scientific reports

OPEN

Forecasting admissions in psychiatric hospitals before and during Covid-19: a retrospective study with routine data

J. Wolff [1,2]✉, A. Klimke[3,4], M. Marschollek[1] & T. Kacprowski[5,6]

The COVID-19 pandemic has strong effects on most health care systems. Forecasting of admissions can help for the efficient organisation of hospital care. We aimed to forecast the number of admissions to psychiatric hospitals before and during the COVID-19 pandemic and we compared the performance of machine learning models and time series models. This would eventually allow to support timely resource allocation for optimal treatment of patients. We used admission data from 9 psychiatric hospitals in Germany between 2017 and 2020. We compared machine learning models with time series models in weekly, monthly and yearly forecasting before and during the COVID-19 pandemic. A total of 90,686 admissions were analysed. The models explained up to 90% of variance in hospital admissions in 2019 and 75% in 2020 with the effects of the COVID-19 pandemic. The best models substantially outperformed a one-step seasonal naïve forecast (seasonal mean absolute scaled error (sMASE) 2019: 0.59, 2020: 0.76). The best model in 2019 was a machine learning model (elastic net, mean absolute error (MAE): 7.25). The best model in 2020 was a time series model (exponential smoothing state space model with Box-Cox transformation, ARMA errors and trend and seasonal components, MAE: 10.44). Models forecasting admissions one week in advance did not perform better than monthly and yearly models in 2019 but they did in 2020. The most important features for the machine learning models were calendrical variables. Model performance did not vary much between different modelling approaches before the COVID-19 pandemic and established forecasts were substantially better than one-step seasonal naïve forecasts. However, weekly time series models adjusted quicker to the COVID-19 related shock effects. In practice, multiple individual forecast horizons could be used simultaneously, such as a yearly model to achieve early forecasts for a long planning period and weekly models to adjust quicker to sudden changes.

Abbreviations

ARIMA	Autoregressive integrated moving average
ETS	Exponential smoothing state space models
ICU	Intensive care unit
IQR	Interquartile range
MAE	Mean absolute error
PROPHET	Additive models with non-linear trends fitted by seasonal effects
RMSE	Root-mean- squared-error

[1]Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover Medical School, Carl-Neuberg-Straße 1, 30625 Hannover, Germany. [2]Marienstift Hospital, Helmstedter Straße 35, 38102 Braunschweig, Germany. [3]Vitos Hochtaunus, Friedrichsdorf, Emil-Sioli-Weg 1-3, 61381 Friedrichsdorf, Germany. [4]Heinrich-Heine-University Duesseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany. [5]Division Data Science in Biomedicine, Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, TU Braunschweig, Rebenring 56, 38106 Braunschweig, Germany. [6]Braunschweig Integrated Centre of Systems Biology (BRICS), TU Braunschweig, Rebenring 56, 38106 Braunschweig, Germany. ✉email: wolff.jan@mh-hannover.de

sMASE	Seasonal mean absolute scaled error
SVM	Support vector machines
TBATS	Exponential smoothing state space models with screening for Box-Cox transformation, ARMA errors and trend and seasonal components
XGB	Gradient boosting with trees

Health care systems need to balance potentially unlimited demand for services with scarce health care resources[1]. This balance might become even more difficult in the future due to an aging population that leads to increased demand for health care services and reduced medical work force[2]. Hospitals often consume a large part of total health care budgets[3]. A possible way to mitigate resource constraints in hospitals is the use of modern technologies to make services more cost-effective[4]. For instance, the increasing availability of data allows to support medical decision making in hospitals with information derived from machine learning algorithms[5,6].

Efficient resource allocation in hospitals requires the management of volatile demand and available resources. This management is more critical in hospitals than in other areas since lack of timely and sufficient services can lead to negative patient outcomes[7]. Insufficient staff can lead to increased morbidity and mortality[8,9]. Due to the shortage of trained medical staff in many health care systems, securing sufficient medical staff to permanently meet patient needs becomes a critical objective[10]. Inpatient mental health care is more staff intensive than other medical disciplines due to the personal nature of many interventions[11,12].

The Covid-19 pandemic has strong effects on most health care systems and individual services providers[13]. A sudden surge in patients requiring intensive respiratory care and the possible shortage of ICU capacities led to political supply side interventions in many health care systems[14]. In Hesse, Germany, somatic and psychiatric hospitals were required to restrict new admissions to urgent care cases and avoid elective patients from 16th March 2020. Furthermore, new hospital hygiene regulations and the requirement to quarantine patients reduced hospital capacities.

Forecasting of admissions can help for the efficient organisation of hospital care and for the adjustment of resources to sudden changes in patient volumes. The demand for health care services is usually relatively stable. Several recent studies have compared methods to forecast the spread of the Covid-19 pandemic[15,16]. However, there is still a lack of evidence considering the appropriate accounting for external shocks, which might be impossible to be directly and prospectively accounted for in modelling approaches, in forecasting hospital service volumes. Therefore, we aimed to forecast the number of admissions to psychiatric hospitals before and during the COVID-19 pandemic and we compared the performance of machine learning models and time series models.

Methods

Data.
We included all inpatient admissions from 01 January 2017 to 31 December 2020 to nine hospitals in Hesse, Germany. These hospitals are part of a common service provider and account for about half of all inpatient mental health care in the state of Hesse. Aggregated admission numbers per day were obtained from the hospital administrations and did not contain individual patient data. Returns after planned interruptions, such as home leave, were excluded. Multiple separate admissions of the same patient were counted individually. Admissions to the departments of child and adolescent psychiatry and admissions to the departments for psychosomatic medicine were excluded.

We obtained weather and climate data from the Climate Data Centre of Germany's National Meteorological Service[17]. We used the gtrendsR package version 1.5.1 to query Google trend data for Hesse, Germany[18]. School holidays and public holidays were obtained from publicly available calendars.

Analyses.
We used machine learning and time series models to predict the number of hospital admissions for each day of year 2019 and 2020. The machine learning models were (a) gradient boosting with trees (XGB)[19], (b) support vector machines (SVM)[20] and (c) elastic nets[21]. The time series models were a) exponential smoothing state space models (ETS)[22], (b) exponential smoothing state space models with screening for Box-Cox transformation, ARMA errors and trend and seasonal components (TBATS)[23] and (c) additive models with non-linear trends fitted by seasonal effects (PROPHET)[24]. The selection of modelling approaches was based on their performance in previous research and cannot be exhaustive. However, several other examples were successfully used for forecasting in the corona context and might be relevant to the interested reader[16,25,26]. We compared models forecasting a week in advance, a month in advance and a whole year one week in advance.

Features.
Our machine learning model used calendrical variables, climate and weather data, google trend data, Fourier terms and lagged number of admissions as features. All features are provided with a detailed explanation in Table S1. The calendrical features were day of day of the week, weekend, public holiday, school holiday, quarter of the year, month of the year, bridge days, i.e. days between a public holiday and the weekend and the end of the year, i.e. the days between Christmas and new year's eve. The climate and weather data were wind speed, cloudiness, air pressure, precipitation depth and type, duration of sunshine, snow height, air temperature and humidity. Since the weather of future days was unknown at the point of prediction we used lagged values, i.e. the weekly model used the weather 7 days before the predicted day and the monthly models used the weather data 28 days before the predicted day. We did not use weather data for the yearly model.

Google trend data were retrieved using the gtrendsR package[18] in the R environment for statistical computing[27]. We used the German translations of the following keywords in google trend data: depression, sadness, sad, suicide, mania, fear, panic, dread, addiction, dependence, alcohol, drugs, schizophrenia, psychosis and hallucinations. The relative frequency of searches for these key words in the region of Hesse, Germany, was used as feature. As for the weather data, we used lagged values of google trend data. The weekly models used the

number of admissions 14 days before the predicted day, because the number of admissions was not known yet on day 7 before prediction, as additional feature and the monthly model used these values with a lag of 35. Our time series models did not use feature variables.

Training and testing. We used prospectively sliding time windows to validate (2018) and test (2019 and 2020) model performance. The final weekly models predicted each day of one full week of hospital admissions seven days in advance. We tested one model for each week and study site in 2019 and 2020, thereby incrementally prolonging the training period and forwarding the 7-day testing period each by one week. The monthly models each predicted 28 days of hospital admissions in advance and the incremental slides were 28 days. In the yearly models, we predicted the whole year of 2019 and 2020 each one week before the years started.

We compared model performance with the Root-Mean- Squared-Error (RMSE), the R^2, the Mean Absolute Error (MAE) and a seasonal Mean Absolute Scaled Error (sMASE) as follows[28]:

$$Observation\ at\ time\ t = Y_t$$

$$Forecast\ of\ Y_t = F_t$$

$$Forecast\ error = e_t = Y_t - F_t$$

$$MSE = mean\left(e_t^2\right)$$

$$RMSE = \sqrt{MSE}$$

$$R^2 = correlation(Y_t, F_t)^2$$

$$MAE = mean(|e_t|)$$

$$sMASE = \frac{MAE}{seasonally\ adjusted\ naive\ MAE}$$

The sMASE was calculated by dividing the MAE of our weekly, monthly and yearly forecasts by the MAE derived from a naïve forecast based on the number of admissions observed 14, 35 and 364 days before the predicted day, respectively. Variable importance was calculated for each variable in the best performing model using model specific metrics, i.e. in the case of elastic nets the absolute value of the coefficients after standardizing each feature. An advantage of model-specific metrics compared to model-agnostic measures is that they should be better in accounting for collinearity between features[29].

Ethics approval and consent to participate. Our study did not involve individual patient data but summed numbers of admissions per day. The ethics committee of the Medical School Hannover confirmed that our study did not require ethical oversight.

Results

The total number of admissions showed a relatively strong weekly seasonality and a yearly seasonality. Figure 1 provides the results of a multiple seasonal decomposition of the number of daily admissions by loess[30]. There was no strong trend in admission numbers during the first three years, until the commencement of the Corona hospital regulation on March 16th had a clear negative effect on the number of admissions.

Table 1 shows the forecasting performance in 2019 and in 2020 at all study sites combined. The naïve seasonal forecasts were based on the number of admissions 14, 35 and 364 days before the predicted day for the weekly, monthly and yearly models, respectively. In absolute terms, the best model in 2019 was the weekly elastic net, which achieved a MAE of 7.25 days and an explained variance of 90%. Compared to a naïve forecast based on the number of admissions two weeks in advance, this model achieved a forecast improvement of 38% (sMASE = 0.62). In absolute terms, the best model in 2020 was the weekly TBATS model. However, compared to the monthly possible naïve forecast, i.e. the number of admission 35 days in advance, the highest improvement was achieved with the monthly SVM.

The error accumulation in 2019 and 2020 at all study sites combined is shown in Fig. 2. While model performance was relatively similar in 2019, errors diverged after commencement of the Corona hospital regulation on March 16th, 2020. Weekly time series models adjusted quicker to the new circumstances and accumulated less error until the end of year 2020.

The forecasting models showed variation in performance between study sites. Figure 3 shows differences in percentage errors between study sites per week derived from the overall best performing weekly machine learning and time-series models (see Table 1), respectively. Both models performed similar in year 2019. However, the elastic net caused less error peaks, for instance at easter Monday and during Christmas time because it had these holidays as features. In contrast, the TBATS model adjusted quicker to the corona regulations and adjusted to the new level of admission numbers during the rest of 2020 better than the elastic net.

Figure 4 shows the top 25 feature variables ordered by their importance in forecasting the number of admissions with the elastic net, which was the best performing machine learning algorithm in our comparison. Variable
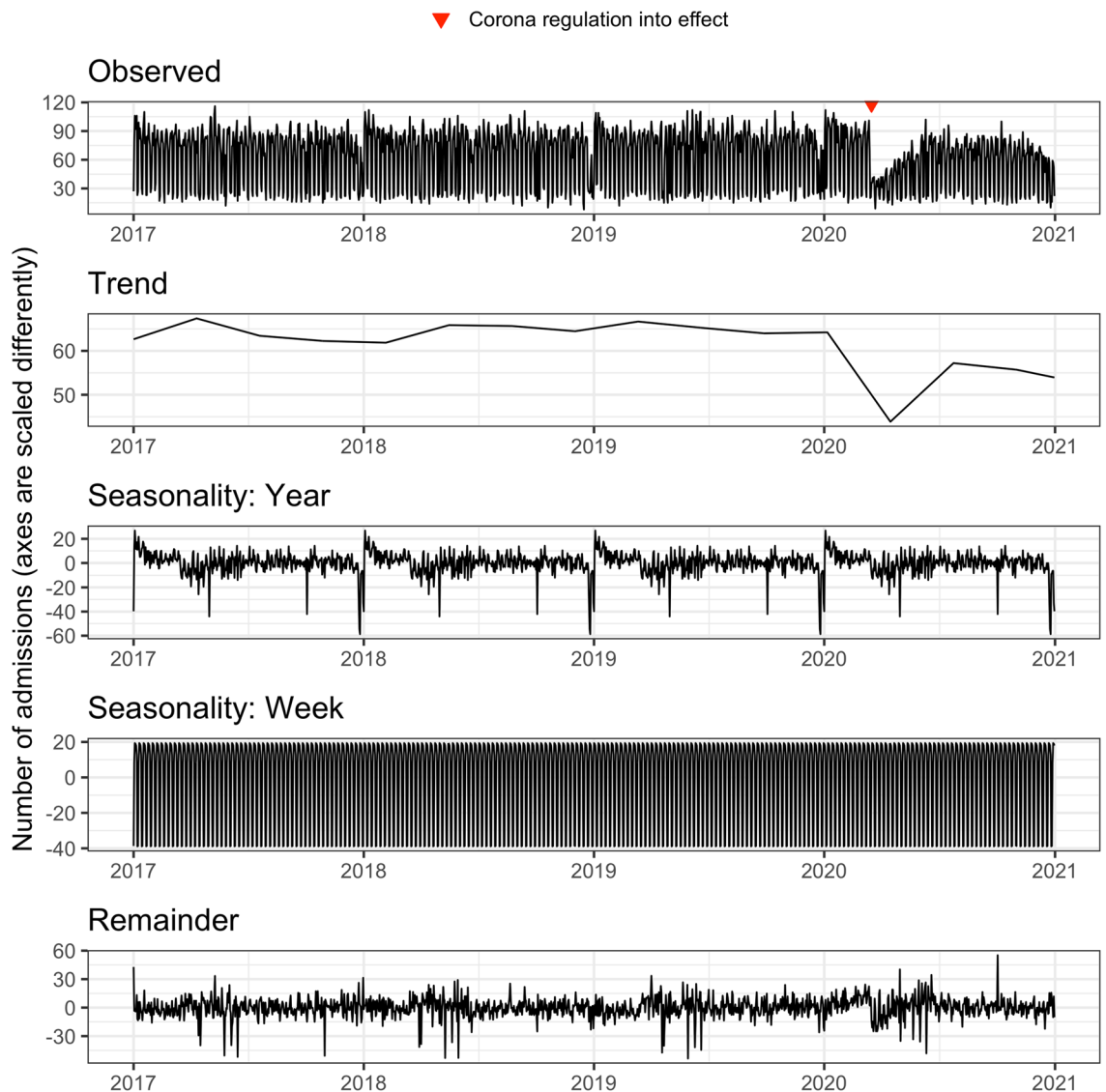
Figure 1. Multiple seasonal decomposition by loess. The y-axes show the number of days and are scaled differently between the facets. Loess = Locally weighted scatterplot smoothing.

importance represents the influence of each feature on the forecast performance relative to the other variables[31]. The strongest influence on forecast performance was found in calendrical variables.

Discussion

Key findings. We aimed to forecast the number of admissions to psychiatric hospitals before and during the COVID-19 pandemic and we compared the performance of machine learning models and time series models. This would eventually allow to support timely resource allocation for optimal treatment of patients. Model performance did not vary much between different modelling approaches before the COVID-19 pandemic. Established forecasts were substantially better than seasonal naïve forecasts. The most important features were calendrical variables that did not require short term adjustments in weekly and monthly models. However, weekly time series models adjusted quicker to the COVID-19 related shock effects than monthly and yearly models and the machine learning models. This is to be expected based on the theory and mechanics underlying the different modelling approaches, since longer forecasting horizons made the models less flexible and slower to adapt to radical changes, and the machine learning models, in contrast to the time series models, based their predictions on many data points from the past, which made them slower to adapt to radical changes.

Strength and weaknesses. A strength of our study were the data of four years from nine hospitals representing about half of all inpatient psychiatric admissions in Hesse, Germany. This allowed both to give a representative picture of inpatient psychiatric care in Germany and to show how the forecasting approaches work at different study sites. Furthermore, it was possible to analyse the effect of sudden changes in hospital admissions

		RMSE		R²		MAE		sMASEª	
		2019	2020	2019	2020	2019	2020	2019	2020
Week	XGB	11.38	16.21	0.86	0.71	8.40	12.06	0.72	1.08
	SVM	10.45	15.68	0.88	0.73	7.56	10.78	0.65	0.97
	Elastic net	**9.65**	15.63	**0.90**	**0.75**	**7.25**	11.42	0.62	1.03
	ETS	13.92	16.19	0.78	0.66	8.65	10.76	0.74	0.97
	TBATS	14.10	**15.43**	0.77	0.70	8.93	**10.44**	0.77	0.94
	PROPHET	13.54	16.13	0.79	0.68	8.81	11.03	0.76	0.99
Month	XGB	11.45	16.50	0.86	0.70	8.44	12.30	0.68	0.84
	SVM	10.49	15.90	0.89	0.73	7.69	11.17	0.62	**0.76**
	Elastic net	9.74	16.36	**0.90**	0.73	7.32	11.94	**0.59**	0.81
	ETS	13.89	18.10	0.78	0.60	8.65	12.31	0.70	0.84
	TBATS	14.15	18.24	0.77	0.60	9.15	12.45	0.74	0.85
	PROPHET	14.18	18.41	0.77	0.59	8.91	12.38	0.72	0.84
Year	XGB	11.31	16.95	0.87	0.71	8.37	12.80	0.67	1.02
	SVM	10.60	16.52	0.88	0.72	7.77	11.61	0.62	0.93
	Elastic net	9.81	16.60	**0.90**	0.74	7.33	12.33	**0.59**	0.98
	ETS	13.77	18.83	0.78	0.66	8.62	12.98	0.69	1.04
	TBATS	13.78	18.08	0.78	0.67	8.85	12.59	0.71	1.00
	PROPHET	13.80	18.83	0.78	0.68	8.63	13.42	0.69	1.07

Table 1. Forecasting performance in 2019 and in 2020. Best values per column are in boldface. *RMSE* Root-mean-squared-error, *MAE* Mean absolute error, *sMASE* Seasonal mean absolute scaled error. ªThe naïve seasonal forecasts were based on the number of admissions 14, 35 and 364 days before the predicted day for the weekly, monthly and yearly models, respectively.
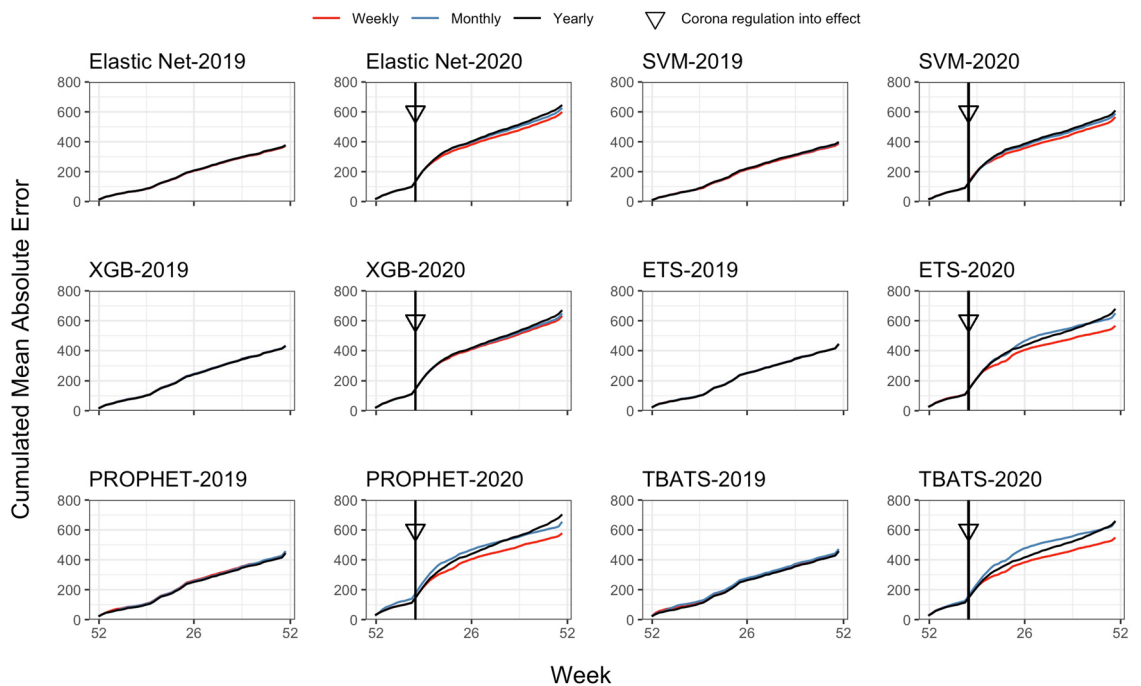

Figure 2. Cumulated mean absolute error in 2019 and 2020 by machine learning and time series models (days). XGB = Gradient boosting with trees, SVM = Support vector machines, ETS = Exponential smoothing state space models, TBATS = Exponential smoothing state space models with Box-Cox transformation, ARMA errors and trend and seasonal components, PROPHET = Additive models with non-linear trends fitted by seasonal effects.

to the performance of different modelling approaches due to the commencement of the Corona hospital regulation in March 2020.

A limitation of our study was the lack of data to differentiate between causes of reduced hospital admissions after the corona regulation came into effect in March 2020. The reduced admissions could have been a result of different supply side and demand side effects, such as avoidance of elective admissions, reduced capacities due
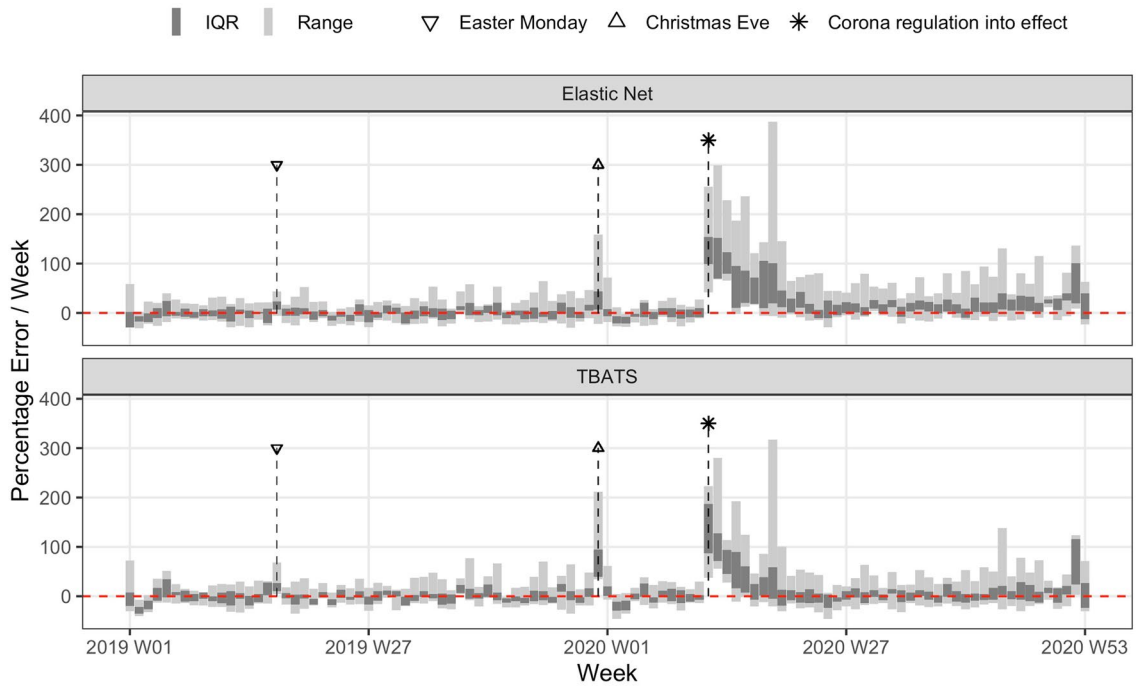
Figure 3. Variation of percentage error between study sites. TBATS = Exponential smoothing state space models with Box-Cox transformation, ARMA errors and trend and seasonal components, IQR = Interquartile range.
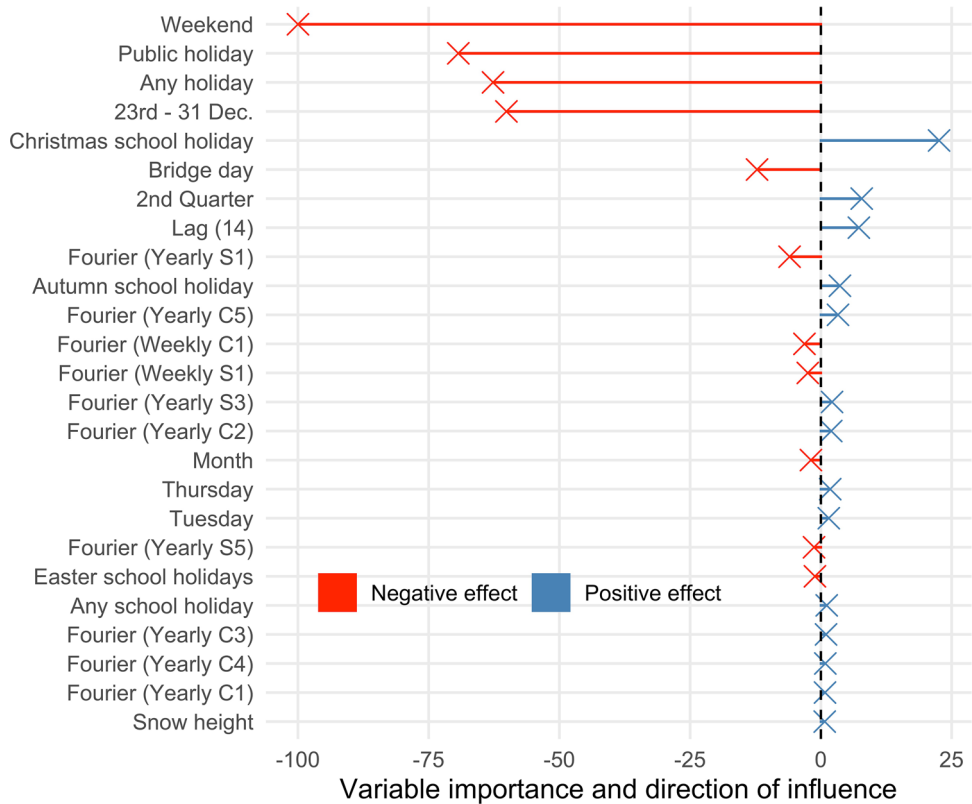

Figure 4. Variable importance of TOP 25 features in machine leaning models. Positive and negative effects represent increases and decreases in the number of admissions, respectively. Dec = December. Bridge day: Day between a holiday and the weekend, vice versa. Lag (14) = The number of admissions fourteen days before the predicted day. The Fourier series accounted to a yearly and a weekly seasonality with sine (S) and cosine (C) waves with an maximum order of 2 (weekly) and 5 (yearly).

to isolation and quarantine requirements and unwillingness of patients to enter hospitals during the Corona crisis. Another limitation of our study was its restriction to one large German provider of inpatient mental health care, which requires a lot of care when translating to different healthcare systems or different clinical settings.

Comparison to previous research. Previous studies in the field of forecasting admissions in hospitals often focused on emergency departments[32] and there were no previous studies that analysed forecasting of psychiatric hospital admission comparable to our study in scale and scope.

Vollmer et al.[33] predicted admission numbers in the emergency departments of two hospitals London with data from 2011 to 2018. They compared machine learning models to more traditional time series models to make forecasts of admissions one, three and seven days in advance. The forecasts of different time horizons, i.e., one, three and seven days in advance, performed very similar. This is comparable to our findings of relatively similar results between weekly, monthly and yearly predictions, although at a different scale. In contrast to our study, lagged admissions from previous weeks were among the strongest predictors, probably related to the stronger increase and decrease of admission number levels during the study period at the different hospitals in comparison to our study. As in our study, Vollmer et al. also found that calendrical variables were among the features with the strongest influence on forecasting performance. Weather and climate data and google search data had a relatively low influence on forecasting performance.

Similar results were found by Boutsioli et al.[34], who used a simple OLS regression to forecast hospital admissions to the emergency departments of ten public hospitals in Greece. They only used the calendrical variables weekend, summer holiday, public holiday and the participation in emergency care in their model and explained a relatively high variance in hospital admissions of up to 88%.

Jones et al.[35] forecasted the admission numbers at three emergency departments in the USA one, seven, fourteen, twenty-one and thirty days in advance. They used autoregressive integrated moving average (ARIMA) models, time series regression, exponential smoothing, and artificial neural network models to predict admissions per day. They also found that admissions were characterised by yearly and weekly seasonality (see for comparison our Fig. 1). As in our study, they found a relatively low improvement in forecasting performance in the shorter forecasting horizons in comparison to the longer horizons. Similar to our study and to the study of Vollmer et al.[33], weather and climate had a relatively low influence on forecasting performance.

McCoy et al.[36] forecasted hospital discharge numbers at two academic medical centers in the USA. They compared the performance of a PROPHET model with a seasonal ARIMA model and a one-step naïve seasonal forecast and compared monthly models to yearly models. The best performance was achieved by a PROPHET model. Comparable to our study, they also found relatively low to none improvement of forecasting accuracy in refitting their models monthly in comparison to yearly models.

A main similarity between our study and previous studies was the relatively strong influence of calendrical variables in comparison to other potential forecasting features, such as weather and google trend data. The finding is most likely related to the strong dependence of health care systems service patterns to the work-day/weekend difference in service unit performance on the one hand and the relatively weak influence of other factors on actual number of admissions.

Conclusions

Accurate forecasting of hospital admissions can help for the efficient organisation of hospital care and for the adjustment of resources to sudden changes in patient volumes. We found a substantial improvement of forecasting accuracy in comparison to a seasonally adjusted naïve baseline forecast. Model performance did not vary much between different modelling approaches and different forecasting horizons before the COVID-19 pandemic. However, weekly time series models adjusted quicker to the COVID-19 related shock effects. In practice, multiple individual forecast horizons could be used simultaneously, such as a yearly model to achieve early forecasts for a long planning period and weekly models to provided more precise forecasts that adjust quicker to sudden changes.

Data availability

The datasets generated and/or analysed during the current study are not publicly available due data confidentiality by German and EU data protection regulations but could be made available in parts from the corresponding author on reasonable request.

References

1. Hurst, J. Challenges for health systems in member countries of the organisation for economic co-operation and development. *Bull. World Health Organ.* **78**, 751–760 (2000).
2. Dall, T. M. *et al.* An aging population and growing disease burden will require alarge and specialized health care workforce by 2025. *Health Aff.* **32**, 2013–2020. https://doi.org/10.1377/hlthaff.2013.0714 (2013).
3. WHO. Global spending on health 2020: Weathering the storm (Geneva: World Health Organization, 2020).
4. Ekman, B. Cost analysis of a digital health care model in sweden. *PharmacoEconomics Open* **2**, 347–354. https://doi.org/10.1007/s41669-017-0059-7 (2018).
5. Wolff, J. *et al.* Predicting patient outcomes in psychiatric hospitals with routine data: A machine learning approach. *BMC Med. Inform. Decis. Mak.* **20**, 21. https://doi.org/10.1186/s12911-020-1042-2 (2020).
6. Wolff, J. *et al.* Predicting the risk of drug–drug interactions in psychiatric hospitals: A retrospective longitudinal pharmacovigilance study. *BMJ Open* **11**, e045276. https://doi.org/10.1136/bmjopen-2020-045276 (2021).

7. Jack, E. P. & Powers, T. L. A review and synthesis of demand management, capacity management and performance in health-care services. *Int. J. Manag. Rev.* **11**, 149–174. https://doi.org/10.1111/j.1468-2370.2008.00235.x (2009).
8. Bernstein, S. L. *et al.* The effect of emergency department crowding on clinically oriented outcomes. *Acad. Emerg. Med.* **16**, 1–10. https://doi.org/10.1111/j.1553-2712.2008.00295.x (2009).
9. Needleman, J. *et al.* Nurse-staffing levels and the quality of care in hospitals. *N. Engl. J. Med.* **346**, 1715–1722. https://doi.org/10.1056/NEJMsa012247 (2002).
10. Michel, J.-P. & Ecarnot, F. The shortage of skilled workers in Europe: its impact on geriatric medicine. *Eur. Geriatr. Med.* **11**, 345–347. https://doi.org/10.1007/s41999-020-00323-0 (2020).
11. Wolff, J. *et al.* Work-time distribution of physicians at a German University Hospital. *Dtsch. Arztebl. Int.* **114**, 705–711. https://doi.org/10.3238/arztebl.2017.0705 (2017).
12. Wolff, J. *et al.* A work time study analysing differences in resource use between psychiatric inpatients. *Soc. Psychiatry Psychiatr. Epidemiol.* **50**, 1309–1315. https://doi.org/10.1007/s00127-015-1041-2 (2015).
13. Blumenthal, D. *et al.* Covid-19—implications for the health care system. *N. Engl. J. Med.* **383**, 1483–1488. https://doi.org/10.1056/NEJMsb2021088 (2020).
14. Netters, S. *et al.* Pandemic ICU triage challenge and medical ethics. *BMJ Support. Palliat. Care* **11**, 133–137. https://doi.org/10.1136/bmjspcare-2020-002793 (2021).
15. Elsheikh, A. H. *et al.* Artificial intelligence for forecasting the prevalence of COVID-19 pandemic: An overview. *Healthcare* **9**, 1614. https://doi.org/10.3390/healthcare9121614 (2021).
16. Saba, A. I. & Elsheikh, A. H. Forecasting the prevalence of COVID-19 outbreak in Egypt using nonlinear autoregressive artificial neural networks. *Process Saf. Environ. Prot.* **141**, 1–8. https://doi.org/10.1016/j.psep.2020.05.029 (2020).
17. National Meteorological Service. Climate data centre. https://cdc.dwd.de/portal/ (2021).
18. Massicotte, P., Eddelbuettel, D. gtrendsR: Perform and display google trends queries. *Version 1.4.8.* https://CRAN.R-project.org/package=gtrendsR (2021).
19. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
20. Hearst, M. A. *et al.* Support vector machines. *IEEE Intell. Syst. Appl.* **13**, 18–28. https://doi.org/10.1109/5254.708428 (1998).
21. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B* **67**, 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x (2005).
22. Hyndman, R. J. *et al.* A state space framework for automatic forecasting using exponential smoothing methods. *Int. J. Forecast.* **18**, 439–454. https://doi.org/10.1016/S0169-2070(01)00110-8 (2002).
23. De Livera, A. M., Hyndman, R. J. & Snyder, R. D. Forecasting time series with complex seasonal patterns using exponential smoothing. *J. Am. Stat. Assoc.* **106**, 1513–1527. https://doi.org/10.1198/jasa.2011.tm09771 (2011).
24. Taylor, S. J. & Letham, B. Forecasting at scale. *PeerJ* https://doi.org/10.7287/peerj.preprints.3190v2 (2017).
25. Al-qaness, M. A. A. *et al.* Efficient artificial intelligence forecasting models for COVID-19 outbreak in Russia and Brazil. *Process Saf. Environ. Prot.* **149**, 399–409. https://doi.org/10.1016/j.psep.2020.11.007 (2021).
26. Elsheikh, A. H. *et al.* Deep learning-based forecasting model for COVID-19 outbreak in Saudi Arabia. *Process Saf. Environ. Prot.* **149**, 223–233. https://doi.org/10.1016/j.psep.2020.10.048 (2021).
27. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (2018).
28. Hyndman, R. J. & Koehler, A. B. Another look at measures of forecast accuracy. *Int. J. Forecast.* **22**, 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001 (2006).
29. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1–26. https://doi.org/10.18637/jss.v028.i05 (2008).
30. Cleveland, R. *et al.* STL: A seasonal-trend decomposition procedure based on loess. *J. Off. Stat.* **6**, 3–37 (1990).
31. Greenwell, B. M., Boehmke, B. C., McCarthy, A. J. A simple and effective model-based variable importance measure. *arXiv:180504755 [cs, stat].* http://arxiv.org/abs/1805.04755. Accessed 14 Jun 2021.
32. Gul, M. & Celik, E. An exhaustive review and analysis on applications of statistical forecasting in hospital emergency departments. *Health Syst.* **9**, 263–284. https://doi.org/10.1080/20476965.2018.1547348 (2020).
33. Vollmer, M. A. C. *et al.* A unified machine learning approach to time series forecasting applied to demand at emergency departments. *BMC Emerg. Med.* **21**, 9. https://doi.org/10.1186/s12873-020-00395-y (2021).
34. Boutsioli, Z. Forecasting the stochastic demand for inpatient care: The case of the Greek national health system. *Health Serv. Manag. Res.* **23**, 116–120. https://doi.org/10.1258/hsmr.2009.009025 (2010).
35. Jones, S. S. *et al.* Forecasting daily patient volumes in the emergency department. *Acad. Emerg. Med.* **15**, 159–170. https://doi.org/10.1111/j.1553-2712.2007.00032.x (2008).
36. McCoy, T. H., Pellegrini, A. M. & Perlis, R. H. Assessment of time-series machine learning methods for forecasting hospital discharge volume. *JAMA Netw. Open* **1**, e184087. https://doi.org/10.1001/jamanetworkopen.2018.4087 (2018).

Author contributions

J.W., A.K., M.M. and T.K. conceived and designed the study. J.W. analysed and interpreted the data. J.W. and T.K. initiated the research.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-20190-y.

Correspondence and requests for materials should be addressed to J.W.

Reprints and permissions information is available at www.nature.com/reprints.

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