Predictive Modelling of Respiratory Syncytial Virus burden on Hospital Services

2022

Note: This document is a mini version of my complete dissertation submitted to University of Manchester

Aim

The purpose of this study is to analyze laboratory-confirmed RSV admissions for a 5-year cohort from all sites under the Manchester Foundation Trust, including Manchester Royal Infirmary, Trafford General Hospital, Wythenshawe Hospital, and North Manchester General Hospital, and forecast RSV arrivals at A&E and admissions to acute care using time-series forecasting.

The primary objective is to develop a clinical predictive model that could be applied in practice and aid hospitals in winter planning by identifying and gaining insight into the future trends of RSV-associated hospitalizations. Using the original RSV time-series and COVID-19 adjusted RSV time-series, the secondary objective is to assess the impact on the future forecast.

Abstract

Introduction. During 2020–2021, COVID-19 mitigating efforts disrupted seasonal RSV outbreaks. A rise in respiratory syncytial virus (RSV) hospitalizations was recorded when COVID-19 restrictions were eased in several countries, outpacing yearly patterns prior to the pandemic. Changes in RSV epidemiology may have unanticipated impacts on healthcare systems and communities throughout the world, increasing the burden brought on by the pandemic and the need for more resources. Making clinical and public health decisions requires an understanding of the timing and severity of re-emerging RSV outbreaks. It is postulated that by using the aggregated RSV admissions data over the last five years, it is possible to develop a time-series prediction model that can forecast RSV admissions in UK hospitals on a monthly basis and identify emerging patterns in RSV hospitalizations.

Methods. The burden in hospitals in the United Kingdom (UK) was estimated using time-series modelling of monthly time-series of aggregated RSV admissions data collected from all hospitals within the NHS Manchester Foundation Trust from June 2017 until June 2022. This burden was measured in terms of RSV admissions, age-group, frequency of patients who would need critical care, and frequency of emergency cases related to RSV. To determine which time-series analysis model has the best accuracy in in predicting RSV case trend several sub-goals need to realized:

- Collect RSV data and transform them into reasonable formats that can depict the case trend.
- Select relevant time-analysis models and model parameters to model the case trend.
- Evaluate the performance of different models by comparing them to the each other. Decide the model with the best prediction accuracy.

Results. Forecasting was done using the original RSV time-series in the first stage. The best fit models chosen after the model evaluation process projected monthly from July 2022 to January 2023. According to the forecast, there would be an average of 48.50 RSV-associated hospital admissions, 47.16 admitted patients would not need critical care, 40.77 emergency admissions, and 23.74 admissions for children aged 1 to 5 years.

Forecasting was done using the COVID-19 corrected RSV time-series in the second stage. According to the prediction, there would be an average of 79.99 RSV-related hospital admissions, 76.97 admitted patients won't need critical care, 45.35 admissions will be of the "Emergency" classification, and 36.84 admissions would fall into the "1 to 5 years" age group.

After doing univariate time-series forecasting using the original RSV time-series, predictions obtained for all the variables show greater peaks during the summer and smaller peaks during the winter. Using COVID-19 corrected RSV time-series, the forecast for all the variables reveals lower peak during the summer season and greater peak during the winter season.

Conclusions. This study shows that time-series models can be utilized to make predictions with high accuracy. According to the findings, hospitals should be on the alert for RSV outbreaks in the upcoming seasons, and improved surveillance is advised for managing hospital capacity. These models can contribute to lessening the burden on hospital services, even though external validation is required to evaluate the model's employability.

Analytic Approach

The primary outcome of interest is forecasting RSV A&E arrival and admissions into acute care. The prediction was done by performing univariate time series forecasting on variables of interest from every dataset by using appropriate time-series models. RSV displays strong time-series seasonality. However, the COVID pandemic added significant complexity in predicting upcoming trends. The secondary outcome of interest is to understand the trend and pattern in the forecast after doing the COVID-19 adjustment in the time-series.

A two-step analytical process was employed to put the outcomes of interest into practise. Forecasting was done with the original COVID-19 influenced time-series in the first stage. The COVID-19-affected portion of the time-series was replaced in the second stage with the predicted values for that time period.

Results

'RSV spells by critical care' dataset

Time series forecasting for 'Grand.Total' variable:

Stage 1: Without COVID-19 replacement

The best time-series model after model evaluation was SARIMA with hyperparameters (p=1, d=0, q=1, P=1, D=0, Q=0, S=12), which had the lowest SMAPE score of 27.97%.

Year	Point Forecast	Lower 80	Higher 80	Lower 95	Higher 95
Jul 2022	78.60916	39.2812335	117.93708	18.46229	138.7560
Aug 2022	55.43968	0.1272053	110.75215	-29.15345	140.0328
Sep 2022	36.95753	-21.53354	95.44861	-52.49685	126.4119
Oct 2022	40.01180	-19.2140760	99.23768	-50.56636	130.5900
Nov 2022	45.09023	-14.3100033	104.49047	-45.75459	135.9351
Dec 2022	46.80850	-12.6333535	106.25035	-44.09997	137.7170
Jan 2023	36.62657	-22.8252277	96.07837	-54.29711	127.5503

Figure 5: Stage-1 model forecast for 'Grand.Total' with 80% CI

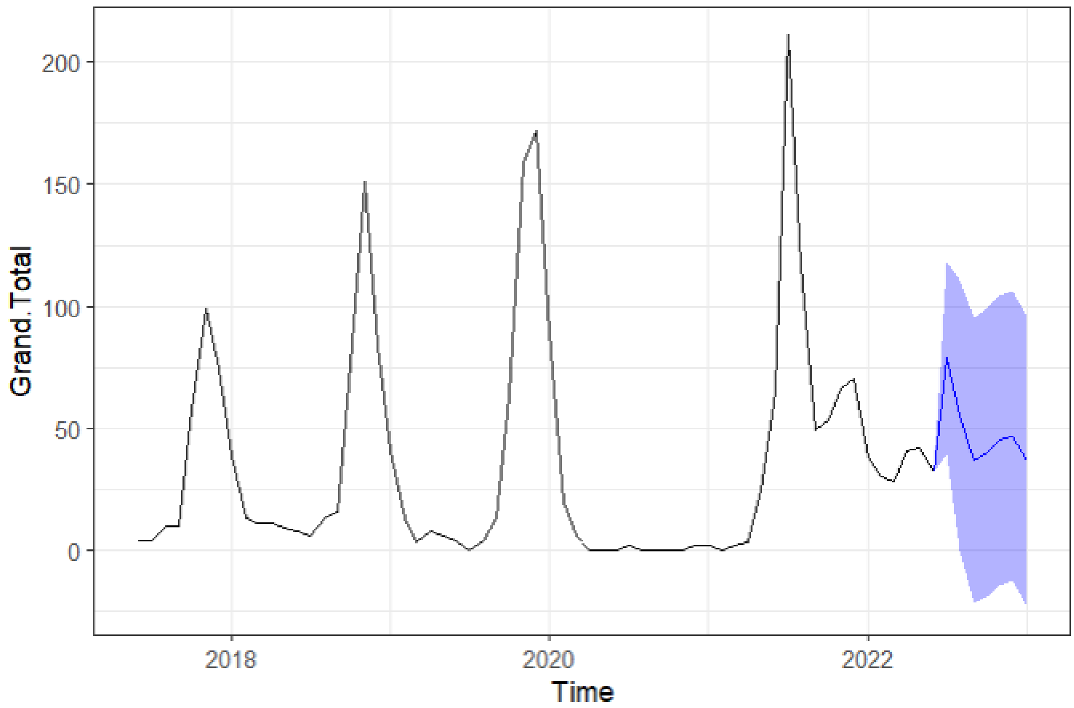

Table 1: Stage-1 forecast values for 'Grand.Total' with CI

The SARIMA model was used to forecast the total number of RSV cases for the following seven months, as shown by the blue line and confidence intervals in Figure 5. (Table-1). According to the time-series, the model forecasts peaks in RSV cases, with a higher peak occurring in the middle of the year and a lower peak occurring before the end of the year.

Stage 2: With COVID-19 replacement

The best time-series model identified was Neural Network with Hyperparameters (p=1, P=1, one hidden layer with 2 neurons, and lambda = 1), which had the lowest SMAPE score (40.52%). The values predicted by this model were utilized to replace the data that COVID-19 had affected.

With a SMAPE score of 71.9%, STL was the best-performing time-series model after the model evaluation process utilizing the COVID-19 adjusted time-series.

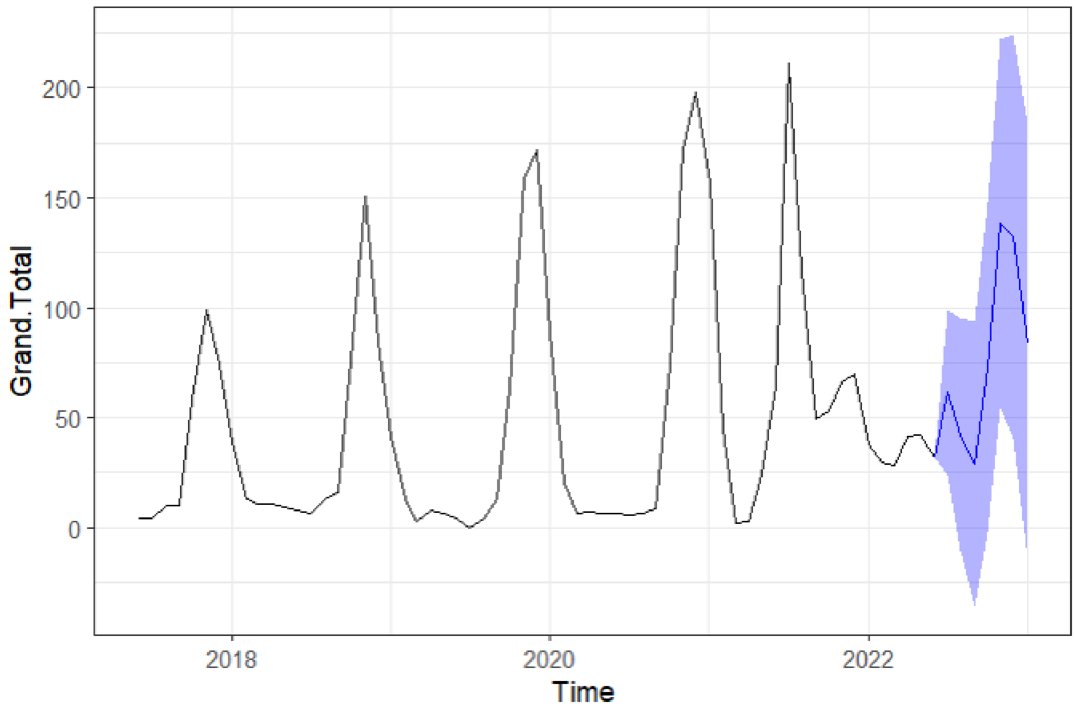

Figure 6: Stage-2 model forecast for 'Grand.Total' with 80% CI

Year	Point Forecast	Lower 80	Higher 80	Lower 95	Higher 95
Jul 2022	61.58416	24.114616	99.05371	4.279434	118.8889
Aug 2022	42.31293	-10.674365	95.30022	-38.724144	123.3500
Sep 2022	29.12882	-35.766013	94.02366	-70.119263	128.3769
Oct 2022	72.77608	-2.157398	147.70955	-41.824784	187.3769
Nov 2022	138.29740	54.519642	222.07515	10.170378	266.4244
Dec 2022	132.20051	40.426889	223.97414	-8.155134	272.5562
Jan 2023	83.67801	-15.448602	182.80462	-67.923059	235.2791

Table 2: Stage-2 forecast values for 'Grand.Total' with CI

For RSV admissions, the STL model forecasts peaks with the smaller peak occurring during the mid-term and the higher peak occurring before the end of the year, as shown by the time-series in (Figure-6).

Time series forecasting for 'False' variable:

Stage 1: Without COVID-19 replacement

The best time-series model after model evaluation process was SARIMA with hyperparameters (p=1, d=0, q=1, P=1, D=0, Q=0, S=12), which has the lowest SMAPE score of 27.88%.

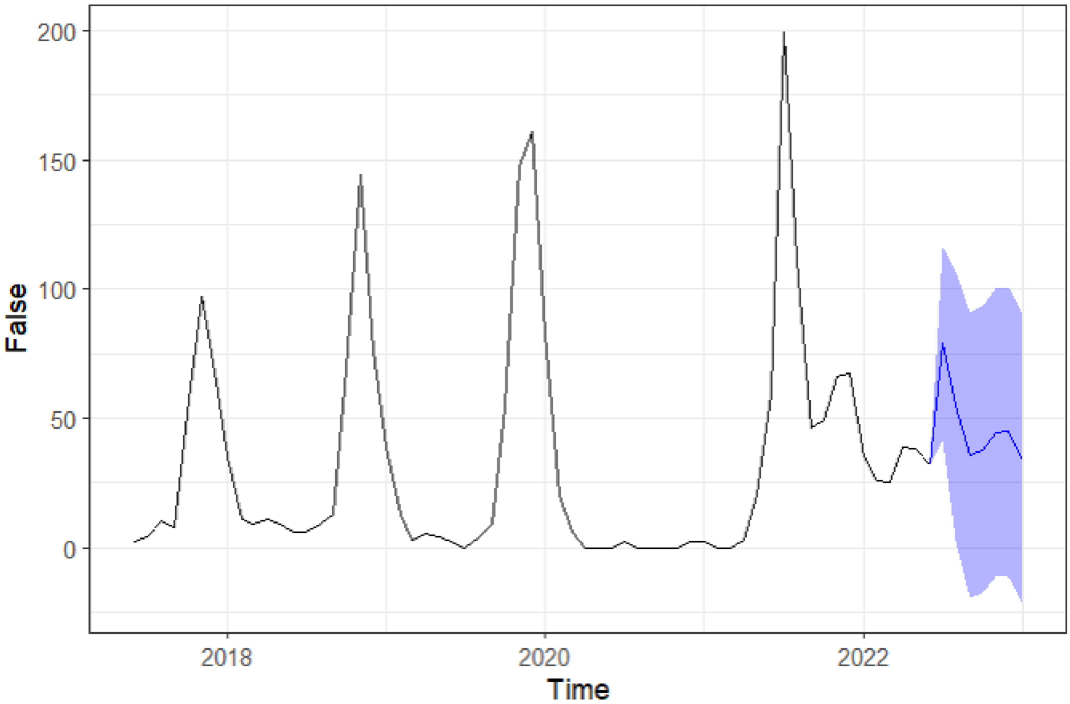

Figure 7: Stage-1 model forecast for 'False' with 80% CI

Year	Point Forecast	Lower 80	Higher 80	Lower 95	Higher 95
Jul 2022	79.00588	41.508656	116.50309	21.65883	136.3529
Aug 2022	54.84249	2.491762	107.19322	-25.22104	134.9060
Sep 2022	35.52367	-19.629667	90.67701	-48.82608	119.8734
Oct 2022	37.79797	-17.968607	93.56455	-47.48965	123.0856
Nov 2022	44.14307	-11.761022	100.04716	-41.35486	129.6410
Dec 2022	44.76560	-11.169488	100.70069	-40.77974	130.3109
Jan 2023	34.08422	-21.857868	90.02631	-51.47182	119.6403

Table 3: Stage-1 forecast values for 'False' with CI

A similar pattern can be seen in the time-series for the "False" and "Grand.Total" variables. According to the time-series (Figure-7), the SARIMA model forecasts peaks in RSV cases that do not require critical care, with a higher peak occurring in the middle of the year and a lower peak occurring before the end of the year.

Stage 2: With COVID-19 replacement

The best time-series model identified was Neural Network with hyperparameters (p=1, P=1, one hidden layer with 2 neurons and lambda = 1) as the best performing time-series model with the least SMAPE score of 34.08% and used the values forecasted by this model for replacing the data affected by COVID-19.

After the model evaluation process by using the COVID-19 adjusted time-series, STL was identified as the best performing time-series model with the least SMAPE score of 73.83%.

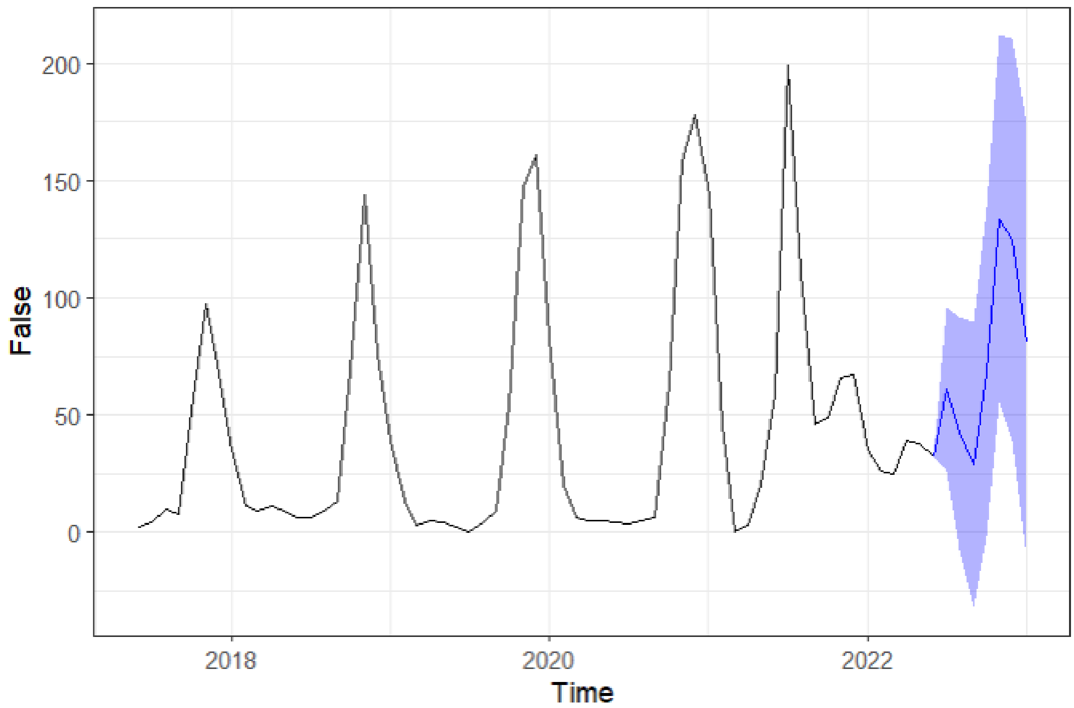

Figure 8: Stage-2 model forecast for 'False' with 80% CI

Year	Point Forecast	Lower 80	Higher 80	Lower 95	Higher 95
Jul 2022	60.92851	25.925495	95.93152	7.396019	114.4610
Aug 2022	42.39886	-7.100399	91.89812	-33.303725	118.1014
Sep 2022	28.81348	-31.809471	89.43644	-63.901325	121.5283
Oct 2022	67.96937	-2.031404	137.97015	-39.087576	175.0263
Nov 2022	133.54219	55.279332	211.80504	13.849479	253.2349
Dec 2022	124.70318	38.970801	210.43555	-6.413177	255.8195
Jan 2023	80.51439	-12.086939	173.11572	-61.107121	222.1359

Table 4: Stage-2 forecast values for 'False' with CI

As observed from the time-series (Figure-8), the STL model predicts peaks in RSV cases that will not require critical care with the lower peak during the mid-term and higher peak before the end of the year.

Forecasted values for 'True' variable:

Because the time-series for this variable does not show pattern, trend, or auto-correlation by its lags, the predicted values of the "False" variable is subtracted from the "Grand.Total" variable to determine the future values of the "True" variable.

Stage 1: Without COVID-19 replacement

Year	Forecasted Values
Jul 2022	-0.3967181
Aug 2022	0.5971894
Sep 2022	1.4338588
Oct 2022	2.2138307
Nov 2022	0.9471616
Dec 2022	2.0428953
Jan 2023	2.5423527

Table 5: Stage-1 forecast values for 'True'

Stage 2: With COVID-19 replacement

Year	Forecasted Values
Jul 2022	0.65565585
Aug 2022	-0.08593234
Sep 2022	0.31533807
Oct 2022	4.80670717
Nov 2022	4.75520961
Dec 2022	7.49733703
Jan 2023	3.16361707

Table 6: Stage-2 forecast values for 'True'

As observed (Table-5 & Table-6), for both stages the peaks in RSV admissions that will require critical care are towards the end of the year.

'RSV spells by admission type' dataset

Time series forecasting for 'Emergency' variable:

Stage 1: Without COVID-19 replacement

SARIMA with hyperparameters (p=1, d=0, q=1, P=1, D=0, Q=0, S=12) was identified as the best performing time-series model with the least SMAPE score of 31.08%.

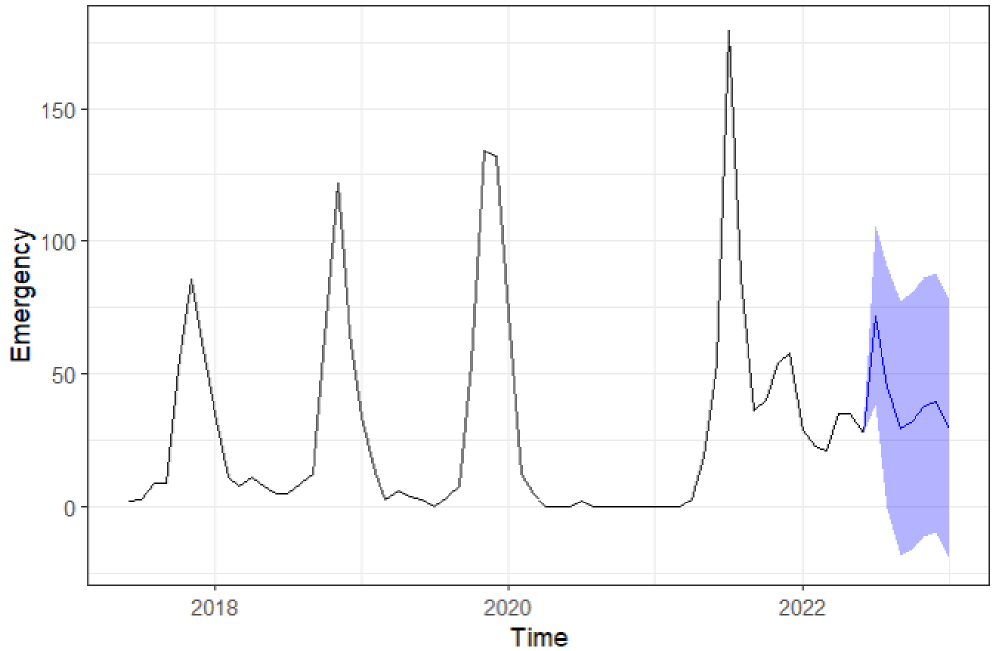

Figure 9: Stage-1 model forecast for 'Emergency' with 80% CI

Year	Point Forecast	Lower 80	Higher 80	Lower 95	Higher 95
Jul 2022	72.04580	38.4636847	105.62791	20.68639	123.4052
Aug 2022	45.01592	-0.7198364	90.75169	-24.93089	114.9627
Sep 2022	29.57993	-18.4538228	77.61368	-43.88135	103.0412
Oct 2022	32.23502	-16.2975237	80.76755	-41.98910	106.4591
Nov 2022	37.77741	-10.8659205	86.42074	-36.61615	112.1710
Dec 2022	39.47053	-9.1975342	88.13859	-34.96085	113.9019
Jan 2023	29.33409	-19.3395023	78.00768	-45.10574	103.7739

Table 7: Stage-1 forecast values for 'Emergency' with CI

From the time-series (Figure-9), the SARIMA model predicts peaks in RSV admissions of type 'Emergency' with higher peak during the mid-term and lower peak before the end of the year.

Stage 2: With COVID-19 replacement

Neural Network with hyperparameters (p=1, P=1, one hidden layer with 2 neurons and lambda = 1) was selected as the best performing time-series model with the least SMAPE score of 29.57% and used the values forecasted by this model for replacing the data effected by COVID-19.

After the model evaluation process by using the COVID-19 adjusted time-series, TBATS was selected as the best performing time-series model with the least SMAPE score of 79.14%.

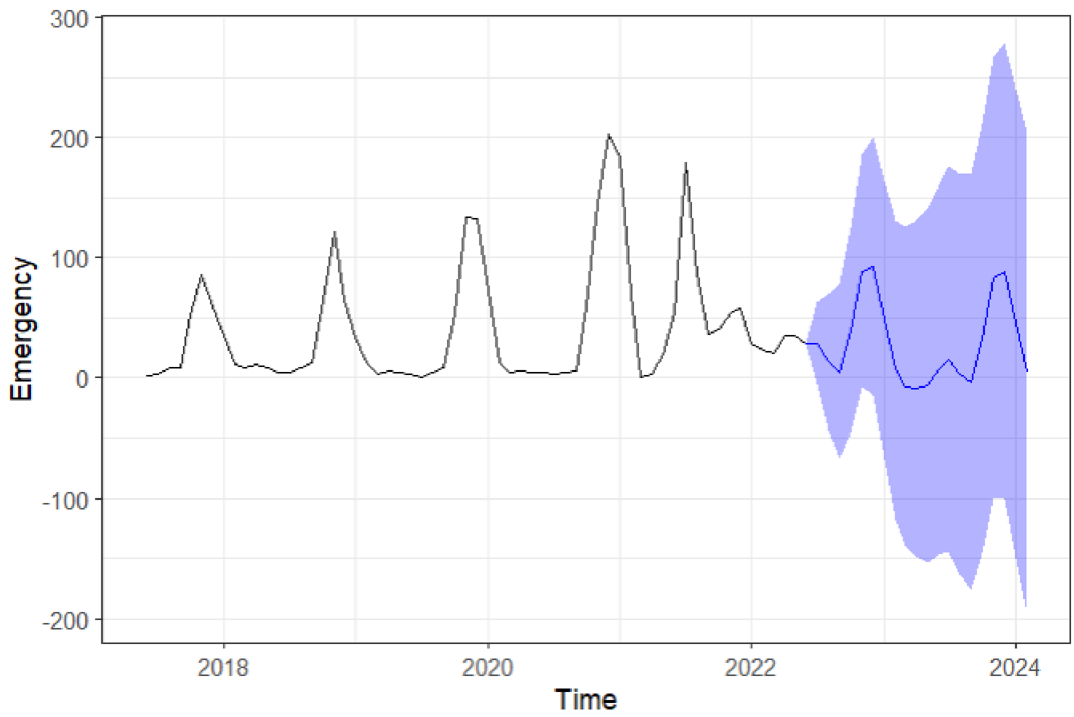

Figure 10: Stage-2 model forecast for 'Emergency' with 80% CI

Year	Point Forecast	Lower 80	Higher 80	Lower 95	Higher 95
Jul 2022	29.006729	-5.002020	63.01548	-23.00516	81.01862
Aug 2022	13.543703	-43.043501	70.13091	-72.99896	100.08636
Sep 2022	5.186503	-67.246330	77.61934	-105.58996	115.96296
Oct 2022	39.249744	-46.161987	124.66148	-91.37623	169.87572
Nov 2022	88.610391	-8.057242	185.27803	-59.23000	236.45078
Dec 2022	93.085262	-13.653665	199.82419	-70.15784	256.32836
Jan 2023	48.788957	-67.156716	164.73463	-128.53465	226.11256

Table 8: Stage-2 forecast values for 'Emergency' with CI

To better understand the trend represented by the blue line in Figure-10, total number of Emergency cases for twenty months is projected into the future. The table displays the forecast numbers with their confidence intervals for the following seven months (Table-8). According to the time-series, the TBATS model forecasts smaller peaks in the mid-term and greater peaks toward the end of the year for RSV admissions of type "Emergency."

'RSV spells by age' dataset

Time series forecasting for '1-5 Years' variable:

Stage 1: Without COVID-19 replacement

SARIMA with hyperparameters (p=2, d=0, q=0, P=1, D=0, Q=0, S=12) was selected as the best performing time-series model after the model evaluation process with the least SMAPE score of 27.19%.

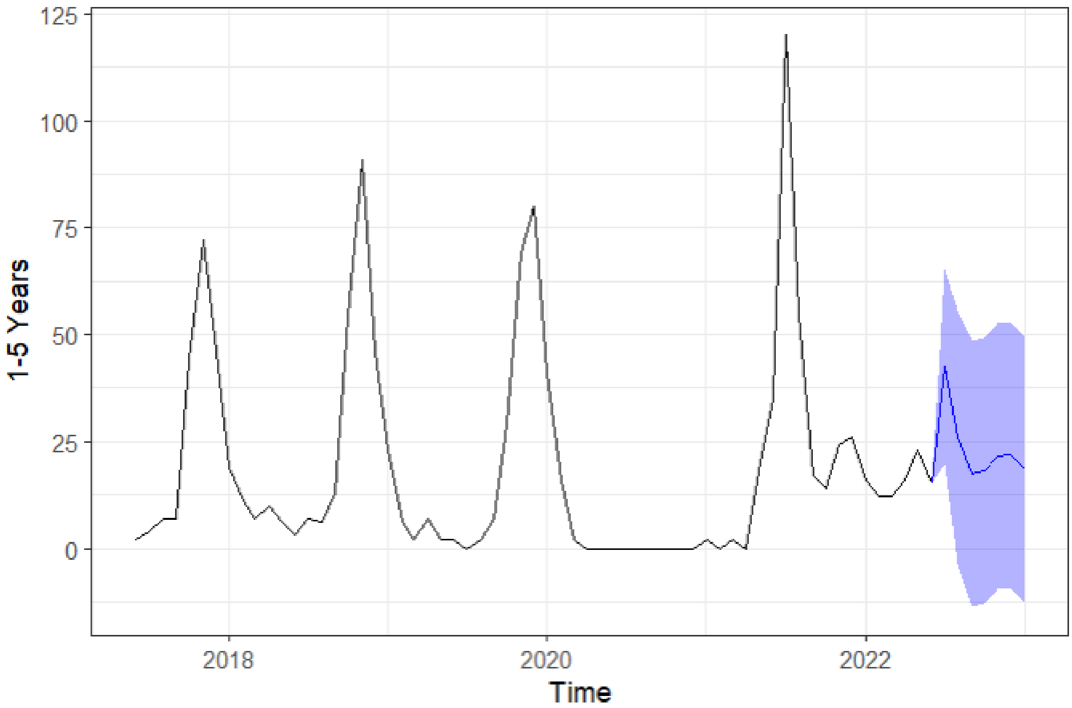

Figure 11: Stage-1 model forecast for '1-5 Years' with 80% CI

Year	Point Forecast	Lower 80	Higher 80	Lower 95	Higher 95
Jul 2022	42.67607	19.825276	65.52686	7.728799	77.62333
Aug 2022	26.12530	-3.606340	55.85694	-19.345319	71.59592
Sep 2022	17.43301	-13.597560	48.46358	-30.024151	64.89017
Oct 2022	18.08318	-12.989136	49.15550	-29.437828	65.60420
Nov 2022	21.57057	-9.531569	52.67270	-25.996045	69.13718
Dec 2022	21.94709	-9.197372	53.09154	-25.684252	69.57842
Jan 2023	18.42487	-12.733683	49.58342	-29.228023	66.07776

Table 9: Stage-1 forecast values for '1-5 Years' with CI

In the age group "1-5 Years," the SARIMA model predicts peaks in RSV admissions, with a higher peak during the mid-term and a lower peak before the end of the year, as shown by the time-series (Figure-11).

Stage 2: With COVID-19 replacement

The best time-series model with the lowest SMAPE score is the Neural Network with Hyperparameters (p=1, P=1, one hidden layer with 2 neurons, and lambda = 1), and the values predicted by this model were used to replace the data that COVID-19 affected.

The COVID-19 adjusted time-series used in the model evaluation process yielded STL as the top-performing time-series model with the lowest SMAPE score.

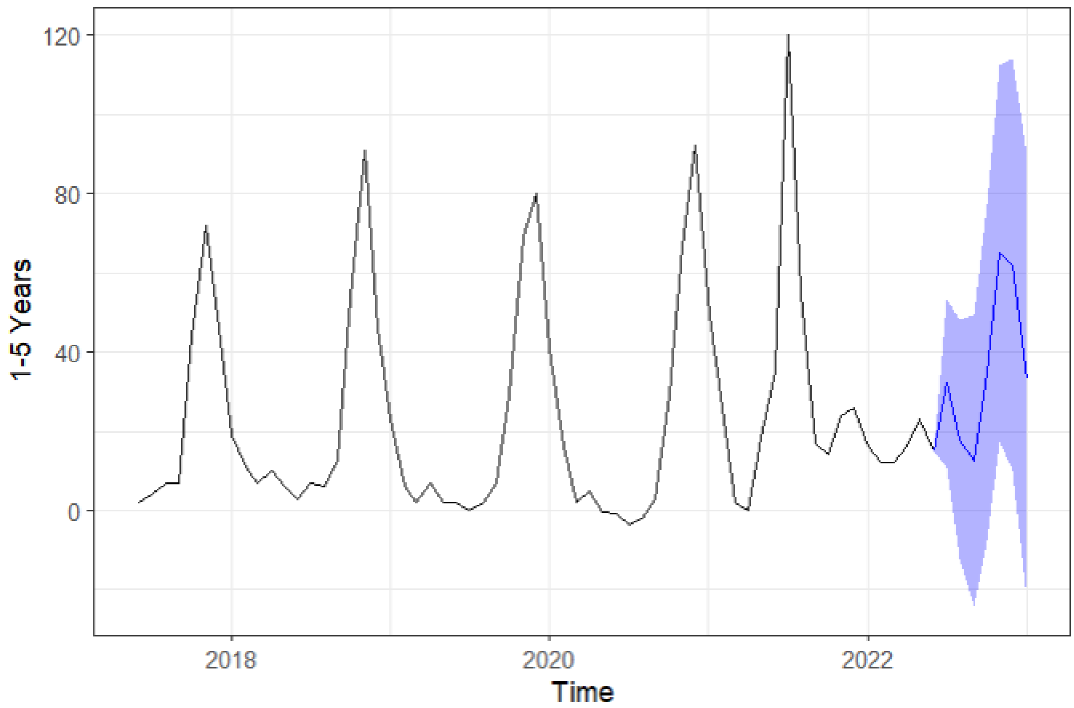

Figure 12: Stage-2 model forecast for '1-5 Years' with 80% CI

Year	Point Forecast	Lower 80	Higher 80	Lower 95	Higher 95
Jul 2022	32.24099	10.985584	53.49639	-0.2663474	64.74833
Aug 2022	17.97387	-12.084304	48.03205	-27.9961403	63.94388
Sep 2022	12.61192	-24.201062	49.42490	-43.6886766	68.91251
Oct 2022	34.93797	-7.569645	77.44558	-30.0718163	99.94775
Nov 2022	64.99670	17.471979	112.52141	-7.6860866	137.67948
Dec 2022	61.98722	9.926679	114.04777	-17.6325091	141.60696
Jan 2023	33.13384	-23.097844	89.36552	-52.8650978	119.13277

Table 10: Stage-2 forecast values for '1-5 Years' with CI

For the age range "1-5 Years," the STL model predicts peaks in RSV admissions, with a lesser peak around the mid-term and a bigger peak before the end of the year, as shown by the time-series (Figure-12).

‘RSV-Deprivation’ dataset

‘Index of Multiple Deprivation’ (IMD):

In order to understand patients from which decile of deprivation are admitted and stay in the hospital for longer with RSV, "IMD" variables have been grouped with respect to the "Length of Spell" variable. The term "IMD" denotes the general deprivation of the patient's area of origin.

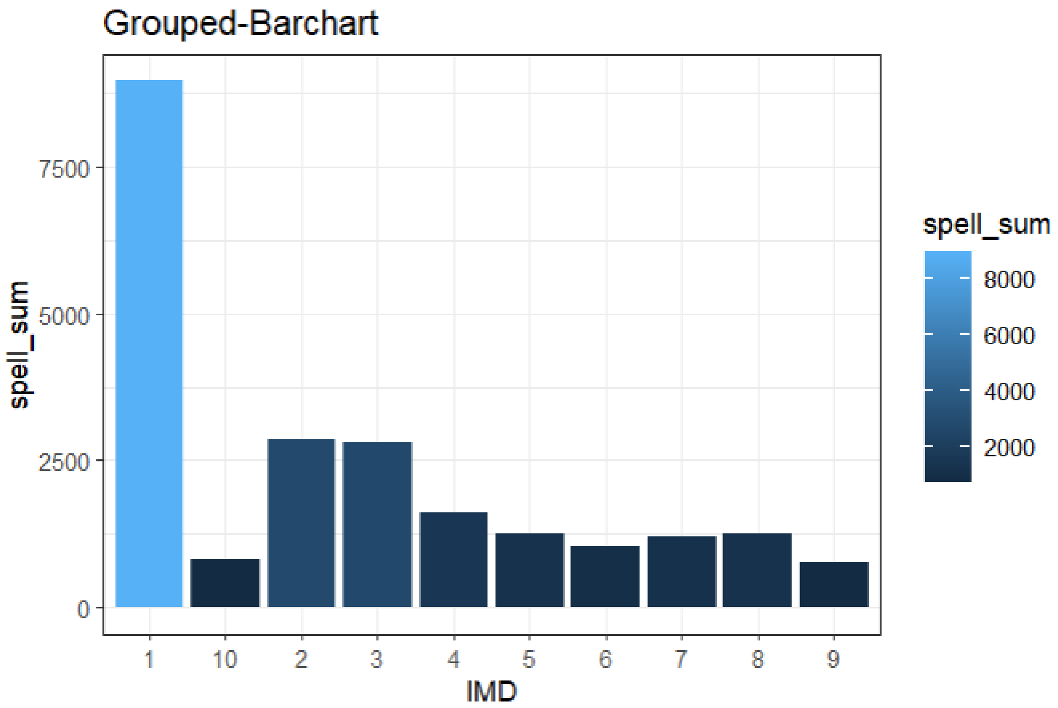

Figure 13: Total length of spell for each decile

IMD	spell_sum
1	8963
2	2878
3	2803
4	1616
5	1268
6	1043
7	1192
8	1251
9	769
10	825

Table 11: Tabular format for the total length of spell for each decile

Deciles of Deprivation	Interpretation
1	10% most deprived
2	20% most deprived
3	30% most deprived
4	40% most deprived
5	50% most deprived
6	50% least deprived
7	40% least deprived
8	30% least deprived
9	20% least deprived
10	10% least deprived

Table 12: Deciles of Deprivation with Interpretation

As observed (Figure-13), that for RSV, patients from decile 1, or the 10% most deprived area, stay in hospitals for a longer period of time than patients from deciles 9 and 10, which correspond to the 20% and 10% least deprived areas, respectively.

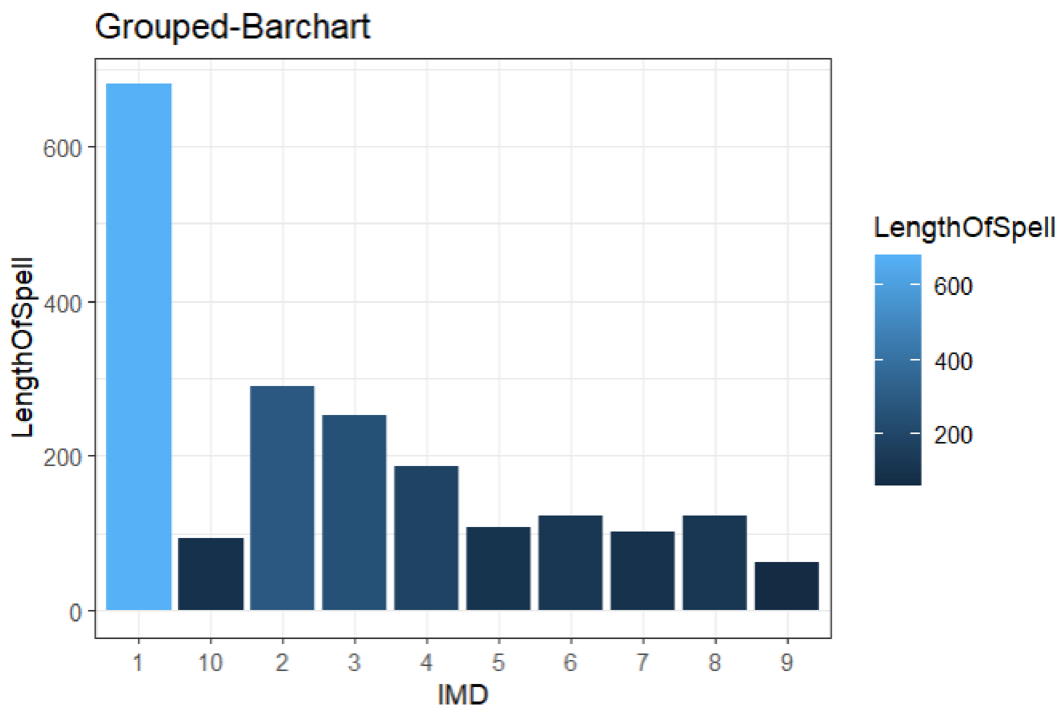

Figure 14: Total count for each decile

IMD	Total Count
1	680
2	289
3	252
4	186
5	108
6	122
7	101
8	123
9	63
10	94

Table 13: Total count for each decile in tabular format

The first decile category has the most admissions overall, whereas the ninth and tenth decile categories have the fewest admissions, according to the calculation of the total number of RSV admissions for each decile (Figure-15 & Table-13).

Discussion

From the first stage according to the forecast, there would be an average of 48.50 RSV-associated hospital admissions, 47.16 admitted patients would not need critical care, 40.77 emergency admissions, and 23.74 admissions for children aged 1 to 5 years.

From the second stage according to the predictions, there would be an average of 79.99 RSV-related hospital admissions, 76.97 admitted patients won't need critical care, 45.35 admissions will be of the "Emergency" classification, and 36.84 admissions would fall into the "1 to 5 years" age group.

As observed in the second stage, all of the best time-series models with SMAPE scores more than 70% are obtained after the model evaluation procedure utilizing the adjusted COVID-19 time-series. During the model evaluation process COVID-19 time-series was adjusted from (start= [2017,6] to (end= [2021,2]) to train the model, and evaluated the predicted results on the testing set (start= [2021,3] to (end= [2022,6]) (Figure-4). Since the training series follows a repetitive seasonal pattern and the seasonal pattern is distorted post lockdown, it is obvious that the forecast obtained after training the model using the training time-series will not match or resemble the testing series. As a result, a very high SMAPE score was achieved.

The whole time-series (start= [2017,6]; end= [2022,6]) with complex seasonality is employed when predicting seven months in the future. Time-series models like STL, ETS, and TBATS excel in extracting important data from time series with complicated seasonality and producing precise forecasts.

In both stages, a difference in the forecast pattern is noted. The models are projecting a greater peak in the middle of the year, or during the summer season, and a lower peak before the end of the year, or during the winter season, in the first stage. In the second stage, the models are projecting lower summer peak and greater winter peak, which is consistent with the actual seasonality of RSV cases in the UK.

Conclusions

The results of this two-stage time-series modelling analysis suggest that RSV outbreaks may resurface outside of the typical RSV season. The length of any further mitigation measures and the degree of viral importation from other areas may have a significant impact on the RSV outbreak. Improved surveillance is advised for managing hospital capacity, especially for emergency admissions and admissions for children aged 1 to 5 years, in order to handle emergency admissions and admissions for children with excellent management and to provide the best care for the patients. Hospitals must be aware of the possibility of RSV epidemics outside of the typical season. The activities and policies connected to prevention and treatment in public health could be guided by this information.

The models can make monthly forecasts and projections with emerging patterns and trends for RSV hospitalizations based on the type of time-series employed, even though external validation is necessary to assess the model's viability for application.