Probabilistic forecasting of daily COVID-19 admissions using machine learning

Bahman Rostami-Tabar[†]

Cardiff Business School, Cardiff University, 3 Colum Drive, Cardiff CF10 3EU, UK
[†]*Corresponding author. Email: rostami-tabarb@cardiff.ac.uk*

Siddharth Arora

Saïd Business School, University of Oxford, Park End Street, Oxford OX1 1HP, UK

Juan F. Rendon-Sanchez

Kingston University, Kingston Hill, Kingston upon Thames, Surrey KT2 7LB, UK

AND

Thanos E. Goltsos

Cardiff Business School, Cardiff University, 3 Colum Drive, Cardiff CF10 3EU, UK

Accurate forecasts of daily Coronavirus-2019 (COVID-19) admissions are critical for healthcare planners and decision-makers to better manage scarce resources during and around infection peaks. Numerous studies have focused on forecasting COVID-19 admissions at the national or global levels. Localized predictions are vital, as they allow for resource planning redistribution, but also scarce and harder to get right. Several possible indicators can be used to predict COVID-19 admissions. The inherent variability in the admissions necessitates the generation and evaluation of the forecast distribution of admissions, as opposed to producing only a point forecast. In this study, we propose a quantile regression forest (QRF) model for probabilistic forecasting of daily COVID-19 admissions for a local hospital trust (aggregation of 3 hospitals), up to 7 days ahead, using a multitude of different predictors. We evaluate point forecast accuracy as well as the accuracy of the forecast distribution using appropriate measures. We provide evidence that QRF outperforms univariate time series methods and other more sophisticated benchmarks. Our findings also show that lagged admissions, total positive cases, daily tests performed, and Google grocery and Apple driving are the most salient predictors. Finally, we highlight areas where further research is needed.

Keywords: Forecasting; COVID-19; admissions; human mobility; time series; machine learning; probabilistic forecasting.

1. Introduction

As a response to Coronavirus-2019 (COVID-19) pandemic, governments worldwide have enacted a series of public health and social measures aiming to restrict the spread of the virus (see WHO, 2021). Integral to the reasoning for the measures is the notion of protecting healthcare provision from reaching capacity, both for COVID-19 patients and other essential health services. There are many instances

where hospital capacity has been at least locally reached, resulting in redirection of patients (to other hospitals) and cancellations of essential operations (as doctors are moved to COVID-19 specific wards) (Bobashev *et al.*, 2020; Weissman *et al.*, 2020). Even when this is not the case, 'many countries face health workforce challenges, including shortages, maldistribution and misalignment between population health needs and health worker competencies' (World Health Organization, 2020, pp. 12).

To allocate resources intelligently then, hospitals and regional authorities need to have forward visibility of estimates of admissions (Gallivan & Utley, 2005). These forecasts need to be localized (region level, e.g. area of coverage by 1–3 hospitals) and short- to mid-term in terms of the time they cover. Localized rather than the aggregate (country level) daily forecasts allow for planning and reshuffling of resources between different wards and jurisdictions as needed.

This need for localized, daily forecasts, however, presents challenges in terms of modelling. Daily time series admissions are count data i.e. non-negative integer time series. As we move to finer granularities (e.g. from global/country to local/region, and from monthly to daily), and during periods when a virus is subdued, these count data can often become low volume. Crucially, given the inherent variability in the admissions time series, it is meaningful to focus on modelling the full forecast distribution of admissions for planning, rather than to focus only on a point forecast (in line with recent recommendations in the wider forecasting literature, see Goltsos *et al.*, 2021).

Recent peaks of infections of the Omega family of variants tell us that forecasting COVID-19 admissions remains relevant (BBC, 2022). It has also been pointed out that 'changes in climate and land use will lead to opportunities for viral sharing among previously geographically isolated species of wildlife' (Carlson *et al.*, 2022). This increased circulation of viruses might very well trigger different emergencies, perhaps similar to what COVID-19 has brought. Accurate long-term prediction of the number of COVID-19 patients needing admission is challenging, see Ioannidis *et al.* (2022). Prediction accuracy for COVID-19 hospital admissions at shorter horizons has been encouraging [see, for example, (Bekker *et al.*, 2023)]; however, the literature on short-term forecasting is still rather scarce, and our study aims to contribute to this area of research.

In this paper, we aim at forecasting daily COVID-19 admissions up to 7 days ahead by utilizing various predictors, all of which are publicly available. We examine whether these variables can explain the variability in daily COVID-19 admissions and use significant predictors to build a forecasting model. We propose a Quantile Regression Forest (QRF) model and show that it outperforms univariate (exponential smoothing and ARIMA families) and features enabled sophisticated benchmarks [linear regression (LR), ARIMA-X, Facebook Prophet]. We generate both point and probabilistic forecasts and evaluate model performance using the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) for point forecasting and the Ranked Probability Score (RPS) to assess probabilistic forecast accuracy.

The remainder of this article is structured as follows: In section 2, we review the literature and position our work within it; in section 3, we present various data sets used in this study, followed by discussing the modelling approach and benchmark methods in section 4. We then describe the experimental setup and performance evaluation metrics in section 5; in section 6, we present and discuss our results; in section 7, we summarize our findings and present ideas for future research.

2. Related research

Forecasting COVID-19-related variables has evolved rapidly with numerous mathematical, statistical and computational intelligence models proposed to forecast epidemic-related variables at global, country and regional levels.

The mathematical models have a long history (Kermack & McKendrick, 1927), whereby the modelling relies on dividing the populations into compartments such as Susceptible, Exposed, Infected, Removed and Dead. They are referred to as SEIRD models, and are comprised of a series of differential equations (Fanelli & Piazza, 2020; Hamzah et al., 2020; Martelloni & Martelloni, 2020; Sardar et al., 2020). Statistical models include LR, ARIMA, exponential smoothing and trend and seasonal components (Benvenuto et al., 2020; Feroze, 2020; Rostami-Tabar & Rendon-Sanchez, 2021). Computational intelligence models include neural networks, deep-learning, clustering techniques and agent-based models (Balcan et al., 2009; Fong et al., 2020; Hazarika & Gupta, 2020; Moftakhar et al., 2020; The Gleam Project, 2020; Nikolopoulos et al., 2021; Qian et al., 2021). There are also hybrid models, such as the ones proposed by Zaplotnik et al. (2020), that combine social networks (to mimic the connections between people) and a virus transmission model in a framework of data assimilation.

The forecasts are usually aimed at supporting government decision-making and planning in hospital networks, although other key aspects such as supply chain stress are emerging (Ivanov, 2020; Nikolopoulos et al., 2021). Forecasting tools for local operations (1–3 hospitals) appear to be less common than those for region (4–10) or country levels (10+), despite their importance in the context of strained and changing resource planning and operation. At a local level, forecasts of admissions-related variables are produced following two general approaches. The first involves starting from the national level and downscaling to the region and city to finally reach the local level. The second involves starting at the local level, by supporting or guiding the models with the aid of national or regional estimates. The rest of this section summarizes relevant research on forecasting Covid-19 admissions-related variables, including admissions, during COVID-19. To grasp the context, we include approaches for higher levels above hospitals (Table 1 summarizes the overview of models).

Zhao et al. (2020) modelled admissions at the hospital level in Switzerland by use of geographical downscaling of projections from the regional level. The total occupancy of a hospital was divided in groups covering existing and COVID patients. For modelling COVID patients, the total number of ICU projected through a SEIR models was spatially disaggregated. Future patients requiring ICU were assigned geographically (by using latitude and longitude) to each administrative subdivision (canton) via stratified random sampling inside the canton with a weighting factor based on population density. Each case was assigned to the nearest hospital in travel time. The probability of attending a hospital was calculated through Monte Carlo simulations. The authors strike a balance between complexity and precision that is insightful: model at national and regional level with epidemiological models, and then disaggregate to model admissions and ICU usage at hospital level via geographic location and population density.

Deasy et al. (2020) forecast bed occupancy in England. COVID-19 diagnosed cases per commissioning regions are used to forecast the likely distributions of new COVID-19 infections. Subsequently, mortality rates and required percentage of ICU by region obtained from previous literature are used to estimate ICU admission and ICU mortality per case. They employed Monte Carlo simulation to build a link between the expected daily incidence distribution and the expected excess ICU bed occupancy due to COVID-19.

Qian et al. (2021) built Capacity Planning and Analysis System (CPAS), where projections are made first of admissions at the hospital level, then at the regional level and finally at the national level in the UK. Patient distribution is calculated first, followed by daily admissions based on trend forecaster (which uses an enhanced compartmental model). An agent-based model simulates outcomes for patients admitted, based on risk profiles produced by an individual risk predictor (fed with patients records) and then simulated outcomes are aggregated to the desired level. The compartmental model that drives the trend forecast is embedded in Bayesian hierarchical models and Gaussian processes to cope with the

Table 1   *Forecasting approaches for Covid-19 admissions-related variables*

Approach	Target variables	Freq. of data	Horizon	Scope	Reference
Compartmental model for rates of hospital and ICU admission	ICU and non-ICU beds required at the peak; time to peak	Weekly	53 weeks	Country (USA)	Moghadas *et al.* (2020)
Growth models and hierarchical forecasting	Number of ICU admissions	Daily	14 days	Country (Italy) and region	Gitto *et al.* (2021)
Linear and exponential models	Number of ICU admissions	Daily	24 days	Region (Lombardy, Italy)	Grasselli *et al.* (2020)
Ordinary least squares to extrapolate daily incidence	Expected daily incidence and expected ICU bed occupancy	Daily	14 days	Region (commissioning regions in England)	Deasy *et al.* (2020)
Compartmental model with hospital compartments applied to key cities to predict occupation of ICU	Hospital demand	Daily	Unspecified	Region (Mexico city metropolitan areas)	Capistran *et al.* (2020)
Compartmental models at national and regional level with geographical downscaling	ICU bed occupancy	Daily	4 days	Region and hospital (Switzerland)	Zhao *et al.* (2020)
Agent-based simulation with compartmental model and individual risk predictor to predict hospital admissions and ICU admissions	New hospital admissions, ICU demand	Daily	30 days	Country, region and hospital (UK)	Qian *et al.* (2021)
Cox regression model with demographics and data from regional patient database	Number of admissions and deaths	Daily	4 weeks	Region (Scotland)	Shah *et al.* (2021)

scarcity of observations. This is an example of a system driven by a high-level model (compartmental) that produces forecasts from the hospital to the national level. The utility of short-term forecasts of COVID-19 hospital activity has been demonstrated during the pandemic to inform public health policy (Funk *et al.*, 2020; Leclerc *et al.*, 2021).

Growth models (Harvey, 1984) and hierarchical forecasting methods (Nenova & May, 2016) have been used by Gitto *et al.* (2021) to predict the demand for ICU beds at the national and regional level in Italy during the most severe period of the outbreak. The hierarchical models were used to guarantee consistency between national and regional forecasts.

From these forecasting approaches, several observations can be made. It seems that models for admissions at the local level benefit from complex systems. However, the study of their accuracy is not systematic, probably because of their use in earlier stages of the pandemic and the complexity inherent in their training. In addition, the dependencies between different levels, from national to hospital, are likely to make them more challenging to maintain than simpler models. Amid the changing circumstances, with assumptions and applicability of models being constantly challenged, it is important to be able to directly forecast admissions at the local level (1–5 hospitals) without having to rely on complex hierarchies and associated data which may be hard to obtain. This is the focus of the present article.

3. Data

This study has been motivated by a real-life forecasting problem faced by data analysts in a Clinical Commissioning Group (CCG), managing three hospitals in the East Midlands region of England (local level).

The variable we aim to forecast is the total daily number of COVID-19 patients admitted to three hospitals of a region (covered by the CCG), we refer to this as daily COVID-19 admissions or simply admissions. To that end, this study utilizes various sources of daily time series data collected at the CCG level, including (i) daily COVID-19 admissions, (ii) COVID-19 confirmed cases based on age groups, (iii) calls to the UK COVID-19 hot-line (NHS111) based on age groups, (iv) Google and Apple human mobility, (v) COVID-19 tests conducted and (vi) percentage of the population vaccinated with the first and second doses. We share the collected data used in this article to improve reproducibility (Boylan, 2016).

The admission data and all predictors are collected for a region corresponding to a CCG in the East Midlands region of England between 18 March 2020 and 30 June 2021. Therefore, they all correspond to the same area/geographical location. We use the time series from 18 March 2020 up to the end of March 2021 as in-sample (training) and from the first of April 2021 until the end of June 2021 as out-of-sample (test). In this section, we briefly describe these data sets and their sources.

3.1 *Daily COVID-19 admissions*

The daily COVID-19 admissions data are obtained from the official UK government website[1] for data and insights on Coronavirus (COVID-19). This database contains the total number of daily COVID-19 admissions in three hospitals of a CCG in the East Midlands region of England. We have aggregated daily admissions across three hospitals to obtain the total daily COVID-19 admissions time series that is illustrated in Fig. 1. There are three peaks in the time series corresponding to three observed waves. Therefore, the series contains multiple short piecewise trends. The time series does not show any day-of-week seasonality and there is no difference in hospital admission between working days and weekends. However, there are significant autocorrelation coefficient, as shown in Fig. 2. This indicates that the number of admissions in the future is associated with its past values.

It should be noted that the number of admissions is very low for many days; in fact, there are many days without any COVID-19 patient admitted to the hospitals. This adds another level of complexity to forecast daily admissions at the local level, compared with national or global levels.
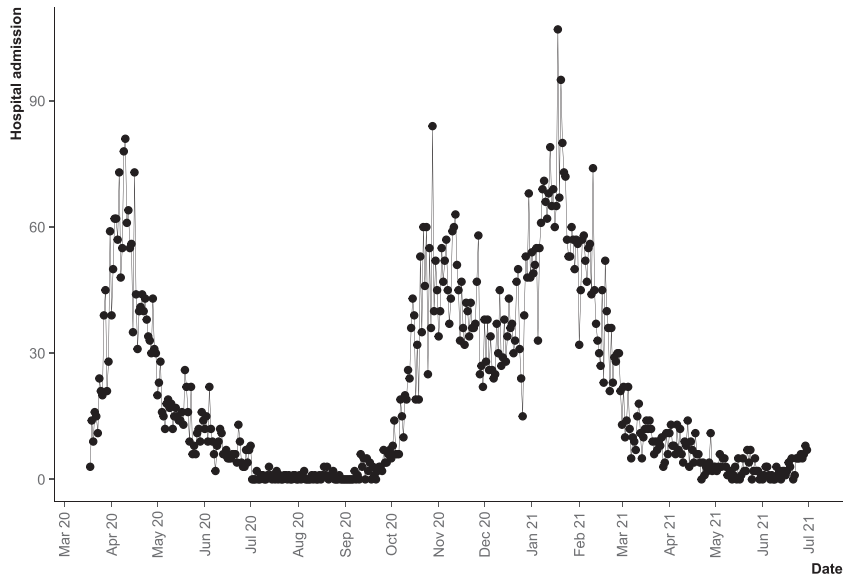
---

[1]https://coronavirus.data.gov.uk/details/healthcare

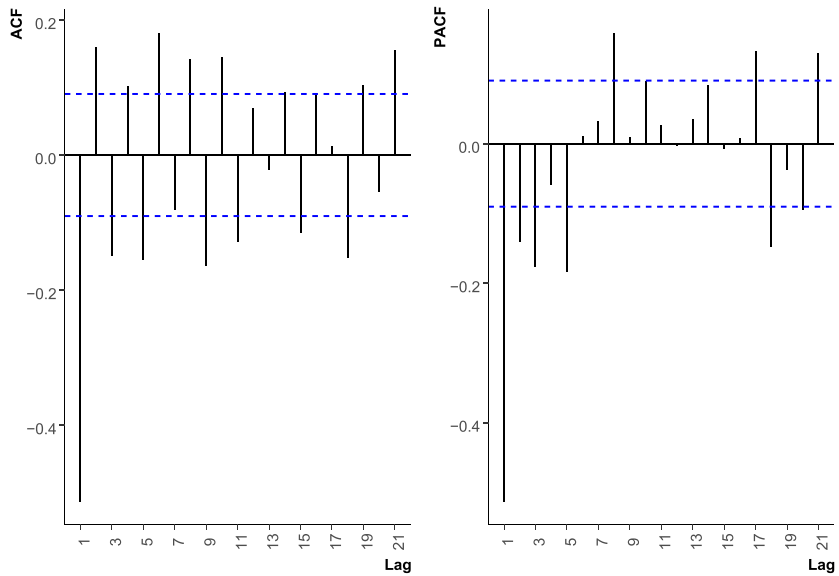Fɪɢ. 1. Daily time series of COVID-19 admissions.

Fɪɢ. 2. Autocorrelation of daily time series of COVID-19 admissions predictor.

3.2   *Potential predictors*

Various predictors are used to forecast future COVID-19 daily admissions, which are described below.

- Confirmed positive cases

    The number of daily COVID-19 confirmed cases is extracted from the Public Health UK[2]. Data are obtained for age grouped, and confirmed cases are calculated for age band over and under 70. We consider total positive cases and total cases per age group $age \geq 70$ and $age < 70$.

- Phone call data (NHS111 call)

    The number of daily phone calls received at the National Health Service 111 (NHS 111) corresponding to the CCG is substantial. Data are publicly available and extracted from NHS Digital[3]. Data are obtained for different age band, and they are aggregated to calls from over ($age \geq 70$) and under 70 ($age < 70$). The threshold of $age \geq 70$ for NHS111 call and positive cases is chosen based on the data granularity provided for age groups and also recommendations from the literature (Bonanad *et al*., 2020; Sattar *et al*., 2020)

- Human mobility

    We use publicly-available data on human mobility from Google and Apple. Both data sets contain aggregated and anonymized information on users' movement on their online platform. Google data show the percentage change in the length of time people spend in different types of locations (i.e. residential, retail, transit, grocery and workplace). Apple mobility data indicate the change in volume of people driving, walking or taking public transit in their communities.

    The mobility data provided by Google and Apple describe the change in people's behaviour after the implementation of social distancing measures with respect to a baseline day. The baseline day is the median value of each measure from the 5-week period 3 January—6 February 2020 for Google and 13 January 2020 for Apple. The provided data are normalized using pre-pandemic data to adjust for differences in population and mobility levels across the area considered in this study. Detailed definitions of human mobility measures are freely accessible for Google[4] and Apple[5].

- Daily tests performed.

    This data set contains the total number of daily tests conducted including (rapid) lateral flow device and polymerase chain reaction tests. It includes confirmed positive, negative or void COVID-19 virus test results. The data set has been downloaded from Public Health England Coronavirus website.[6]

- Percentage of population vaccinated

    We also use the time series of percentage of population from the region receiving the first and the second dose of COVID-19 vaccines. The data set is extracted from the UK government website[7].

---

[2]https://coronavirus.data.gov.uk/downloads/demographic/cases/specimen_date-latest.csv
[3]https://digital.nhs.uk/data-and-information/publications/statistical/mi-potential-covid-19-symptoms-reported-through-nhs-pathways-and-111-online/latest
[4]https://www.google.com/covid19/mobility/
[5]https://aws.amazon.com/marketplace/pp/prodview-ucdg4nudicxzi#offers
[6]https://coronavirus.data.gov.uk/details/testing
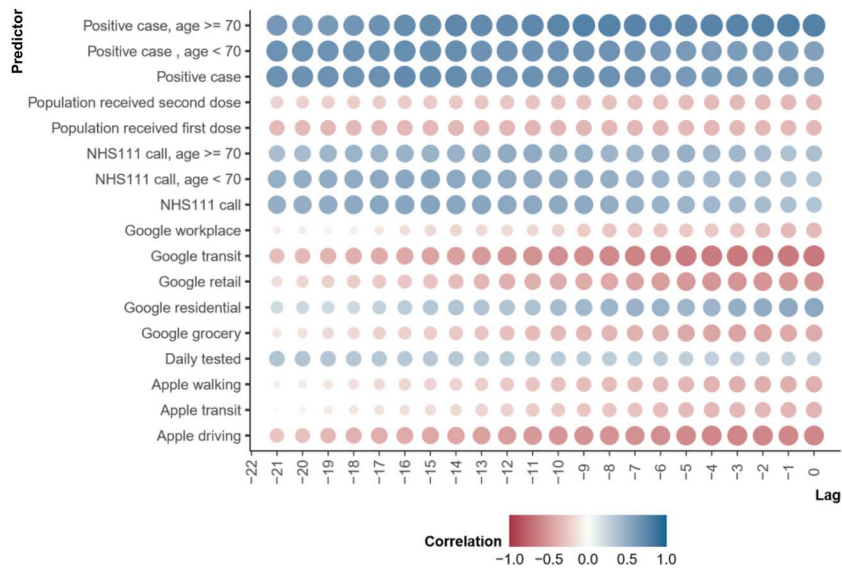[7]https://coronavirus.data.gov.uk/details/download

Fig. 3. Association between daily COVID-19 admissions and possible lagged predictors.

3.3   *Relationship between daily admissions and potential predictors*

Figure 3 indicates the correlation between daily COVID-19 admissions and potential predictors. The number of daily admissions on any particular day might be associated with various predictors described in section 3.2 and their past values/lags. The sample cross-correlation function (CCF) illustrated in Fig. 3 can be used to identify lags of predictors with significant autocorrelation coefficients. Each dot indicates the linear association between the time series of daily COVID-19 admissions and one lagged predictor that ranges from minus one to plus one. The Y-axis shows different predictors, and X-axis the number of lags. For example, consider *Google transit* and $lag = -1$, the CCF value shows the correlation between the COVID-19 admissions on day t and the Google transit percentage change on day $t - 1$. If $lag = 0$, we measure the correlation on day t.

Figure 4 presents the correlation between potential predictors at the same period (i.e. lag = 0). It is clear that some predictors are highly correlated and provide similar information. This should be considered when selecting predictors to forecast daily COVID-19 admissions to avoid multicollinearity.

In addition to investigating the linear association between predictors and admissions illustrated in Figs 2–4, we also explore the potential non-linear association between these variables. To that end, we looked at the scatterplot of each predictor versus admissions at the same period t (i.e. lag = 0), which is visualized in Fig. 5. We observe that the association between some predictors such as percentage of population vaccinated, total number of test conducted, Apple mobility-transit and hospital admission is not necessarily linear. We have also investigated the association between lagged predictors [e.g. $t - 1(lag1), t - 2(lag2),...,t - 21(lag21)$] and COVID-19 admissions at time t, and the shape of the plots and strength of relationship does not seem to be materially different and are omitted.

Based on the preliminary analysis of the collected data, we should consider a toolbox of forecasting models that can take into account the following features:
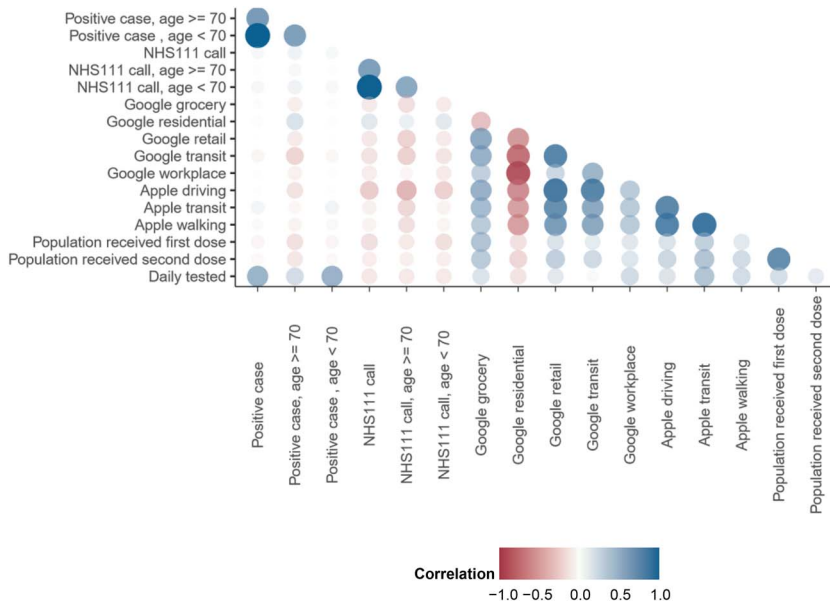
Fig. 4. Linear association between predictors at the same time period (lag = 0).

- The past values (lags) of daily COVID-19 admissions.

- Multiple predictors, their lags and their linear/non-linear association with daily COVID-19 admissions.

- Linear trends.

We note that the preliminary data analysis is based on the in-sample time series of admissions and all predictors. We consider several forecasting models that can capture the systematic information outlined above.

4. Forecasting models

In this section, we first present the constructed model to forecast daily admissions, followed by a brief description of benchmark methods.

4.1 *QRFs*

In this study, we employ a QRF, as proposed by Meinshausen (2006), to generate probabilistic forecasts of COVID-19 daily admissions. QRF are an extension of random forests that were proposed by Breiman (2001). In random forests, each tree is grown using a bagged version of the training data (i.e. selecting a random subset of the data with replacement), and a random subset of features are employed for split point selection at each node. Random forests comprise an ensemble of trees, such that the final prediction for a new test feature is the averaged response of all trees.

We use regression trees as they have been shown to be accurate in generating forecasts for a range of applications, including predicting the risk of emergency admission using electronic health records
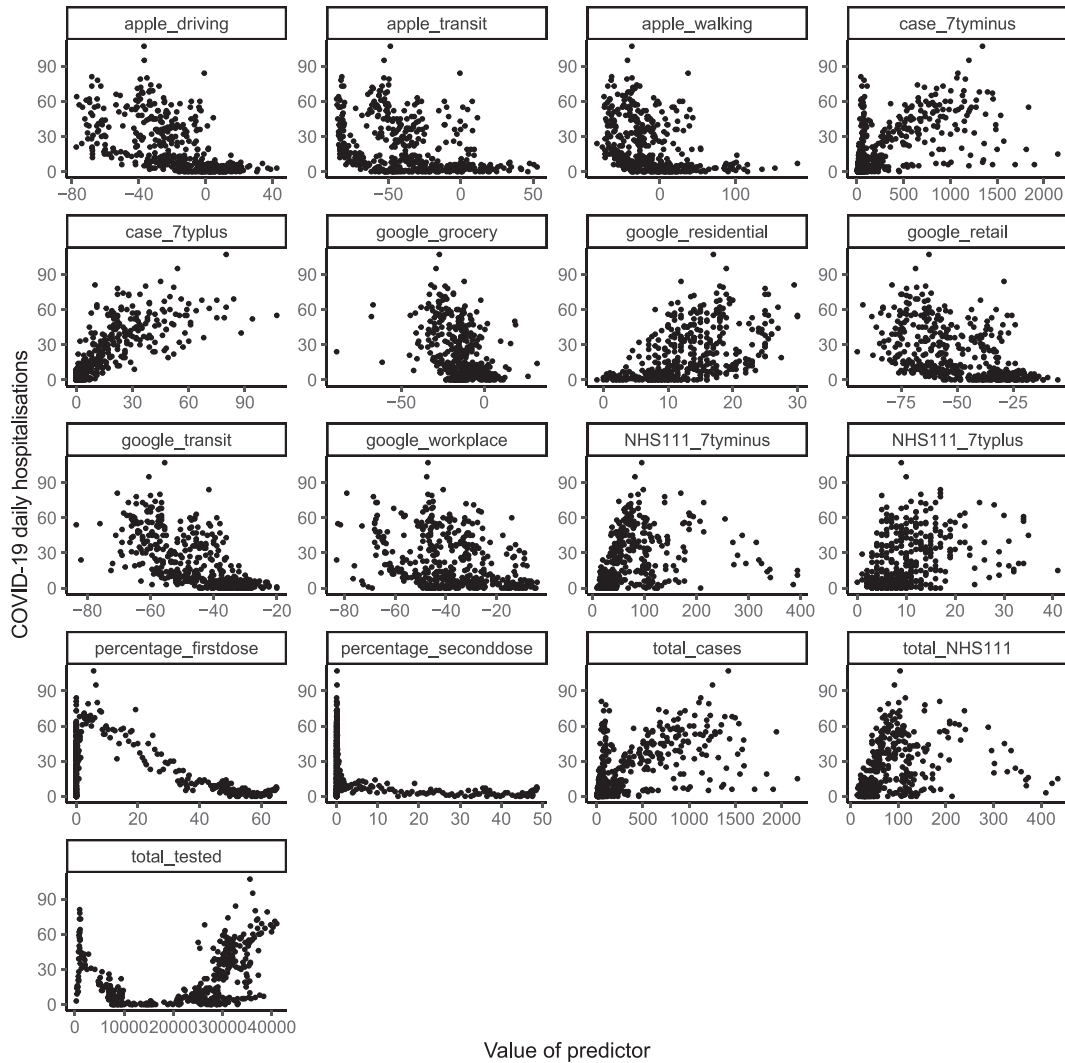
Fig. 5. Scatterplot of daily COVID-19 admissions vs. predictors at the same time period (lag = 0).

(Rahimian *et al*., 2018), forecasting patient waiting times in an emergency department (Arora *et al*., 2020), real-time prediction of patient readiness for discharge (Bishop *et al*., 2021) and predicting the risk of hospital admission (Hong *et al*., 2018). In the context of this work, an attractive feature of QRF is that the forecasts from this method are based solely on realizations of the historical data, which helps ensure that the predictions of admissions are positive and integer count values. This attribute of QRF is particularly useful as it helps avoid the issue of truncating the distribution values below zero and rounding of fractional prediction values. Although random forests have been commonly used for approximating the conditional mean of the target variable, QRF allow for estimating the full conditional distribution using a non-linear and non-parametric modelling framework.

Using QRF, our main objective is to model the total number of admissions for n days, $y = \{y_1, y_2, ..., y_n\}$ (target vector size $n \times 1$), as a function of a m features $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m$ (feature matrix size $n \times m$). Note that y_i denotes the number of admissions for day i, which are modelled using a m-dimensional feature vector $\mathbf{x}_i = \{x_{i1}, x_{i2}, ..., x_{im}\}$, where $\mathbf{x}_i \in R^m$. Specifically, we aim to train a model Ψ, with model parameters denoted θ, to estimate the conditional distribution of admissions for day i as, $\hat{F}(y_i|\mathbf{x}_i) = \Psi(\mathbf{x}_i, \theta)$. Using regression trees, for a given test feature vector say $\mathbf{x}_{test}$ that belongs to a leaf node $l(\mathbf{x}_{test}, \Psi)$, we calculate weight associated with a tree as:

$$wi\,(\mathbf{x}test, \Psi) = \frac{I\big(\mathbf{x}_i \in l\,(\mathbf{x}_{test}, \Psi)}{\big\{j : \mathbf{x}_j \in l\,(\mathbf{x}_{test}, \Psi)\big\}} \qquad (1)$$

Using the above weights, the forecast from the single tree Ψ is obtained as a weighted average of the target variable i.e. $\sum_{i=1}^{n} w_i\,(X_{test}, \psi) \times y_i$. To obtain forecasts using an ensemble of regression trees, the weights are averaged over all K trees ($\Psi_1, \Psi_2, ..., \Psi_K$) as $w_i(X_{test}) = 1/K \times \sum_{j=1}^{K} w_i(X_{test}, \psi_j))$. The conditional distribution for admissions is then calculated as $\hat{F}(y_i|X = X_{test}) = \sum_{i=1}^{n} w_i(X_{new})I(y_i < y)$.

For further details on QRF, please refer to Meinshausen (2006). In this study, we employed 25 features ($m = 25$) in the model to generate a one-step-ahead forecast for admissions ($t + 1$). We used lagged admission values up to a week before [t, (t-1),..., (t-6)], while for all other variables, we considered the most recent lagged value (t-1). Our feature matrix is rich, as it includes a range of different predictors including, lagged admissions, confirmed COVID-19 cases, NHS 111 calls, along with human mobility data obtained from Google and Apple, as discussed earlier in Section 3. The names and brief description of the 25 features used in the modelling are provided in Table 2.

To generate forecasts of the daily admissions for horizons (h) varying from 1 day to 7 days ahead, we construct a horizon-specific QRF model. For forecasting admissions, for say, the next day (horizon $h = 1$), we estimate a QRF model by minimizing the 1-day-ahead error on the out-of-bag (OOB) samples using only the training data. For a given horizon (or model), we estimate the QRF hyperparameters, namely the: (1) number of trees, (2) minimum leaf size and (3) number of features used for split-point selection. The hyperparameter tuning was based on the minimization of the mean absolute deviation calculated using the OOB observations. Each bagged tree makes use of around two-thirds of the randomly selected observations used for constructing the tree, whereas the remaining one-third of the observations not used for constructing the tree is referred to as the OOB data that were used only for hyperparameter tuning. Having identified the optimal hyperparameters, the QRF model was trained using the entire training data.

This modelling strategy results in a total of seven QRF models i.e. one model for each horizon. Note that the hyperparameters were tuned separately for each QRF model. Once estimated, the hyperparameter values were kept fixed and used to retrain each of the seven QRF models using the entire training data, and these models were then subsequently used to generate forecasts for the out-of-sample data. Note that the QRF model estimated by minimizing h-step-ahead error on the OOB observations was used to generate h-step-ahead forecasts for the complete out-of-sample period. An attractive feature of this modelling scheme is that it negates the need to generate forecasts of the independent variables. It can be argued that the shortcomings of this modelling approach, which relies on building horizon-specific models, are that the forecasts for different horizons are treated as being independent, and this approach is computationally expensive.

Figure 6 shows a simple example of probabilistic forecasts of 1-day-ahead COVID-19 admissions for the complete out-of-sample period, generated using QRF. In the figure, the actual admission counts

Table 2 *List of features used in QRF modeling. The features are presented according to their importance scores (most salient feature at the top) obtained using random forests (for 1-day-ahead forecasting)*

Predictor	Description
Total admission lag2	Hospital admission 2 days ago
Case 7tyminus lag1	Positive cases for the age group below 70 yesterday
Total cases lag1	Total positive cases yesterday
Total tested lag1	Total tests conducted yesterday
Case 7typlus lag1	Positive cases for the age group above 70 yesterday
Total admission lag3	Hospital admission 3 days ago
Total admission lag1	Hospital admission yesterday
Total admission lag4	Hospital admission 4 days ago
Total admission lag6	Hospital admission 6 days ago
Google grocery lag1	Google mobility index for grocery yesterday
Total admission lag7	Hospital admission 7 days ago
Apple driving lag1	Apple mobility index for driving yesterday
Total admission lag8	Hospital admission 8 days ago
Google workplace lag1	Google mobility index for workplace yesterday
Total admission lag5	Hospital admission 5 days ago
NHS111 7tyminus lag1	NHS111 calls received for the age group below 70 yesterday
Total NHS111 _lag1	Total NHS111 calls received yesterday
Google retail lag1	Google mobility index for retail yesterday
Apple transit lag1	Apple mobility index for transit yesterday
Percentage first dose lag1	Percentage of the population vaccinated with first dose yesterday
Google residential lag1	Google mobility index for residential yesterday
Google transit lag1	Google mobility index for transit yesterday
Apple walking lag1	Apple mobility index for walking yesterday
Percentage second dose lag1	Percentage of the population vaccinated with second dose yesterday
NHS111 7typlus lag1	NHS111 calls received for the age group above 70 yesterday

are denoted by black dots, the median of the probabilistic forecasts is issued as the point forecast, whereas the shaded regions correspond to 50 and 90% prediction intervals of the forecast distribution.

The probabilistic forecasting perspective attempts to quantify the uncertainty of the forecast admissions. Risky decisions tend to be concentrated on the extremes: although the unexpectedly high hospital admission creates crowding (Au *et al.*, 2009) and inefficiency, the unexpectedly low admissions may create a waste of resources. Probabilistic forecasts are thus tools to support decision-makers/planners not only with most likely outcomes but also with these low probability, high-cost situations (Rostami-Tabar & Ziel, 2022). Different hospital managements may have different attitudes to risk, depending upon the availability of resources, and thus it becomes imperative to consider the whole distribution while making decisions. The probabilistic forecasts can and should be seen as a core ingredient of any risk management practice within healthcare (Rostami-Tabar *et al.*, 2023).

4.2 *Benchmarks*

We use various benchmark methods to compare their forecast accuracy with employed QRF model. Exponential smoothing and ARIMA models are among the most widely used approaches in time series
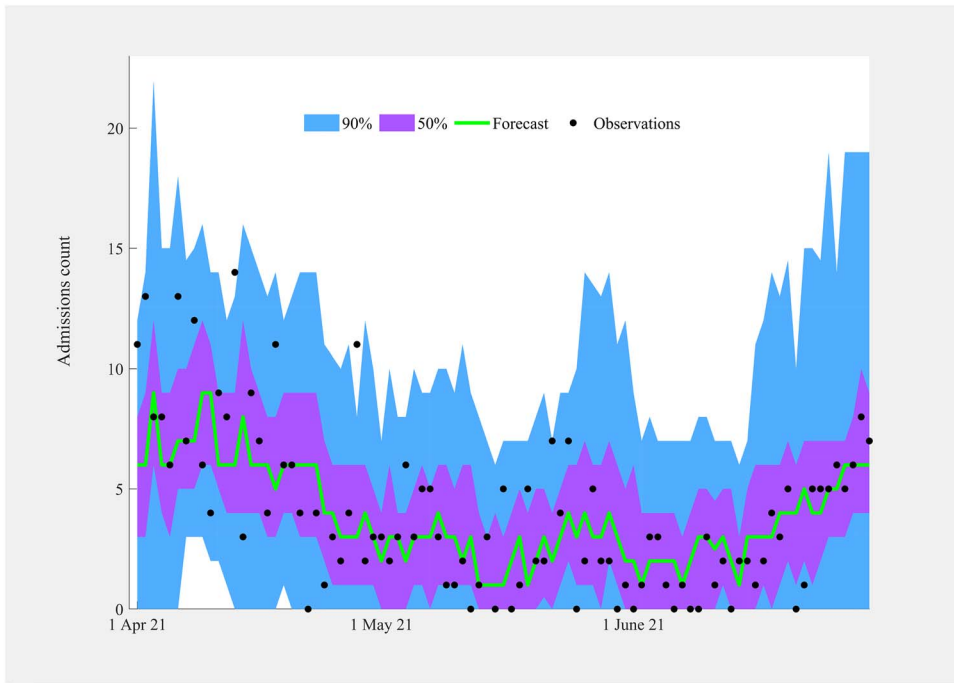
Fig. 6. Probabilistic forecasts of COVID-19 daily admissions generated using QRF.

forecasting. We use the *ETS*() and *ARIMA*() functions in the fable package (O'Hara-Wild *et al*., 2020) in R to produce daily forecasts of COVID-19 admissions using these methods. For further description of these time series methods, please refer to Hyndman & Athanasopoulos (2021).

We also consider time series LR, ARIMAX and Prophet to account for the association between predictors and daily admissions.

- Time Series LR

Time series LR is widely used in practice and is a common benchmark for even complex forecasting tasks. The LR algorithm learns to construct a weighted sum of its input features. We use TSLM() function in fable package (O'Hara-Wild *et al*., 2020). The LR model uses trend(), season() and all predictors described in Table 2. The function trend() accounts for linear trend and the season() creates dummy variables for the day of week. We denote Y_t as the number of patients admitted to the hospital on a day t. The selected predictors to be included in the LR model are described in Table 2 are presented using the following notations:

$$
\begin{aligned}
Y_t = \beta 0 + \beta 1 \text{ trend} t + \sum_{k=1}^{8} \beta k + 1 Y t - k + \beta 10 case7\ typlust - 1 + \beta 11 total\ NHS 111 t - 1 \\
+ \beta 12 total\_tested t - 1 + \beta 13 percentage\_firstdose_{t-1} + \beta 14 percentage\_seconddose_{t-1} \\
+ \beta 15 apple\_driving_{t-1} + \beta 16 google\_grocery_{t-1} + \beta 17 google\_residential_{t-1} \\
+ \beta 17 google\_transit_{t-1}
\end{aligned}
\tag{2}
$$

In building the LR model, we first create a LR model to forecast one step ahead.

To select the predictors to include in the model, a forward stepwise selection based on Akaike Information Criterion Corrected (AICc) is used. In choosing predictors, we account for linear association with admissions and the existing collinairity between predictors. These decisions are informed by the data analysis presented in Figs 3–5. To forecast the admissions for h-step-ahead, we use the lags of predictors chosen for one-step-ahead forecast rather than forecasting the predictors. Therefore, we always use the actual information until day t to forecast the future. It is interesting to note that we also tested the approach of forecasting the predictors, which led to worse behaviour.

• ARIMAX

Although ARIMA only uses time series information, an ARIMAX model can be constructed by adding predictors to the ARIMA model. We use the same information as ARIMA model plus all predictors used in the LR model, except the hospital admission lagged values. These are removed as ARIMA models account for autoregressive lagged values. We employ the ARIMA() function in fable package in R to train and estimate the parameters of ARIMAX models and forecast daily COVID-19 admissions (O'Hara-Wild *et al.*, 2020).

• Prophet

Prophet is a forecasting procedure created by Facebook. The method shown to be accurate for daily time series, especially on daily data seen on Facebook. The model accounts for multiple seasonality, piecewise trend and any number of predictors. The idea behind piecewise trend is that if the time series follows different linear trends over different regions of the series, then it could be modelled as connected linear pieces (Taylor & Letham, 2018).

The response variable on a day t can be written as the sum of three functions of time plus an error term:

$$y_t = g(t) + s(t) + h(t) + Xreg(t) + \varepsilon_t,$$

where $g(t)$ represents a piecewise linear trend, $s(t)$ indicates various types of seasonality, $h(t)$ captures the holiday effects, $Xreg(t)$ accounts for predictors used in the LR model and ε_t is the error term. We should note that given the short length of the daily time series admissions, modelling the effect of holiday effects is not feasible and this has not been considered in this study.

Prophet is an automated method, which makes it easy to use in many organizations. It is robust to missing data and shifts in the trend, and typically handles outliers well (Taylor & Letham, 2018). We use the fable interface for prophet() in the fable.prophet package in R (O'Hara-Wild, 2020).

The experiment could benefit from including other benchmarks such as Theta (Assimakopoulos & Nikolopoulos, 2000) to reduce the running time or in cases where more data are available, it would be worth investigating methods based on artificial neural networks (ANNs), including deep learning.

5. Experimental setup

We first split the data set into training (80%) and test (20%) sets. We apply the forecasting models on the training set and evaluate the forecast performance of the proposed QRF approach and benchmarks on the test set. Evaluation is conducted using a rolling origin forecasting study on the test set with re-estimation. Forecasting horizon is chosen up to $h = 7$ which corresponds to 7 days that allows decision makers to use forecast to inform decisions on resource allocation and planning in the health service. It should be noted that, given the nature of the daily time series of COVID-19 admissions where a considerable number observations are zero or low volume, the benchmark methods may generate negative forecast

values. Therefore, we first transform the time series using sqrt(), then train and produce the forecasts using all benchmarks. Next, the generated forecasts are transformed back to the original scale (rounded to nearest integer) and the truncation at zero is applied to avoid producing negative forecasts.

5.1   *Forecast accuracy evaluation*

In this study, we consider point forecast accuracy metrics, as well as a metric to evaluate the accuracy of the forecast distribution. We evaluate the point forecasts using RMSE and MAE. MAE is appropriate for the median and RMSE for the mean forecast.

For evaluating the forecasts of the discrete probability distributions, we employ the ranked probability score (RPS), which was introduced by Epstein (1969). The RPS quantifies the difference between the actual observations and the cumulative distribution of forecasts, such that a smaller score is better.

For both point and probabilistic accuracy metrics, we first report the overall performance of each model by averaging the score of each measure computed for all rolling origins and all horizons. We also report the accuracy measure for each model and each forecasting horizon. To that end, we aggregate the score of different rolling origins for a given horizon.

6. Empirical results and discussion

In this section, we first discuss the important predictors of the daily COVID-19 admissions derived from QRF, and report the forecasting performance of QRF model and different benchmarks for point and probabilistic forecasts.

In the context of this study, identifying the most salient features can help provide valuable insights into the key drivers of admission counts, which can potentially be of practical benefit to hospitals and policymakers.

In Fig. 7, we plot the feature importance scores for the most salient features, whereby the ranking of features was obtained using random forests (for horizon equal to 1 day). The feature rankings were calculated using only the OOB observations. The feature rankings were obtained using the permutation feature importance measurement; for details, see Breiman (2001). Specifically, to measure the influence of a feature in predicting the response, we calculate the change in model error over the OOB observations when a feature is permuted. Note that permuting the more salient features will result in a higher increase in the prediction error. We considered features that could potentially improve the forecast accuracy without making the feature matrix high dimension. We considered multiple lags (more than 1 day) for other exogenous features, but it did not result in any noticeable improvement in the modelling. It could be surmised that some of the features (such as vaccination status, number of cases etc.) tend to vary smoothly and this variation could potentially be captured in the hospital admissions data over longer lead times. Interestingly, to predict admission count for a given day, lagged admission counts turn out to be strong predictors, which is broadly in keeping with the autocorrelation and partial autocorrelation in the admissions time series (Fig. 2).

Using the feature rankings obtained from RF for each horizon, we generated out-of-sample forecasts using only the 5, 10, 15 and 20 most salient features and compared their performance with a model that employed all 25 features. Specifically, the QRF model performance for a different number of most salient features was investigated for each horizon. Using only 5 or 10 most salient features resulted in the worst performance overall across all horizons. Moreover, 20 most salient features provided a slightly better performance for 1- and 2-day-ahead forecasts than using all 25 features. For longer horizons
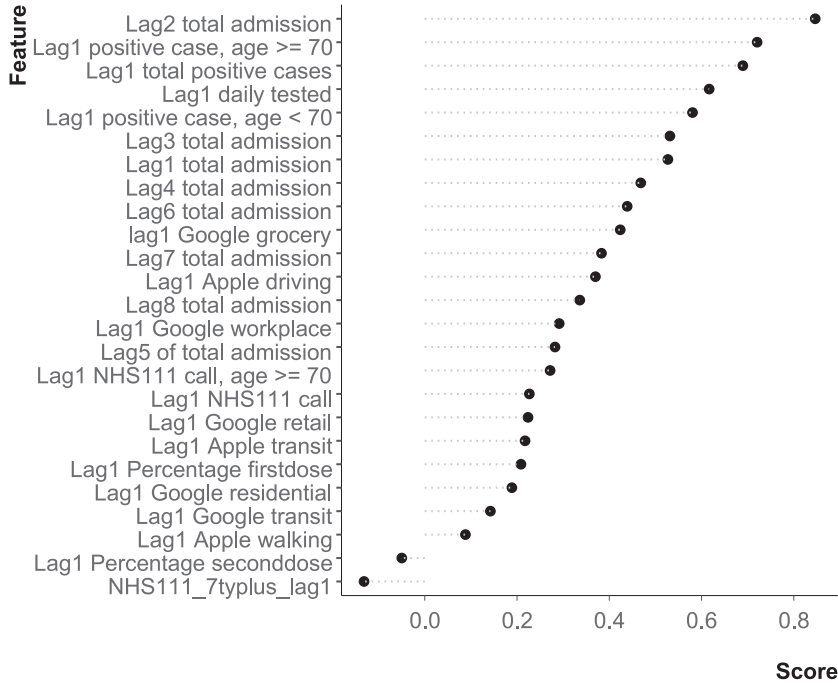
FIG. 7. Importance of the predictors in forecasting COVID-19 admissions.

(3- to 7-day-ahead forecasts), employing all 25 features during the modelling gave the most accurate performance.

Features relating to the total number of positive COVID-19 cases and the number of daily tests performed were also salient, as one might expect. For mobility, features corresponding to Google grocery and Apple driving were salient; however, the mobility features ranked much lower compared with lagged admission counts. Based on these results, it could be argued that the temporal correlations in the admissions count time series can be accommodated using lagged admissions count at a local hospital level.

Table 3 presents an overall forecasting performance of all models for both point forecast and forecast distribution. The result indicates that QRF model outperforms benchmarks for all forecast accuracy measures. Next, we look at the performance across different forecasting horizons.

In Fig. 8(a and b), we plot the RMSE and MAE for all models considered in this study for forecast horizons ranging from 1 day to 7 days ahead. As discussed earlier, we generate forecasts using a rolling forecast origin. Thus, the error metrics for an h-step-ahead forecasts are calculated using n-h + 1 test observations (where n denotes the total number of observations in the test data). From Fig. 8(a and b), it is evident that QRF and ETS are the most competitive models, with QRF being the best performing model by far, except for the horizon of 1 day when the difference between QRF and EST and LR is negligible.

The model rankings were consistent when the evaluation was based on the RPS, see Fig. 8(c). We used the RPS to quantify the accuracy of the forecasts of the discrete probability distributions, as it is

Table 3   *Overall forecast accuracy for admissions*

Forecasting model	MAE	RMSE	RPS
QRF	**2.0254**	**2.6974**	**0.1070**
ETS	2.2688	2.9255	0.1136
LR	2.2751	2.9283	0.1159
ARIMAX	2.3204	2.9702	0.1177
ARIMA	2.3479	2.9955	0.1186
Prophet	2.5471	3.1157	0.1205

(a) RMSE accuracy measure

(b) MAE accuracy measure

(c) RPS accuracy measure
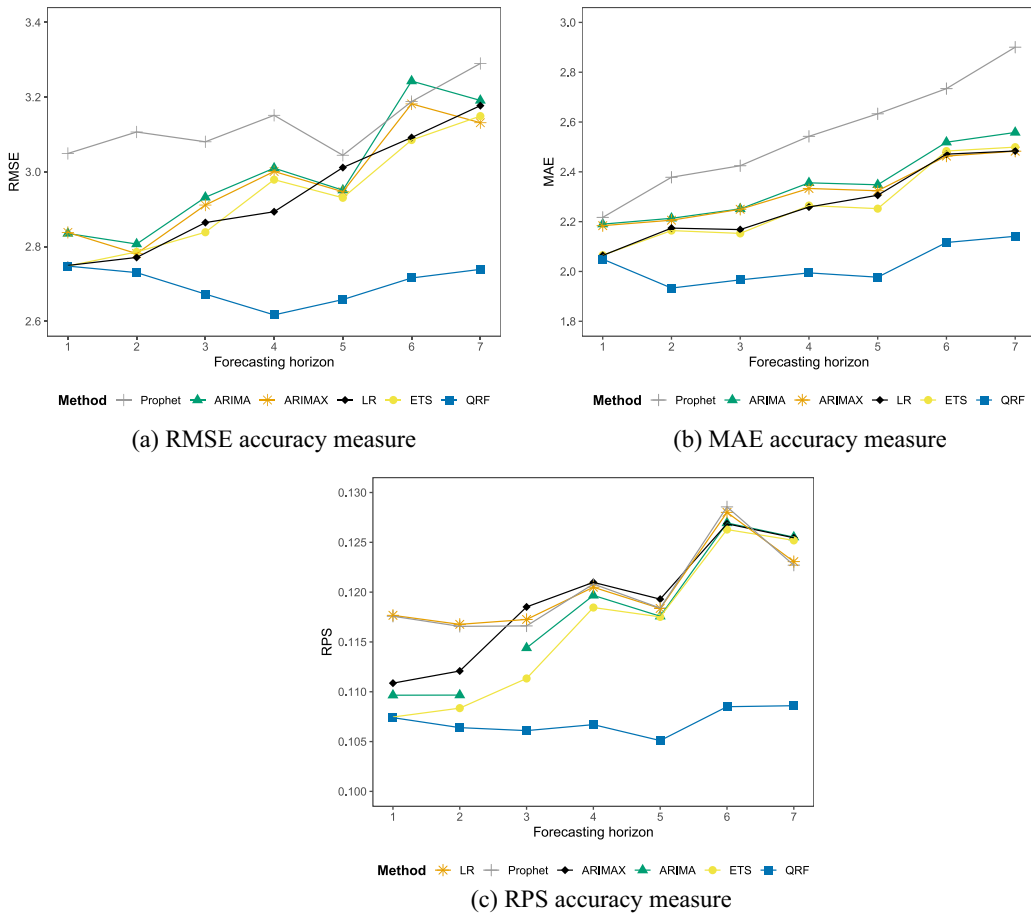
FIG. 8. Point and probabilistic forecast accuracy measure over 7-days forecast horizon.

strictly proper and sensitive to distance [for details, see (Murphy, 1969, 1970)]. The performance of the QRF was considerably better than all the other models used in this study.

Compared with the other methods, the superior performance of the QRF could potentially be attributed to the following facts: (1) QRF appropriately treated the admissions data as being count and

not continuous. QRF generated forecasts that were non-negative integer counts (as the model predictions were realizations of admission counts from the historical data). Since actual admission counts cannot be negative or fractions, admission forecasts from all methods except QRF had to be truncated to exclude any negative predictions, whereas the positive fractional predictions had to be rounded to the nearest integer. (2) QRF provided a non-parametric modelling approach and did not rely on any assumptions for the underlying residual distribution, and crucially, allowed for a non-linear relation between the predictors and the response, as evident from Fig. 5. (3) We used horizon-specific modelling approach for the QRF. Specifically, we estimated seven different horizon-specific QRF models, whereby the hyperparameters of each model were tuned separately using the training data. Whereas for the time series models, such as ARIMA, ARIMAX, ETS, we used the traditional approach of estimating a single model to generate forecasts across multiple horizons. However, the model parameters are re-estimated for each rolling origin.

The proposed modelling framework could be used in practice as the predictors are publicly accessible, and the models employed in this study can generate predictions in real time. We use two broad categories of predictors: (i) lagged values of daily COVID-19 admissions, which are accessible via electronic health record systems, and (ii) population level predictors, such as: confirmed cases of COVID-19, vaccination status, call center data, mobility data, etc., all of which are freely available online. In terms of computational speed, having identified the optimal hyperparameters, the total time to train a QRF model and generate forecasts for the next day was 0.47 s (in Matlab 2021b ©Mathworks, using a 2.2 GHz processor with 32 GB RAM). It was also worth noting that the policy and operations of the hospitals considered in this study are determined by the NHSs, which is the publicly funded healthcare system in England. Thus, the methods developed for the CDG considered in this study could potentially be adapted for other groups in England that operate under the operational framework of the NHS.

7. Conclusion

An accurate estimation of COVID-19 patients admitted to hospitals will help hospitals and regional authorities to allocate resources more effectively. Although many research studies have considered forecasting COVID-19 hospital admissions, less attention has been given to admissions at local levels. The accurate forecast of COVID-19 admissions at the local level is crucial but it presents various forecasting challenges, which might not be suited for many time series forecasting methods.

The daily time series of COVID-19 admissions has the following attributes which make the forecasting challenging: (i) admissions are count data, which require the forecasts to be integer as well; and (ii) they are low and contain many zero values which may result in negative forecasts when using time series forecasting methods. Moreover, the main focus of the research has been on modelling the full probability distribution of COVID-19 admissions and not just a point estimate.

We use the total daily time series of COVID-19 admissions in a Commissioning Group in England to train the models and evaluate the forecast accuracy. The performance of the proposed QRF model is compared with ETS, ARIMA, LR, ARIMAX and Prophet using two point forecast error measures (i.e. RMSE, MAE) and a probabilistic forecasting accuracy measures (RPS). The methodology developed in this study is based on the aggregated data for a subset of hospitals. However, we believe that it could be adopted for each individual hospital given some common features such as trends, low volume observations and presence of zeros.

In this study, we built a QRF model which guarantees the generation of strictly positive forecasts. A limitation of this study is that we only have daily data for slightly more than a year. In hospitals where

feature-rich data may be available for longer, an interesting line of future work would be to employ ANNs, including deep learning, for probabilistic forecasting of admission counts.

We should note that given the uncertainty around the future COVID-19 patients admitted to hospitals, it is crucial to generate and provide a probabilistic forecast, which allows a decision-maker to assess the potential risk of making a decision, hence making a more informed decision. Therefore, in addition to point forecasts, we also provide the distribution of the future daily COVID-19 admissions. Moreover, we investigate the usefulness of various predictors, including COVID-19 phone call data, confirmed cases, number of virus tests performed, percentage of the population vaccinated and human mobility data from Google and Apple, in forecasting daily COVID-19 admissions.

Results indicate that the proposed QRF model outperforms the benchmarks in terms of both point forecast and predictive distribution accuracy. We observe that the accuracy of the QRF model is consistent across different horizons, and the gain in forecast accuracy becomes more prominent as the forecasting horizon increases. Our findings indicate that lagged COVID-19 admissions, number of positive COVID-19 cases, the number of daily tests performed, Google grocery and Apple driving are among the most strong predictors of COVID-19 admissions for a given day.

The generalizability of our findings could be considered in (a) the context of modelling demand under unprecedented conditions, such as the pandemic, and (b) using population-level features to improve forecast accuracy at a local hospital level. Since before Covid-19, it has been observed that the emergence of novel pandemic agents seems unpredictable (Morse *et al.*, 2012). Given the increased cross-species viral transmission risk (Carlson *et al.*, 2022), this might not be the last pandemic. Although the generalizability of our findings might be difficult to assess, but the overall conditions in which the research framework was designed and implemented can serve as a reference for similar future events. Forecasting Covid-19 admissions at the hospital level is a challenging task, and our results can provide insights for similar situations, where the demand for hospital resources due to a highly transmissible virus was reduced with an increase in the level of vaccination status in the population. Another important insight was related to incorporating the mobility data, which captures how people move around and interact, which can result in forecast accuracy improvement. This suggests that population-level features can be essential predictors in emergency situations, as they can capture the spread of an infectious disease or the movement of people during a natural disaster. In addition, this study highlights the importance of identifying and using leading predictors, such as calls to emergency telephone lines, the rate of positive cases and the population vaccinated, to improve prediction accuracy. Another point worth considering is the handling of low-volume data, which can be an issue in emergency situations. We note that traditional time series methods may not be effective when dealing with count data, especially when values are near zero. In such cases, rounding or truncation may be necessary to use conventional time series methods. However, we found that the Quantile Random Forest (QRF) method performed better with count data, suggesting that it may be a useful approach for emergency situations where low-volume data are a concern. In summary, the paper provides several insights that can be applied to other emergency situations. By incorporating population-level features, identifying appropriate leading predictors and using appropriate modelling techniques, it is possible to develop accurate forecasts that can help emergency personnel better prepare for and respond to emergencies. On a smaller note, it is interesting to observe how a dramatic event such as Covid is quickly forgotten once it is no longer in the spotlight. In emergencies such as a pandemic, our difficulty or incapacity to learn from the past means that we have an increased risk in the future of not having the right pieces in place to confront a problem. The risks of the acceleration of such forgetfulness should be carefully assessed, especially under the climate crisis.

Given the importance of forecasting patients admitted to hospital in an emergency situations such a pandemic, research into any of the following areas would appear to be merited:

- We believe that different predictors might have varying importance during different waves of the pandemic. Therefore, investigating what predictors should be used in the forecasting model as the pandemic evolves is an important question, and developing a flexible forecasting model that can adapt to the change would be valuable.

- Similarly, it may not be prudent to rely on one model to perform well across all horizons or waves. Forecast combinations could prove beneficial as they have in other areas of forecasting (e.g. in the M4 forecasting competition, all 6 top performing methods employed forecast combinations of some kind (Makridakis *et al.*, 2018)]. When it comes to COVID-19-related research, Taylor & Taylor (2021) demonstrated the efficacy of combining interval and distributional forecasts for modeling the cumulative mortality due to COVID-19. It would be worth investigating the accuracy of combined forecasts from different models for each horizon (and/or or wave).

- Forecasting the risk score of admissions i.e. predicting the probability of a threshold being exceeded, instead of predicting the exact number for a given day, is an interesting avenue. This may provide a more straightforward tool for a decision-maker to deal with uncertainty.

- Given the nature of the admission time series at the local level, employing a temporal aggregation approach might improve forecast accuracy (Rostami-Tabar *et al.*, 2013, 2014).

- Finally, although this study focuses on the local level, it is important to investigate the usefulness of the predictors discussed in this paper to forecast COVID-19 admissions at more aggregate (e.g. national) levels.

Data availability statement

Data used in this study are accessible via the GitHub Repository: https://github.com/bahmanrostamitabar/Covid19_hospital_admission_data.

Funding

Conflict of Interest

There is no conflict of interest.

Authors' contributions

Bahman Rostami-Tabar (Conceptualization, Programming, Formal analysis, Model development. Writing—Original draft preparation), Siddharth Arora (Conceptualization, Programming, Formal analysis, Model development. Writing—Original draft preparation), Juan Rendon Sanchez (Conceptualization, literature review, Writing—Original draft preparation), Thanos E. Goltsos (Conceptualization. Writing—Original draft preparation).

REFERENCES

ARORA, S., TAYLOR, J. W. & MAK, H.-Y. (2020) Probabilistic forecasting of patient waiting times in an emergency department. Manufacturing & Service Operations Management, forthcoming. https://doi.org/10.1287/msom.2023.1210.

ASSIMAKOPOULOS, V. & NIKOLOPOULOS, K. (2000) The theta model: a decomposition approach to forecasting. *Int. J. Forecast.*, **16**, 521–530.

AU, L., BYRNES, G., BAIN, C. A., FACKRELL, M., BRAND, C., CAMPBELL, D. A. & TAYLOR, P. G. (2009) Predicting overflow in an emergency department. *IMA J. Manage. Math.*, **20**, 39–49.

BALCAN, D., COLIZZA, V., GONÇALVES, B., HU, H., RAMASCO, J. J. & VESPIGNANI, A. (2009) Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci. U S A*, **106**, 21484–21489. https://www.pnas.org/content/106/51/21484.

BBC. (2022) *UK Covid Cases Jump by Almost 30% as Pressure Grows on NHS*. https://www.bbc.co.uk/news/live/uk-62175000.

BEKKER, R., BROEK, M. U. H. & KOOLE, G. (2023) Modeling covid-19 hospital admissions and occupancy in the Netherlands. *Eur. J. Oper. Res.*, **304**, 207–218.

BENVENUTO, D., GIOVANETTI, M., VASSALLO, L., ANGELETTI, S. & CICCOZZI, M. (2020) Application of the Arima model on the covid-2019 epidemic dataset. *Data Brief*, **29**, 105340. http://www.sciencedirect.com/science/article/pii/S2352340920302341.

BISHOP, J. A., JAVED, H. A., EL-BOURI, R., ZHU, T., TAYLOR, T., PETO, T., WATKINSON, P., EYRE, D. W. & CLIFTON, D. A. (2021) Improving patient flow during infectious disease outbreaks using machine learning for real-time prediction of patient readiness for discharge. *PLoS One*, **16**, e0260476.

BOBASHEV, G., SEGOVIA-DOMINGUEZ, I., GEL, Y. R., RINEER, J., RHEA, S. & SUI, H. (2020) Geospatial forecasting of covid-19 spread and risk of reaching hospital capacity. *SIGSPATIAL Special*, **12**, 25–32.

BONANAD, C., GARCÍA-BLAS, S., TARAZONA-SANTABALBINA, F., SANCHIS, J., BERTOMEU-GONZÁLEZ, V., FÁCILA, L., ARIZA, A., NÚÑEZ, J. & CORDERO, A. (2020) The effect of age on mortality in patients with covid-19: a meta-analysis with 611,583 subjects. *J. Am. Med. Dir. Assoc.*, **21**, 915–918.

BOYLAN, J. E. (2016) Reproducibility. *IMA J. Manage. Math.*, **27**, 107–108.

BREIMAN, L. (2001) Random forests. *Mach. Learn.*, **45**, 5–32.

CAPISTRAN, M. A., CAPELLA, A. & CHRISTEN, J. A. (2020) Forecasting hospital demand during covid-19 pandemic outbreaks. arXiv preprint arXiv:2006.01873.

CARLSON, C. J., ALBERY, G. F., MEROW, C., TRISOS, C. H., ZIPFEL, C. M., ESKEW, E. A., OLIVAL, K. J., ROSS, N. & BANSAL, S. (2022) Climate change increases cross-species viral transmission risk. *Nature*, **607**, 555–562.

DEASY, J., ROCHETEAU, E., KOHLER, K., STUBBS, D. J., BARBIERO, P., LIO, P. & ERCOLE, A. (2020) Forecasting ultra-early intensive care strain from covid-19 in England. MedRxiv.

EPSTEIN, E. (1969) A scoring system for probability forecasts of ranked categories. *J. Appl. Meteorol. Climatol.*, **8**, 985–987.

FANELLI, D. & PIAZZA, F. (2020) Analysis and forecast of covid-19 spreading in China, Italy and France. *Chaos Solit. Fract.*, **134**, 109761. http://www.sciencedirect.com/science/article/pii/S0960077920301636.

FEROZE, N. (2020) Forecasting the patterns of covid-19 and causal impacts of lockdown in top five affected countries using Bayesian structural time series models. *Chaos Solit. Fract.*, **140**, 110196. http://www.sciencedirect.com/science/article/pii/S0960077920305920.

FONG, S. J., LI, G., DEY, N., GONZALEZ-CRESPO, R. & HERRERA-VIEDMA, E. (2020) Finding an accurate early forecasting model from small dataset: a case of 2019-ncov novel coronavirus outbreak. *Int. J. Interact. Multimedia Artif. Intell.*, **6**, 132. http://dx.doi.org/10.9781/ijimai.2020.02.002.

FUNK, S., ABBOTT, S., ATKINS, B. D., et al. (2020) Short-term forecasts to inform the response to the covid-19 epidemic in the UK. medRxiv.

GALLIVAN, S. & UTLEY, M. (2005) Modelling admissions booking of elective in-patients into a treatment centre. *IMA J. Manage. Math.*, **16**, 305–315.

GITTO, S., DI MAURO, C., ANCARANI, A. & MANCUSO, P. (2021) Forecasting national and regional level intensive care unit bed demand during covid-19: the case of Italy. *PLoS One*, **16**, e0247726.

GOLTSOS, T. E., SYNTETOS, A. A., GLOCK, C. H. & IOANNOU, G. (2021) Inventory—forecasting: mind the gap. *Eur. J. Oper. Res.*, **299**, 397–419.

GRASSELLI, G., PESENTI, A. & CECCONI, M. (2020) Critical care utilization for the covid-19 outbreak in Lombardy, Italy: early experience and forecast during an emergency response. *JAMA*, **323**, 1545–1546.

HAMZAH, F. B., LAU, C., NAZRI, H., LIGOT, D., LEE, G., TAN, C., SHAIB, M., ZAIDON, U., ABDULLAH, A., CHUNG, M., et al. (2020) Coronatracker: worldwide covid-19 outbreak data analysis and prediction. *Bull. World Health Organ.*, **1**, 32.

HARVEY, A. (1984) Time series forecasting based on the logistic curve. *J. Oper. Res. Soc.*, **35**, 641–646.

HAZARIKA, B. B. & GUPTA, D. (2020) Modelling and forecasting of Covid-19 spread using wavelet-coupled random vector functional link networks. *Appl. Soft Comput.*, **96**, 106626. http://www.sciencedirect.com/science/article/pii/S1568494620305640.

HONG, W. S., HAIMOVICH, A. D. & TAYLOR, R. A. (2018) Predicting hospital admission at emergency department triage using machine learning. *PLoS One*, **13**, e0201016.

HYNDMAN, R. J. & ATHANASOPOULOS, G. (2021) *Forecasting: Principles and Practice*. OTexts, Melbourne, Australia. https://otexts.com/fpp3

IOANNIDIS, J. P., CRIPPS, S. & TANNER, M. (2022) Forecasting for covid-19 has failed. *Int. J. Forecast.*, **38**, 423–438.

IVANOV, D. (2020) Predicting the impacts of epidemic outbreaks on global supply chains: a simulation-based analysis on the coronavirus outbreak (covid-19/sarscov-2) case. *Transport. Res. E Logist. Transport. Rev.*, **136**, 101922.

KERMACK, W. O. & MCKENDRICK, A. G. (1927) A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A Math. Phys. Sci.*, **115**, 700–721.

LECLERC, Q., FULLER, N. M., KEOGH, R. H., DIAZ-ORDAZ, K., SEKULA, R., SEMPLE, M. G., ISARIC4C INVESTIGATORS; CMMID COVID-19 WORKING GROUP, ATKINS, K. E., PROCTER, S. R. & KNIGHT, G. M. (2021) Importance of patient bed pathways and length of stay differences in predicting covid-19 hospital bed occupancy in England. *BMC Health Serv. Res.*, **21**, 566.

MAKRIDAKIS, S., SPILIOTIS, E. & ASSIMAKOPOULOS, V. (2018) The m4 competition: results, findings, conclusion and way forward. *Int. J. Forecast.*, **34**, 802–808.

MARTELLONI, G. & MARTELLONI, G. (2020) Modelling the downhill of the SARS-CoV-2 in Italy and a universal forecast of the epidemic in the world. *Chaos Solit. Fract.*, **139**, 110064. http://www.sciencedirect.com/science/article/pii/S0960077920304616.

MEINSHAUSEN, N. (2006) Quantile regression forests. *J. Mach. Learn. Res.*, **7**, 983–999.

MOFTAKHAR, L., MOZHGAN, S. & SAFE, M. S. (2020) Exponentially increasing trend of infected patients with covid-19 in Iran: a comparison of neural network and Arima forecasting models. *Iran. J. Public Health*, **49**, 92–100.

MOGHADAS, S. M., SHOUKAT, A., FITZPATRICK, M. C., WELLS, C. R., SAH, P., PANDEY, A., SACHS, J. D., WANG, Z., MEYERS, L. A., SINGER, B. H. & GALVANI, A. P. (2020) Projecting hospital utilization during the covid-19 outbreaks in the United States. *Proc. Natl. Acad. Sci. U S A*, **117**, 9122–9126. https://www.pnas.org/content/117/16/9122.

MORSE, S. S., MAZET, J. A., WOOLHOUSE, M., PARRISH, C. R., CARROLL, D., KARESH, W. B., ZAMBRANA-TORRELIO, C., LIPKIN, W. I. & DASZAK, P. (2012) Prediction and prevention of the next pandemic zoonosis. *The Lancet*, **380**, 1956–1965. https://www.sciencedirect.com/science/article/pii/S0140673612616845.

MURPHY, A. (1969) On the "ranked probability score". *J. Appl. Meteorol. Climatol.*, **8**, 988–989.

MURPHY, A. (1970) The ranked probability score and the probability score: a comparison. *Mon. Weather Rev.*, **98**, 917–924.

NENOVA, Z. D. & MAY, J. H. (2016) Determining an optimal hierarchical forecasting model based on the characteristics of the data set. *J. Oper.Manage.*, **44**, 62–68.

NIKOLOPOULOS, K., PUNIA, S., SCHÄFERS, A., TSINOPOULOS, C. & VASILAKIS, C. (2021) Forecasting and planning during a pandemic: Covid-19 growth rates, supply chain disruptions, and governmental decisions. *Eur. J. Oper. Res.*, **290**, 99–115. https://www.sciencedirect.com/science/article/pii/S0377221720306913.

O'HARA-WILD, S. T. (2020) *fable.prophet: Prophet Modelling Interface for 'Fable'*. R package version 0.1.0. https://CRAN.R-project.org/package=fable.prophet

O'HARA-WILD, M., HYNDMAN, R., WANG, E. & CACERES, G. (2020) *fable: Forecasting Models for Tidy Time Series*. R package version 0.2.1. https://CRAN.R-project.org/package=fable

QIAN, Z., ALAA, A. M. & VAN DER SCHAAR, M. (2021) CPAS: the UK's national machine learning-based hospital capacity planning system for covid-19. *Mach. Learn.*, **110**, 15–35.

RAHIMIAN, F., SALIMI-KHORSHIDI, G., PAYBERAH, A. H., TRAN, J., AYALA SOLARES, R., RAIMONDI, F., NAZARZADEH, M., CANOY, D. & RAHIMI, K. (2018) Predicting the risk of emergency admission with machine learning: development and validation using linked electronic health records. *PLoS Med.*, **15**, e1002695.

ROSTAMI-TABAR, B. & RENDON-SANCHEZ, J. F. (2021) Forecasting covid-19 daily cases using phone call data. *Appl. Soft Comput.*, **100**, 106932.

ROSTAMI-TABAR, B. & ZIEL, F. (2022) Anticipating special events in emergency department forecasting. *Int. J. Forecast.*, **38**, 1197–1213.

ROSTAMI-TABAR, B., BABAI, M. Z., SYNTETOS, A. & DUCQ, Y. (2013) Demand forecasting by temporal aggregation. *Naval Res. Logist.*, **60**, 479–498.

ROSTAMI-TABAR, B., BABAI, M. Z., SYNTETOS, A. & DUCQ, Y. (2014) A note on the forecast performance of temporal aggregation. *Naval Res. Logist.*, **61**, 489–500.

ROSTAMI-TABAR, B., BROWELL, J. & SVETUNKOV, I. (2023) Probabilistic forecasting of hourly emergency department arrivals. *Health Syst*, 1–17.

SARDAR, T., NADIM, S. S., RANA, S. & CHATTOPADHYAY, J. (2020) Assessment of lockdown effect in some states and overall India: a predictive mathematical study on covid-19 outbreak. *Chaos Solit Fract.*, **139**, 110078. http://www.sciencedirect.com/science/article/pii/S0960077920304756.

SATTAR, N., HO, F. K., GILL, J. M., GHOURI, N., GRAY, S. R., CELIS-MORALES, C. A., KATIKIREDDI, S. V., BERRY, C., PELL, J. P., MCMURRAY, J. J., et al. (2020) BMI and future risk for covid-19 infection and death across sex, age and ethnicity: preliminary findings from UK biobank. *Diab. Metab. Syndr. Clin. Res. Rev.*, **14**, 1149–1151.

SHAH, S. A., MOORE, E., ROBERTSON, C., MCMENAMIN, J., KATIKIREDDI, S. V., SIMPSON, C. R., SHI, T., AGRAWAL, U., MCCOWAN, C., STOCK, S., et al. (2021) Predicted Covid-19 positive cases, hospitalisations, and deaths associated with the delta variant of concern, June–July, 2021. *Lancet Digit. Health*, **3**, e539–e541.

TAYLOR, S. J. & LETHAM, B. (2018) Forecasting at scale. *Am. Stat.*, **72**, 37–45.

TAYLOR, J. W. & TAYLOR, K. S. (2021) Combining probabilistic forecasts of Covid-19 mortality in the United States. *Eur. J. Oper. Res.*, **304**, 25–41.

THE GLEAM PROJECT. (2020) *Covid-19 modelling, United States.* https://covid19.gleamproject.org/, accessed 29 September 2020.

WEISSMAN, G. E., CRANE-DROESCH, A., CHIVERS, C., LUONG, T., HANISH, A., LEVY, M. Z., LUBKEN, J., BECKER, M., DRAUGELIS, M. E., ANESI, G. L., et al. (2020) Locally informed simulation to predict hospital capacity needs during the covid-19 pandemic. *Ann. Intern. Med.*, **173**, 21–28.

WHO (2021) *Considerations for Implementing and Adjusting Public Health and Social Measures in the Context of Covid-19:Iinterim Guidance (14 June 2021).* World Health Organization.

World Health Organization (2020) *Covid-19: Operational Guidance for Maintaining Essential Health Services During an Outbreak: Interim Guidance (25 March 2020).* World Health Organization.

ZAPLOTNIK, Z., GAVRIĆ, A. & MEDIC, L. (2020) Simulation of the Covid-19 epidemic onˇ the social network of Slovenia: estimating the intrinsic forecast uncertainty. *PLoS One*, **15**, e0238090.

ZHAO, C., TEPEKULE, B., CRISCUOLO, N. G., WENDEL-GARCIA, P. D., HILTY, M. P., FUMEAUX, T. & VAN BOECKEL, T. P. (2020) Icumonitoring. Ch: a platform for shortterm forecasting of intensive care unit occupancy during the covid-19 epidemic in Switzerland. *Swiss Med. Wkly.*, **150**, w20277.