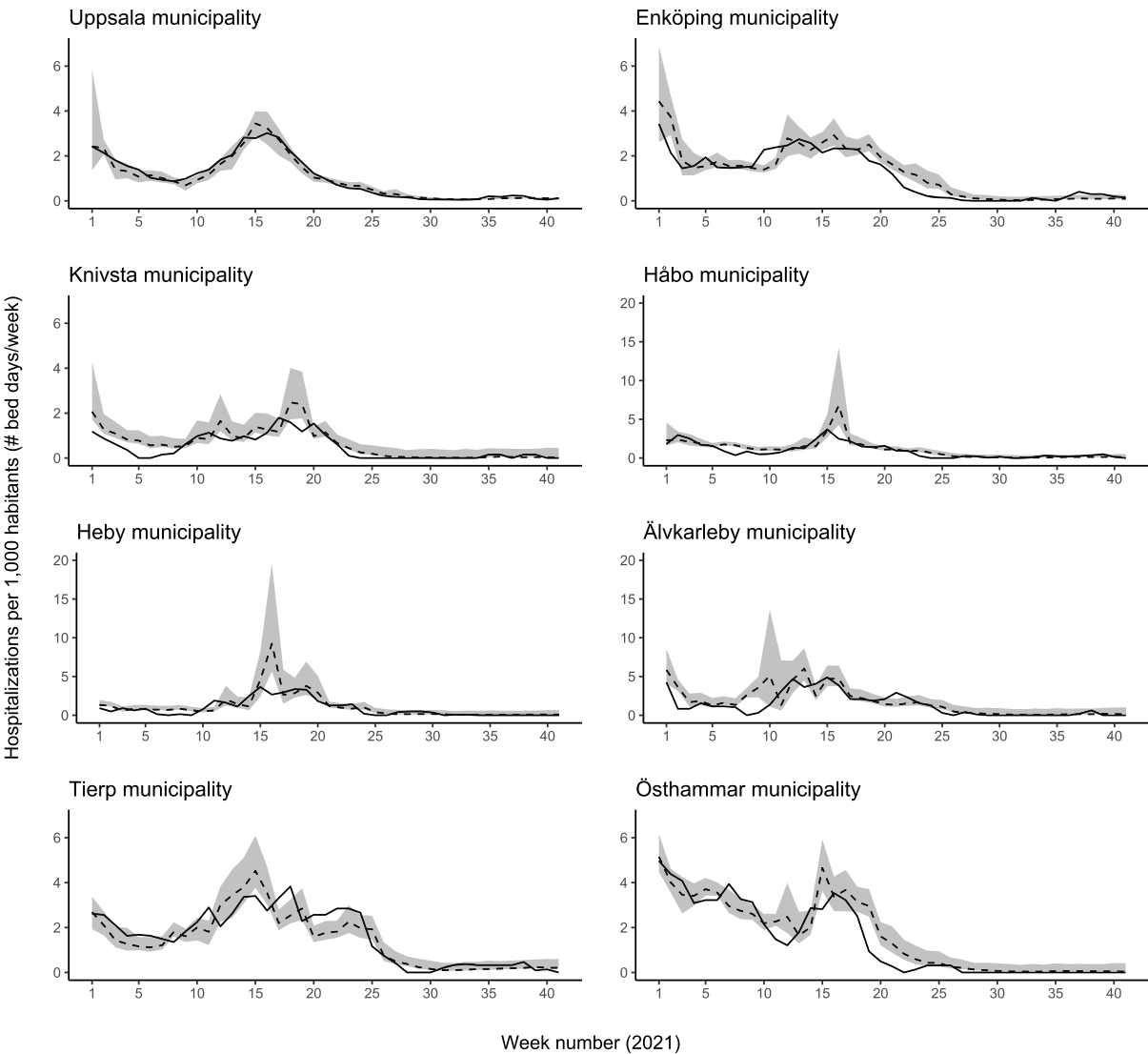


Fig. 2. Time series of predicted (dashed line) vs. observed (filled line) number of hospitalizations per 1000 inhabitants. The gray area indicates the prediction interval (95%) estimated using bootstrapping with 5000 samples.

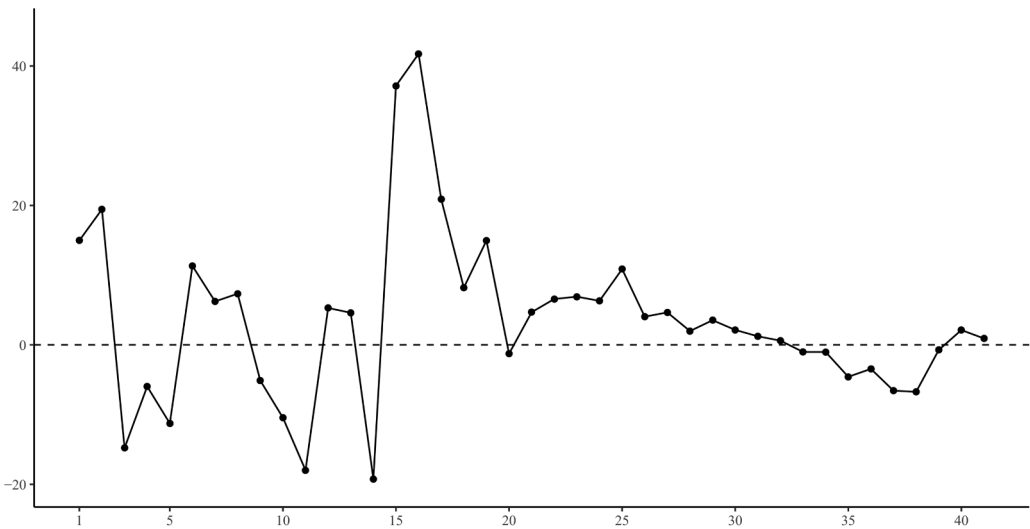


Fig. 3. Difference between the predicted and actual daily average number of hospital beds occupied by COVID-19 patients in Uppsala County. Values above zero indicate overprediction, values below zero indicate underprediction.

Table 4

Model performance comparison. Higher RMSE indicates poorer performance, thus higher variable importance for the variable removed.

| Excluded variable | Mean RMSE |
|---|-----------|
| No variables excluded (full model) | 16.84 |
| Test positivity (t-1) | 22.67 |
| Number of COVID-19 RT-PCR tests (t-1) | 20.82 |
| Hospitalizations (t-1) | 17.53 |
| Vaccinations 50+ (t-1) | 19.91 |
| Number of COVID-19-related calls to 112 (t-1) | 20.48 |
| Number of COVID-19-related calls to 1177 (t-1) | 36.75 |
| Municipality * nr of tests interaction (t-1) | 21.88 |
| Municipality * hospitalizations interaction (t-1) | 23.09 |

these results highlight the critical need for testing during a pandemic. This includes the development of tests as soon as possible after a pandemic outbreak, good availability of tests for all inhabitants, and proper registration of test results. The high collinearity between positivity rates and calls to healthcare hotline 1177 might suggest that, if tests are not yet available, data from a healthcare hotline can provide valuable information for predicting hospitalizations. However, the potential confounding effect of calls related to positive self-tests, coupled with the unavailability of hotline data prior to the introduction of testing, prevents us from isolating the individual predictive contribution of the healthcare hotline.

We noticed that municipalities in the rural areas, e.g. Östhammar and Älvkarleby, which are located on the borders of Uppsala County and far away from the hospitals in Uppsala and Enköping, have a higher number of hospitalizations per capita than the municipalities closer to a hospital (Table 2). These results suggest that people in municipalities further away from the hospital, may be admitted at an earlier stage of disease progression (i.e. with first doubts) due to the distance to the hospital being too great in case of a life-threatening emergency. Furthermore, demography could also play a role, as the mean age of inhabitants in Östhammar and Älvkarleby is higher than the mean age of inhabitants in Uppsala or Enköping (Table 2). The interaction effect between municipality and one-week lagged hospitalizations was not significantly different from zero in Uppsala municipality compared to baseline Enköping, which both have a hospital. The interaction effects of the other municipalities compared to baseline Enköping are all significantly positive throughout the entire time series (Fig. S3, Supplementary

Materials).

We also observed significant differences in the interaction effects between municipality and number of tests (Fig. S2, Supplementary Materials). Although most of the municipalities are not significantly different from baseline Enköping, there is a significant difference between Uppsala municipality and the others. A potential reason may be the difference in baseline testing rates, which is highest in Uppsala (Table 3). The unique population characteristics of Uppsala, as detailed in Table 2, also distinguish it from other municipalities. Furthermore, Uppsala municipality is characterized by a large number of students and academic personnel. Differences in population characteristics could potentially influence both the willingness to undergo testing and the accessibility of tests (Kennedy et al., 2023).

We note several limitations of this study, which may provide insights for future efforts to predict the spread of viral respiratory infections. One challenge we faced during the ongoing data collection was a reporting delay of the vaccination data for certain municipalities, which may have caused a large variability in the increase of vaccinated individuals in certain weeks. Although these issues were resolved retrospectively, they may cause issues for near real-time predictions.

While the model accounted for a large part of the spatio-temporal variability in hospitalizations, more data is needed to account for all variability. Other variables, like environmental temperature, UV light, seasonality, and temporal variability in travel restrictions and social distancing may influence infection spread (Liu et al., 2021; Merow and Urban, 2020; Nichols et al., 2021), but this variability should already be accounted for in the test positivity variable, which shows the effect of these factors before it is reflected in number of hospitalizations. It would have been interesting to include a variable for viral interference between respiratory viruses and its temporal variability (Piret and Boivin, 2022) which could potentially affect hospitalizations independently from test positivity. However, we did not have access to this data. When modeling larger amount of municipalities or a larger amount of smaller spatial units, it would also be useful to consider adding spatio-temporal random effects in the model to account for any remaining spatio-temporal correlation in the residuals (van Zoest et al., 2022).

Our model formulation, and the interpretation of the model coefficients, is under the assumption that the variables are uncorrelated. However, this assumption is likely violated for some variables within the model. As mentioned in the Results section, we observed multicollinearity issues between the ‘number of COVID-19 RT-PCR tests’ and

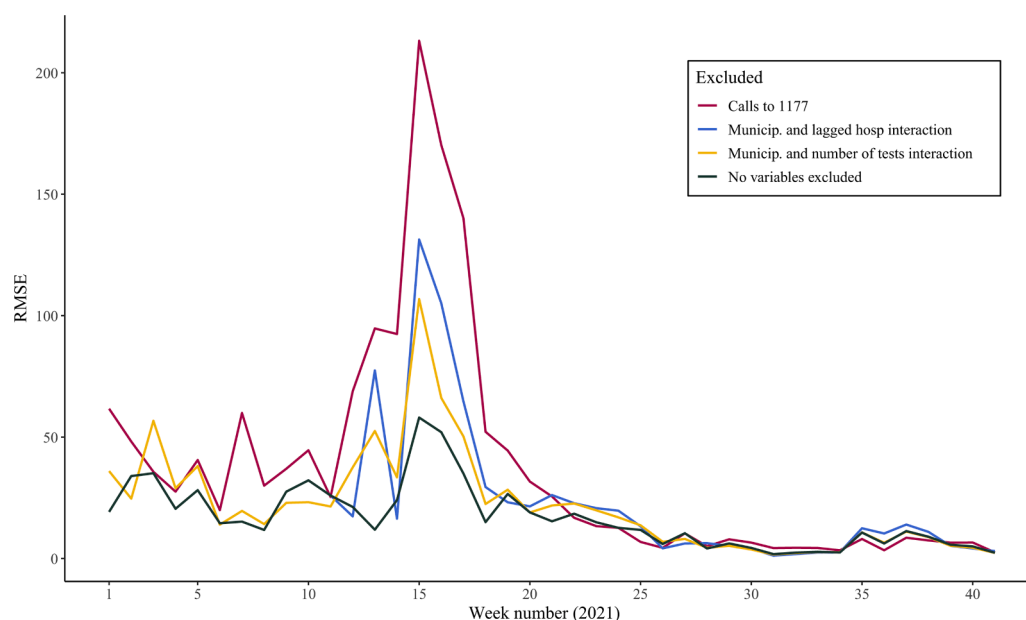


Fig. 4. RMSE time series of the weekly prediction performance of the full model (no variables excluded) vs. the model with one variable excluded.

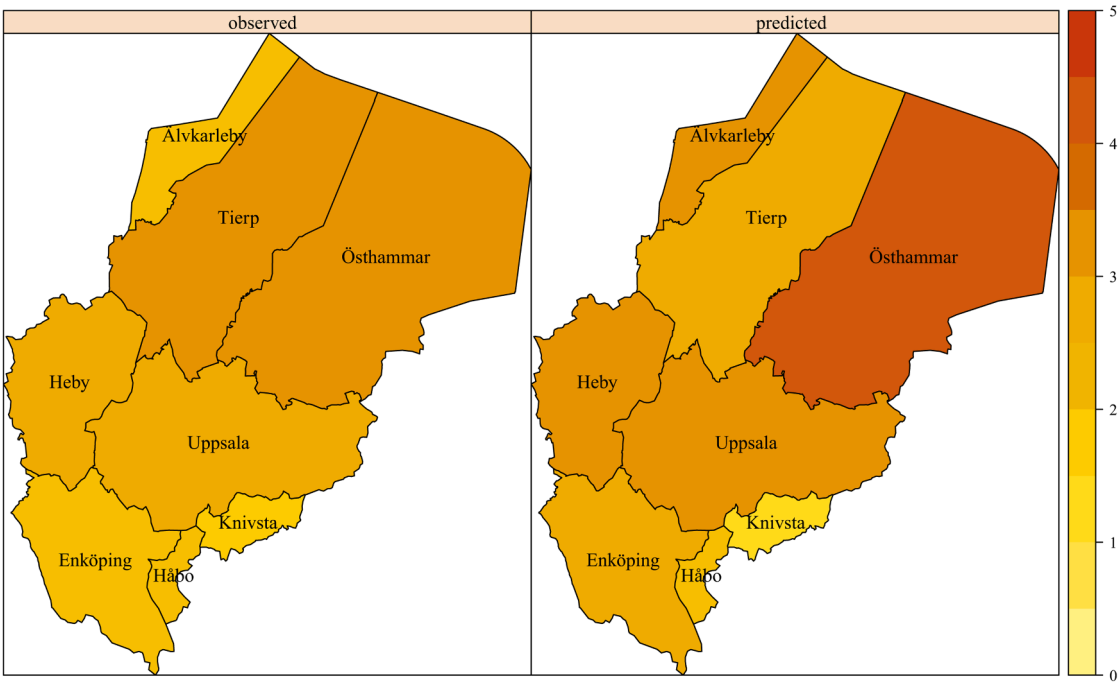


Fig. 5. Spatial variability in observed and predicted number of hospitalizations per 1000 inhabitants during the peak of the third wave of the COVID-19 pandemic (week 17, 2021).

‘calls to 1177 health care advice hotline’ variables. This multicollinearity is not surprising, as both indicate the number of people concerned about symptoms in an early stage of a potential infection. However, removing one of the variables would have a large negative impact on prediction performance, and we therefore decided to keep both variables in the model despite multicollinearity.

The inclusion of the municipality variable reduces the transferability of the trained to other locations, i.e. the model always needs to be retrained to estimate coefficients for “unseen” municipalities. However, excluding this variable strongly reduced the model’s predictive power. Besides that, there were big differences in baseline hospitalizations per municipality. Despite the lack of transferability of the trained model, this study suggests which variables are important to collect for training a model on a location of interest.

The methods used in this paper could be explored for predicting hospitalizations for seasonal respiratory viruses, such as influenza and Respiratory Syncytial Virus (RSV). Several predictors in our model, such as hospitalizations and ambulance calls, are routinely collected by public health authorities and their usefulness could be explored for these purposes.

5. Conclusion

In this paper, we propose a regression-based spatio-temporal prediction model specifically designed for one-week-ahead prediction of COVID-19 hospitalizations. While our modelled is tailored to COVID-19, its underlying principles and methods are applicable to other respiratory diseases. This is due to the fact that hospitalizations for such diseases can often be modelled using a negative binomial distribution to account for overdispersion in count data. The model’s flexibility allows for the inclusion of various spatial, temporal and spatio-temporal covariates, and spatio-temporal interaction effects. The relevance of these covariates may, of course, vary depending on the specific viral respiratory diseases and geographical region under study. However, the approach remains the same: integrating these covariates into a negative binomial regression model. A unique aspect of our work is the evaluation of the temporal variability in the importance of different variables. This is crucial

as the significance of certain variables can change over time, especially during a pandemic. In the early stages of a pandemic, before tests and vaccines become widely available, it is vital to leverage data from alternative sources, such as healthcare hotlines and emergency call data. These sources can provide valuable insights for predicting the spread of the virus. In this manuscript, we provide a comprehensive toolbox for prediction modeling, variable importance analysis, and performance evaluation, thereby enhancing the existing body of literature on this topic.

Ethical declaration

All parts of the study were conducted in accordance with the Declaration of Helsinki by the World Medical Association, as revised in 2013. The study was approved by the Ethical Review Board in Sweden (Etikprövningsmyndigheten, application 2020–04210 and 2021–01915), who waived the need for informed consent since only information on aggregate group level was extracted.

CRediT authorship contribution statement

Vera van Zoest: Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Karl Lindberg:** Data curation, Formal analysis, Methodology, Writing – review & editing. **Georgios Varotsis:** Conceptualization, Data curation, Writing – review & editing. **Frank Badu Osei:** Methodology, Writing – review & editing. **Tove Fall:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

We are grateful for support with accessing the data from the 112 service in Uppsala with the assistance of Douglas Spangler and Hans Blomberg. The study was partly funded by VINNOVA (2020-03173). Tove Fall was supported by grants from Swedish Research Council (2019-01471), the Swedish Heart-Lung Foundation (2019-0505), and the European Research Council (ERC-2018-STG 801965).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.sste.2024.100636](https://doi.org/10.1016/j.sste.2024.100636).

References

- Chelo, D., Nkwelle, I.M., Nguefack, F., Awa, H.D.M., Enyama, D., Nguefack, S., Ndombo, P.O.K., 2021. Decrease in hospitalizations and increase in deaths during the Covid-19 epidemic in a pediatric hospital, yaounde-cameroon and prediction for the coming months. *Fetal. Pediatr. Pathol.* 40 (1), 18–31. <https://doi.org/10.1080/15513815.2020.1831664>.
- Feikin, D.R., Higdon, M.M., Abu-Raddad, L.J., Andrews, N., Araos, R., Goldberg, Y., Patel, M.K., 2022. Duration of effectiveness of vaccines against SARS-CoV-2 infection and COVID-19 disease: results of a systematic review and meta-regression. *Lancet* 399 (10328), 924–944. [https://doi.org/10.1016/S0140-6736\(22\)00152-0](https://doi.org/10.1016/S0140-6736(22)00152-0).
- Galani, A., Aalizadeh, R., Kostakis, M., Markou, A., Alygizakis, N., Lytras, T., Thomaidis, N.S., 2022. SARS-CoV-2 wastewater surveillance data can predict hospitalizations and ICU admissions. *Sci. Total Environ.* 804 <https://doi.org/10.1016/j.scitotenv.2021.150151>.
- Gatto, A., Accarino, G., Aloisi, V., Immorlano, F., Donato, F., Aloisi, G., 2021. Limits of compartmental models and new opportunities for machine learning: a case study to forecast the second wave of COVID-19 hospitalizations in Lombardy, Italy. *Inform.-Basel* 8 (3). <https://doi.org/10.3390/informatics8030057>.
- Gerlee, P., Karlsson, J., Fritzell, I., Brezicka, T., Spreco, A., Timpka, T., Lundh, T., 2021. Predicting regional COVID-19 hospital admissions in Sweden using mobility data. *Sci. Rep.* 11 (1), 24171. <https://doi.org/10.1038/s41598-021-03499-y>.
- Hilbe, J.M., 2014. *Modeling Count Data*. Cambridge University Press.
- Kennedy, B., Varotsis, G., Hammar, U., Nguyen, D., Carrasquilla, G.D., van Zoest, V., Fall, T., 2023. Sociodemographic characteristics and COVID-19 testing rates: spatiotemporal patterns and impact of test accessibility in Sweden. *Eur. J. Public Health*. <https://doi.org/10.1093/eurpub/ckad209>.
- Liu, X., Huang, J., Li, C., Zhao, Y., Wang, D., Huang, Z., Yang, K., 2021. The role of seasonality in the spread of COVID-19 pandemic. *Environ. Res.* 195, 110874. <https://doi.org/10.1016/j.envres.2021.110874>.
- Merow, C., Urban, M.C., 2020. Seasonality and uncertainty in global COVID-19 growth rates. *Proc. Natl. Acad. Sci.* 117 (44), 27456–27464. <https://doi.org/10.1073/pnas.2008590117>.
- Nichols, G.L., Gillingham, E.L., Macintyre, H.L., Vardoulakis, S., Hajat, S., Sarran, C.E., Phalkey, R., 2021. Coronavirus seasonality, respiratory infections and weather. *BMC Infect. Dis.* 21 (1), 1101. <https://doi.org/10.1186/s12879-021-06785-2>.
- Patricio, A., Costa, R.S., Henriques, R., 2021. Predictability of COVID-19 hospitalizations, intensive care unit admissions, and respiratory assistance in Portugal: longitudinal cohort study. *J. Med. Internet Res.* 23 (4) <https://doi.org/10.2196/26075>.
- Perone, G., 2021. Comparison of ARIMA, ETS, NNAR, TBATS and hybrid models to forecast the second wave of COVID-19 hospitalizations in Italy. *Eur. J. Health Econ.* <https://doi.org/10.1007/s10198-021-01347-4>.
- Piret, J., Boivin, G., 2022. Viral interference between respiratory viruses. *Emerg. Infect. Dis.* 28 (2), 273–281. <https://doi.org/10.3201/eid2802.211727>.
- Reno, C., Lenzi, J., Navarra, A., Barelli, E., Gori, D., Lanza, A., Fantini, M.P., 2020. Forecasting COVID-19-associated hospitalizations under different levels of social distancing in lombardy and Emilia-Romagna, Northern Italy: results from an extended SEIR compartmental model. *J. Clin. Med.* 9 (5) <https://doi.org/10.3390/jcm9051492>.
- Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A., Firth, D., 2022. Package 'MASS' (Version 7.3-58.1). CRAN.
- SCB, 2020. Statistical Database - Environment. Statistics Sweden (SCB). Retrieved from. https://www.statistikdatabasen.scb.se/pxweb/en/ssd/START_MI/.
- Simpson, C.R., Robertson, C., Vasileiou, E., Moore, E., McCowan, C., Agrawal, U., Sheikh, A., 2020. Temporal trends and forecasting of COVID-19 hospitalisations and deaths in Scotland using a national real-time patient-level data platform: a statistical modelling study. *Lancet Digit. Health* 3 (8), E517–E525. [https://doi.org/10.1016/S2589-7500\(21\)00105-9](https://doi.org/10.1016/S2589-7500(21)00105-9).
- Spreco, A., Jöud, A., Eriksson, O., Soltesz, K., Källström, R., Dahlström, Ö., Timpka, T., 2022. Nowcasting (Short-term forecasting) of COVID-19 hospitalizations using syndromic healthcare data, Sweden, 2020. *Emerging Infect. Dis.* 28 (3), 564. <https://doi.org/10.3201/eid2803.210267>.
- Turk, P.J., Tran, T.P., Rose, G.A., McWilliams, A., 2021. A predictive internet-based model for COVID-19 hospitalization census. *Sci. Rep.* 11 (1), 5106. <https://doi.org/10.1038/s41598-021-84091-2>.
- van Zoest, V., Varotsis, G., Menzel, U., Wigren, A., Kennedy, B., Martinell, M., Fall, T., 2022. Spatio-temporal predictions of COVID-19 test positivity in Uppsala County, Sweden: a comparative approach. *Sci. Rep.* 12 (1), 15176. <https://doi.org/10.1038/s41598-022-19155-y>.
- Wesner, J.S., Van Peursem, D., Flores, J.D., Lio, Y., Wesner, C.A., 2021. Forecasting hospitalizations due to COVID-19 in South Dakota, USA. *J. Healthc. Inform. Res.* 5 (2), 218–229. <https://doi.org/10.1007/s41666-021-00094-8>.
- Wollenstein-Betech, S., Cassandras, C.G., Paschalidis, I.C., 2020. Personalized predictive models for symptomatic COVID-19 patients using basic preconditions: hospitalizations, mortality, and the need for an ICU or ventilator. *Int. J. Med. Inform.* 142 <https://doi.org/10.1016/j.ijmedinf.2020.104258>.